

دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پایان نامه کارشناسی ارشد
در رشته مهندسی کامپیوتر گرایش هوش مصنوعی

عنوان پایان نامه:

آشکارسازی چهره با الهام از مغز

نگارش:

کوروش مشگی

استاد راهنما:

دکتر سعید شیری

پاییز ۱۳۸۹

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

تقدیم

به پدرم که مردانه مردانگی را به من آموخت،

به مادرم که فراز و نشیب پیمودن را به من یاد داد،

به خواهرم که انسانیت را در من زنده نگاه داشت،

به پدر بزرگ و مادر بزرگم که چراغ راه زندگی بودند،

به خاله ام که محبتش از راه دور مایه دلگرمی بود،

و به او که در دلم جایش بود و از رگ گردنم نزدیکتر

سپاسگزاری

با سپاس از جناب آقای دکتر سعید شیری که با حمایت بی دریغ، پشتیبانی علمی، راهنمایی‌های گرانقدر و دلسوزی پدران خود در مسیر انجام این پژوهش مرا همراهی کردند و راه و رسم موفقیت و شهامت را به من آموختند.

و با تشکر از اساتید گرانقدر، دکتر عبادزاده و دکتر غریب زاده، که مرور و داوری این پایان نامه را به عهده گرفته، مرا از پیشنهادات خود بهره مند نمودند.

و با قدردانی از مهندس محسن مالیر که بدون کمکهای ایشان انجام این پروژه میسر نبود.

تصویب نامه

چکیده

یکی از عملکردهای زیربنایی سیستم بینایی، کدگذاری سنگ بنای تصاویر طبیعی یعنی لبه ها، بافتها و شکلهاست که در عملیاتی مانند تشخیص چهره و فهم صحنه های طبیعی از آن استفاده میشود. برای این منظور باید یک بازنمایی مجرد تشکیل داد که از نمونه های خاصی از ورودیهای بینایی تعمیم یابد. از آنچه درباره بخش اولیه بینایی در قشر خاکستری مغز انسان دانسته شده، برمی آید که این سیستم ترکیبی از ویژگیهای تصویر را بکار می گیرد ولی چگونگی تولید بازنماییهای نامتغیر در مغز انسان همچنان ناشناخته باقیمانده است. نورونهای سطح بالای بینایی برای تعمیم روی تصاویر مشابه، از تنوعهایی آماری که در نواحی محلی تصویر یافت می شود، استفاده می کند. در مدل پیشنهادی این پژوهش تابع توزیعی که با تصاویر ورودی بیشترین همخوانی را دارد پیدا شده و توسط میزان فعالیت نورونهایی که برای کدگذاری آن در نظر گرفته شده اند بازنمایی می شود. این مدل با تصاویر طبیعی آموزش می بیند و با یادگیری یک مجموعه فشرده از توزیعهایی که معمولاً در تصاویر به آنها برخورد می کنیم، قابلیت تعمیم می یابد. نورونهای مدل خصوصیتهای زیادی را از خود نشان می دهد که شباهت زیادی به سلولهای بخش بینایی قشر خاکستری مغز دارد و با برقراری این تناظر میتوان بسیاری از ویژگیهای سلولهای پیچیده از جمله پدیده های غیرخطی مشاهده شده در آنرا توجیه کرد. این مدل از استراتژیهای کدگذاری در قشر اولیه بینایی انسان و لایه های بالاتر آن الهام گرفته است و بازنمایی حاصل از آن قادر است که تصاویری که دارای توزیعی مشابه هستند را با هم خوشه بندی کند. با استفاده از این مفهوم یک روش آشکارسازی چهره نیز پیشنهاد شده است که با استفاده از این مدل و سایر مدلهای برگرفته از مغز، آشکارسازی چهره ها را انجام می دهد. مدل حاصل بنابر آزمایشاتی که شرح داده خواهد شد، در برابر تغییر نورپردازی و حالت چهره مقاوم است و قابلیت آشکارسازی و مکانیابی چهره ها را در تصاویر دارد. از مهمترین ویژگیهای این روش مقاومت آن در برابر پس زمینه های شلوغ و یافتن چند چهره موجود در تصویر حتی با وجود پوشانده شدن بخشی از آنها می باشد.

کلمات کلیدی: فهم تصویرهای طبیعی، کدگذاری توزیع تصویر، کدگذاری بهینه، آشکارسازی چهره تمام رخ، مدل بالا به پایین نظری قشر بینایی مغز

«چه زمانی می توانیم مغزهایی مانند مغز خود بسازیم؟»

هنگامی که فیزیکدانان پرده از رازهای جدید طبیعت بر می دارند، این دانش می تواند برای ساختن دستگاههایی که کارهای جالبی می کنند استفاده شوند، مانند هواپیماهایی که پرواز می کنند و یا رادیوهایی که به میلیونها شنونده می رسند. هنگامی که ما به این آگاهی برسیم که مغز چگونه کار می کند، می توانیم دستگاههای خارق العاده ای با توانایی های شناختی بسازیم، مانند اتومبیلهای باشعور که از ما هم بهتر رانندگی می کنند چرا که با اتومبیلهای دیگر ارتباط برقرار می کنند و درباره اطلاعات ترافیک و جاده دانش خود را به اشتراک می گذارند. در سال ۲۰۰۸ آکادمی ملی مهندسان آمریکا یکی از چالشهای بزرگ خود را مهندسی معکوس مغز انسان قرار داد. اما چه زمانی این امر به تحقق می پیوندد؟ برخی از دانشمندان پیش بینی می کنند که نخستین موج نتایج به لطف پیشرفتهای سریع علوم مغز و علوم کامپیوتر و آن در این دهه بدست می آید. این مسئله در وهله اول بسیار بعید به نظیر می رسید ولی هر چه به پیش می رویم بیشتر دست یافتنی می شود، آنقدر دست یافتنی که از هم اکنون جدال برای کسب عنوانهای تاریخی «اولین نفرها» آغاز شده است. پشت صحنه این زورآزمایی فرایندی دراماتیک نهفته است. عصب شناسان مغز را تا اجزای آن، حتی مولکولها می کاوند و با رویکردی پایین به بالا می کوشند تا پرده از چگونگی عملکرد آنها بردارند. پژوهشگران برای تهیه دیاگرامهای سیم کشی مغزهای بزرگ، از موش گرفته تا گربه و در نهایت انسان در حال رقابت هستند و زمینه ای تحقیقاتی به نام کانکتومیکز^۱ را پایه گذاری کرده اند. تکنیکهای جدیدی برای ضبط اطلاعات به صورت همزمان از تعداد زیادی نورون و همینطور تحریک یا غیرفعال کردن نورون خاصی پدید آمده اند. شور و هیجان زیادی به پا شده است و کم کم می توان حس کرد که ما شروع به درک عملکرد مغز انسان در سطح مدار کرده ایم. مدلسازان مغز تا کنون توانسته اند شبکه های کوچکی با تنها چند هزار نورون مدل کنند ولی این رویه به سرعت در حال تغییر است. در حال حاضر، رایانه های رقمی از لحاظ توان محاسباتی، حافظه و پهنای باند ارتباطی به طور نمایی رشد می کنند. تا همین اواخر اینکار با تسریع نرخ کلاک انجام می شد که در طول حیات خودمان نیز شاهد نمو آن از کیلوهرتز تا گیگاهرتز بوده ایم. اما نرخ کلاک مدتی است به سطح ثابتی رسیده است و اکنون پیشرفتهای در زمینه توان محاسباتی از افزایش تعداد پردازنده ها و ارتقاء توانایی توزیع محاسبات بین آنها نشأت می گیرد. سریعترین ابررایانه ها صدها هزار پردازنده دارند و واحدهای پردازش گرافیکی^۲ به رایانه های شخصی سرعتی معادل ابررایانه های ده سال قبل می دهند. اگر قانون مور^۳ درباره رشد نمایی توان محاسباتی برقرار باشد، زمانی فرا می رسد که رایانه ها

¹ connectomics

² graphics processing units (GPUs)

³ Moore's Law

آنقدر قوی خواهند بود و دانش ما درباره مغز آنقدر کامل خواهد بود که بتوان دستگامهایی بر مبنای اصول پردازش عصبی ساخت. این دستگامها همانند مغز به جای منطق قطعی بر مبنای منطق احتمالاتی استوار خواهند بود و به جای استدلال استقرایی از استدلال استنتاجی و قیاسی بهره خواهند برد.

حال به این نزاع جدید که به «جنگ گربه ای»^۴ معروف شده است می پردازیم. نوامبر قبل، یکی از محققان IBM به نام دارمندرا مودا^۵ در یک کنفرانس ابررایانش اعلام کرد که تیم او برنامه ای نوشته است که مغز یک گربه را شبیه سازی می کند. این اخبار باعث شگفتی بسیاری شد زیرا او از مغز موش فراتر رفته بود و گوی سبقت را در این مهم از سایر گروه ها ربوده بود. برای انجام این کار، مودا جایزه گوردن بل بنیاد ACM را که به دستاوردهای شگفت انگیز در زمینه برنامه های پردازش با کارایی بالا اعطا می شود را از آن خود کرد. با این حال ادعای بی باکانه او توسط هنری مارکرام^۶ به چالش کشیده شد. این دانشمند یک عصب شناس در دانشگاه پلی تکنیک لوزان^۷ و رهبر پروژه مغز آبی^۸ است که در سال ۲۰۰۹ به صراحت گفته بود: «ساخت مغز یک انسان غیر ممکن نیست و ما می توانیم در ۱۰ سال آینده آنرا انجام دهیم.» مارکرام در نامه سرگشاده ای به سرپرست فنی IBM برنارد مایرسون، مودا را به فریب گسترده متهم کرده بود و مقاله او را یک شوخی فریب آمیز و یک شایعه خوانده بود. این قضیه منجر به یک جناح بندی در فضای وبلاگ شد و سوژه داغی برای کسانی که در مرز بین علوم مغزی و رایانه ای حرکت می کردند باقی ماند.

دستمایه این جدال این سوال است که: مدل کردن مغز گربه چه معنایی دارد؟ هر دو گروه تعداد زیادی نورون و اتصالات آنها را مدل کرده اند. هر دو مدل از آنچه در واقعیت اتفاق می افتد بسیار بسیار کندتر هستند. مدل مودا تنها از یک سوما^۹ - بدنه سلول حاوی هسته سلول- و نیزه های ساده سازی شده پاسخ نورونی تشکیل شده است. از طرف دیگر مدل مارکرام بازسازی دقیقی از نورونهاست، دارای سیستم پیچیده اتصالات انشعابی است که دندریتها^{۱۰} خوانده می شوند و حتی طیف کاملی از مکانیسمهای دروازه بندی و ارتباطاتی نظیر کانالهای یونی را در بر می گیرد. سیناپسها و اتصالات بین نورونها در مدل مودا در مقایسه با سیناپسهای بیوفیزیکی دقیق مدل مارکرام، ساده سازی شده است. این دو مدل در دو کرانه سادگی و واقع گرایی پیچیده قرار دارند. این جدال یک کشاکش را نمایان می سازد که از یک سو مدلهای ساده سازی شده نورونها برای اجرای سریعتر شبیه سازیها استفاده می شود و از سوی دیگر به جزئیات بیولوژیکی نورون برای شناخت آنها توجه می شود. در واقع فیزیکدانان و مهندسان با نگاه به نورون به دنبال سادگی هستند درحالی که بیولوژیستها با نگاه به همان نورون به دنبال پیچیدگی می گردند. مشکل مدلهای ساده شده اینست که آنها گاهی اصل را نیز هنگام حذف حواشی دور می ریزند. ایراد مدلهای بیوفیزیکی هم اینست که تعداد جزئیات بیشمار است و بسیاری از آنها همچنان ناشناخته باقی مانده است. چقدر از عملکرد مغز با ساده سازی نورونها و مدارها از دست می رود؟ این سوالی است که اگر بتوانیم مودا و مارکرام را وادار به مقایسه مستقیم مدلهایشان بکنیم خواهیم توانست به آن پاسخ دهیم.

⁴ catfight

⁵ Dharmendra Modha

⁶ Henry Markram

⁷ Ecole Polytechnique Fédérale de Lausanne (EPFL)

⁸ Blue Brain Project

⁹ soma

¹⁰ dendrites

متأسفانه شبیه‌سازهای گسترده‌ای که توسط هر دو حال حاضر توسط هر دو گروه صورت پذیرفته است نمایانگر ریتم خواب یا بیهوشی است و بسیار دور از رفتار گربه است چرا که هر دو فاقد ورودی‌های حسی و خروجی‌های حرکتی است. همچنین هردوی آنها فاقد ساختار بخش‌های مغز - مثلاً مخچه که حرکات را هماهنگ می‌کند، آمیگدالا^{۱۱} که وضعیت‌های عاطفی را پدید می‌آورد، و نخاع که سازمان عضلانی را تحت نظر دارد - می‌باشند. با این حال ما از مدل مودا چگونگی برنامه‌نویسی معماری‌های موازی در مقیاس بالا را می‌آموزیم تا بتوانیم شبیه‌سازهایی که مقیاسی به بزرگی تعداد زیاد نورونها و سیناپسها در مغز واقعی دارند را اجرا کنیم. از مدل مارکرام نیز چگونگی تعبیه سطوح مختلف جزئیات درون مدل را فرا می‌گیریم. مودا در مقاله‌اش پیش‌بینی کرده است که بزرگترین ابررایانه می‌تواند عناصر ابتدایی مغز بشر را در سال ۲۰۱۹ بصورت بلادرنگ شبیه‌سازی کند، بنابراین او و مارکرام بالأخره در یک چیز تفاهم دارند. البته این شبیه‌سازها در بهترین حالت می‌تواند مشابه مغز یک کودک یا یک دیوانه^{۱۲} باشد؛ مغز چیزی بسیار بیش از حاصل جمع قسمت‌های مختلفش است.

مطمئناً ساختن یک گربه یا انسان لزومی ندارد و مطلوب نیست چرا که همین الان هم انسانها و گربه‌هایی با کارکرد کامل وجود دارد! این فناوری می‌تواند کاربردهای دیگری داشته باشد. در سال ۲۰۰۵ سیمون هایکین^{۱۳}، مدیر آزمایشگاه سیستم‌های شناختی در دانشگاه مک‌مستر^{۱۴} مقاله‌ای تأثیرگذار بنام «رادیوی باشعور: ارتباطات بیسیم قدرت گرفته از مغز^{۱۵}» نوشت که بستر را برای نسل جدید شبکه‌های بیسیم فراهم می‌کرد. این شبکه‌ها از اصول محاسباتی مغز برای پیش‌بینی مدل استفاده از طیف الکترومغناطیسی بهره می‌گیرد و در استفاده از پهنای باند از روش‌های فعلی کارا تر است. چنین سیستمی چیزی دور از دسترس نیست. برنامه‌هایی برای مناقصه پروژه پیاده‌سازی این سیستم‌های مخابراتی هوشمند در آمریکا تصویب شده است.

بزودی راه‌های مشابه بیشتری برای توسعه سایر امکانات فعلی بوجود خواهد آمد که می‌توان از آن قبیل به شبکه توزیع برق باشعور و اتومبیل‌های باشعور اشاره کرد. احساس کردن و عمل کردن این سیستم‌های باشعور زیرساخت‌های این دنیا را فراهم خواهد کرد. حسگرها جریان اطلاعات - درباره مصرف برق، وضعیت جاده‌ها، الگوهای هوا، شیوع بیماری‌ها، و ... - را روان خواهند کرد و سیستم از این اطلاعات برای بهینه‌سازی هدف خود - کاهش مصرف برق، زمان سفر، و ... - با منظم کردن جریان منابع سود می‌جوید. قسمتهایی از این سیستم‌هایی همینک نیز موجودند ولی همچنان یک سیستم عصبی مرکزی برای تجمیع این سیل اطلاعات و اتخاذ تدابیر لازم به وجود نیامده است. در روزی نه چندان دور، این امر محقق می‌شود و به تدریج با تقلید از نحوه کار مغزهای ما، دنیای اطراف ما هوشمندانه‌تر و کارا تر می‌شود. همانطور که ای زیرساختها تکامل می‌یابند، ممکن است روزی برسد که با مغزهای ما از لحاظ توان و کمال رقابت کند. در نهایت هوشمندی وارث زمین خواهد بود.

تری سجنوسکی

مجله آمریکن ساینتیست، فوریه ۲۰۱۰

¹¹ amygdala

¹² psychotic

¹³ Simon Haykin

¹⁴ McMaster University

¹⁵ Cognitive Radio: Brain Empowered Wireless Communication

[فهرست]

۱. فهرست عناوین

۲. فهرست شکلها

فهرست عناوین:

صفحه	عنوان	زیربخش
۱	مقدمه	فصل ۱
۲	بینایی رایانه ای و مسئله آشکارسازی چهره	۱-۱
۳	سیستم بینایی انسان به عنوان منشا الهام	۱-۱-۱
۴	مدلهای آماری ساده برای توصیف پدیده هایی پیچیده	۲-۱-۱
۸	اهداف و محدوده پروژه	۲-۱
۱۰	ساختار گزارش	۳-۱
۱۱	مباحث نظری	فصل ۲
۱۲	پایه های ریاضی	۱-۲
۱۲	حوزه فرکانس	۱-۱-۲
۱۲	فرکانس مکانی	۱-۱-۱-۲
۱۲	پنجره های سینوسی	۲-۱-۱-۲
۱۴	استنتاج در مدل های مولد	۲-۱-۲
۱۷	تحلیل مؤلفه اصلی	۳-۱-۲
۱۷	تحلیل مؤلفه مستقل	۴-۱-۲
۱۸	انگیزه	۱-۴-۱-۲
۲۰	تعریف تحلیل مؤلفه مستقل	۲-۴-۱-۲
۲۲	ابهامات تحلیل مؤلفه مستقل	۳-۴-۱-۲
۲۴	تعریف و خصوصیات زیربنایی استقلال	۴-۴-۱-۲
۲۴	متغیرهای ناهمبسته جزئی مستقلند	۵-۴-۱-۲
۲۵	چرا متغیرهای گوسی ممنوع هستند؟	۶-۴-۱-۲
۲۶	مؤلفه های غیرگوسی مستقلند	۷-۴-۱-۲
۲۷	معیارهای غیرگوسی بودن	۸-۴-۱-۲
۳۱	کمینه کردن اطلاعات متقابل	۹-۴-۱-۲

۳۳	تخمین درستمایی بیشینه	۱۰-۴-۱-۲
۳۴	ارتباط ICA و تعقیب افکنش	۱۱-۴-۱-۲
۳۵	نتیجه گیری	۱۲-۴-۱-۲
۳۶	روش مطالعات عصب شناسی	۲-۲
۳۶	مطالعات عملکرد گروهی نورونها	۱-۲-۲
۳۷	مطالعات تک نورونها	۲-۲-۲
۳۷	مطالعات آسیب شناسی	۳-۲-۲
۳۸	استفاده از مدل برای پیش بینی خواص قشر بینایی مغز	۴-۲-۲
۳۹	شبکه های عصبی برای بازنمایی	۳-۲
۳۹	قوانین یادگیری	۱-۳-۲
۳۹	قانون یادگیری هب	۱-۲-۳-۲
۴۱	قوانین یادگیری برای کشف صریح وابستگیهای آماری	۲-۲-۳-۲
۴۳	خواص مغز	۲-۳-۲
۴۴	مقدمه ای بر بینایی انسان	۴-۲
۴۴	فرایند بینایی	۱-۴-۲
۴۴	چشم	۱-۱-۴-۲
۴۷	قسمت انتهایی عصب بینایی	۲-۱-۴-۲
۵۱	قسمتهای گوناگون بینایی قشر خاکستری مخ	۳-۱-۴-۲
۵۵	مدلهای رفتاری بخش بینایی قشر خاکستری مخ	۲-۴-۲
۵۸	لایه V1	۳-۴-۲
۶۳	لایه V2	۴-۴-۲
۶۵	لایه V4	۵-۴-۲
۶۶	لایه IT	۶-۴-۲
۶۸	ویژگیهای آماری تصاویر طبیعی	۵-۲
۶۸	مقدمه	۱-۵-۲
۷۰	آمارگان تک پیکسلی	۲-۵-۲

۷۰	قاعده مندیه‌های آماری	۱-۲-۵-۲
۷۱	ارتباط با خواص پاسخ	۲-۲-۵-۲
۷۱	ملاحظات	۳-۲-۵-۲
۷۲	آمارگان مرتبه دوم	۳-۵-۲
۷۲	قاعده مندیه‌های آماری	۱-۳-۵-۲
۷۳	ارتباط با خواص پاسخ	۲-۳-۵-۲
۷۴	ملاحظات	۳-۳-۵-۲
۷۵	آمارگان مرتبه بالاتر	۴-۵-۲
۷۵	قاعده مندیه‌های آماری	۱-۵-۴-۲
۷۷	ارتباط با خواص پاسخ	۲-۴-۵-۲
۷۹	ملاحظات	۳-۴-۵-۲
۷۹	توسعه های بیشتر	۵-۵-۲
۷۹	مدلهای غیرخطی	۱-۵-۵-۲
۸۰	دخیل کردن محدودیت‌های بیولوژیکی	۲-۵-۵-۲
۸۲	وابستگیها بر اساس زمینه	۳-۵-۵-۲
۸۳	وابستگیها بر اساس کلاس اشیا	۴-۵-۵-۲
۸۵	نتیجه گیری	۶-۵-۲
۸۷	بررسی کارهای پیشین	فصل ۳
۸۸	مدلهای نظری عملکرد قشر بینایی مغز	۱-۳
۸۸	مقدمه	۱-۱-۳
۸۸	کدگذاری بهینه	۲-۱-۳
۸۹	مدلهای پایین به بالا	۳-۱-۳
۹۰	مدل VisNet	۱-۳-۱-۳
۹۲	مدل نقشه های قشری مغز	۲-۳-۱-۳
۹۳	مدل HMax	۳-۳-۱-۳
۹۵	مدل نرمال کردن جانبی واریانس با تقسیم	۴-۳-۱-۳

۹۷	مدلهای بالا به پایین	۴-۱-۳
۹۷	سیر تکاملی	۱-۴-۱-۳
۱۲۸	روند پیشرفت	۲-۴-۱-۳
۱۳۱	ساختار سلسله مراتبی	۳-۴-۱-۳
۱۴۳	مروری بر روشهای آشکارسازی چهره	۲-۳
۱۴۳	مقدمه	۱-۲-۳
۱۴۶	روشهای مبتنی بر دانش	۲-۲-۳
۱۴۸	روشهای مبتنی بر ویژگی	۳-۲-۳
۱۴۹	ویژگی های چهره	۱-۳-۲-۳
۱۵۲	بافت	۲-۳-۲-۳
۱۵۲	رنگ پوست	۳-۳-۲-۳
۱۵۴	چند ویژگی	۴-۳-۲-۳
۱۵۶	روشهای مبتنی بر تطبیق کلیشه	۴-۲-۳
۱۵۶	کلیشه های از پیش تعریف شده	۱-۴-۲-۳
۱۵۸	کلیشه های انعطاف پذیر	۲-۴-۲-۳
۱۵۹	روشهای مبتنی بر ظاهر	۵-۲-۳
۱۶۰	چهره های یگه	۱-۵-۲-۳
۱۶۱	روشهای مبتنی بر توزیع	۲-۵-۲-۳
۱۶۳	شبکه های عصبی	۳-۵-۲-۳
۱۶۶	ماشینهای بردار پشتیبان	۴-۵-۲-۳
۱۶۶	شبکه تنک غربالها	۵-۵-۲-۳
۱۶۷	دسته بندی کننده ساده بیزی	۶-۵-۲-۳
۱۶۷	مدل مخفی مارکوف	۷-۵-۲-۳
۱۶۹	روش نظریه اطلاعاتی	۸-۵-۲-۳
۱۷۰	یادگیری استنتاجی	۹-۵-۲-۳
۱۷۱	روشهای مدرن	۵-۲-۳

۱۷۱	جمع بندی	۶-۲-۳
۱۷۵	طراحی و پیاده سازی	فصل ۴
۱۷۶	معماری مدل پیشنهادی	۱-۴
۱۷۹	مدل کردن وابستگیهای کواریانسی	۲-۴
۱۷۹	شرح مسئله	۱-۲-۴
۱۸۱	کارهای پیشین	۲-۲-۴
۱۸۳	تعاریف	۳-۲-۴
۱۸۷	استنتاج متغیر نهفته	۴-۲-۴
۱۹۳	تخمین پارامترها	۵-۲-۴
۱۹۴	رابطه با مدل مؤلفه های کواریانسی	۶-۲-۴
۱۹۶	نتایج روی تصاویر طبیعی	۳-۴
۱۹۶	نتایج روی تصاویر طبیعی	۱-۳-۴
۱۹۹	تحلیل تک تک واحدهای کواریانسی	۲-۳-۴
۲۰۰	کدگذاری جمعیتی	۳-۳-۴
۲۰۳	تعمیم و تمایز نواحی تصویر	۴-۳-۴
۲۰۴	ساخت تصاویر طبیعی	۵-۳-۴
۲۰۶	مدل کردن آشکارسازی چهره	۴-۴
۲۰۶	شرح مسئله	۱-۴-۴
۲۰۶	معیار شباهت به تکه ای از تصویر چهره	۲-۴-۴
۲۰۸	مدل بازنمایی های همخوان	۳-۴-۴
۲۰۹	آشکارسازی چهره	۴-۴-۴
۲۱۳	نتایج روی تصاویر آزمایشی	۵-۴
۲۱۴	تأثیر پارامترها روی نرخ آشکارسازی چهره ها	۱-۵-۴
۲۱۶	تشخیص چهره در شرایط ویژه نورپردازی	۲-۵-۴
۲۱۸	آشکارسازی چهره های دارای حالت خاص	۳-۵-۴
۲۱۸	آشکارسازی چند چهره و چهره های ناکامل	۴-۵-۴

۲۱۹	جمع بندی	فصل ۵
۲۲۰	جمع بندی	۱-۵
۲۲۴	مسئله و راه حلها	۲-۵
۲۲۸	مقایسه با سایر روشها	۳-۵
۲۲۸	استفاده از تحلیل مؤلفه های اصلی در آشکارسازی چهره	۱-۳-۵
۲۳۰	استفاده از تحلیل مؤلفه های مستقل در آشکارسازی چهره	۲-۳-۵
۲۳۱	تصاویر پایه مستقل آماری	۱-۲-۳-۵
۲۳۳	استقلال در فضای چهره در مقابل فضای پیکسل	۲-۲-۳-۵
۲۳۴	کد چهره فاکتوریل	۳-۲-۳-۵
۲۳۶	مقایسه روشها	۳-۳-۵
۲۳۶	آشکارسازی چهره تمام رخ	۱-۳-۳-۵
۲۳۷	آشکارسازی چهره تمام رخ با حضور غیر چهره ها	۲-۳-۳-۵
۲۳۸	کارهای آتی	۴-۵
۲۴۱	فهرست مراجع	مراجع
۲۵۷	واژگان معادل	نمایه
۲۵۸	واژه نامه	۱-ن
۲۶۷	مخفف ها	۲-ن

فهرست شکلها:

شماره	عنوان	صفحه
۱-۱	مفهوم ساختار توسط وابستگیها پدید می آید	۵
۱-۲	تصویر پنجره ها	۱۳
۲-۲	پنجره های سینوسی	۱۴
۳-۲	سیگنالهای اصلی	۱۷
۴-۲	مخلوط مشاهده شده از سیگنالهای اصلی شکل ۳-۲	۱۸
۵-۲	تخمینی از سیگنالهای اصلی	۱۹
۶-۲	توابع پایه ICA از تصاویر طبیعی	۲۰
۷-۲	توزیع توام مؤلفه های x_1 (محور افقی) و x_2 (محور عمودی) با توزیع یکنواخت	۲۳
۸-۲	توزیع توام مخلوطهای مشاهده شده x_1 (محور افقی) و x_2 (محور عمودی)	۲۳
۹-۲	توزیع چند متغیره دو متغیر مستقل گوسی	۲۵
۱۰-۲	تابع توزیع توزیع لاپلاسین	۲۸
۱۱-۲	نمایشی از روش تعقیب افکنش و جالب بودن افکنشهای غیرگوسی	۳۵
۱۲-۲	برش عرضی چشم	۴۵
۱۳-۲	پدیده شکست نور در چشم (چپ)، پدیده تطابق (راست)	۴۶
۱۴-۲	جایگاه ماکولا، صفحه نوری و فوویا روی شبکیه	۴۷
۱۵-۲	میدان گیرندگی چشم	۴۸
۱۶-۲	ساختار کلی LGN	۴۹
۱۷-۲	مکان عقده های سر راه سیگنالهای بینایی	۵۰
۱۸-۲	نواحی دخیل قشر مخ در ادراک بینایی انسان	۵۱
۱۹-۲	نواحی قشری موجود در مسیرهای بینایی و ارتباطات آنها	۵۲
۲۰-۲	اتصالات نواحی بینایی قشر مخ به یکدیگر	۵۳
۲۱-۲	نواحی قشری موجود در مسیرهای بینایی و کانالهای موازی آنها	۵۴
۲۲-۲	مکان نواحی دخیل قشر مخ در ادراک بینایی انسان	۵۵

۵۶	قشر ابتدایی بینایی	۲۳-۲
۵۶	دو مسیر پردازش داده ها در قشر بینایی	۲۴-۲
۵۷	مسیر بطنی در قشر بینایی	۲۵-۲
۵۸	: نواحی بینایی قشر مخ میمون ماکاگو	۲۶-۲
۵۹	ساختار ناحیه دریافت برای چند نورون در ناحیه V1	۲۷-۲
۶۰	ساختار یک نورون پیچیده ارائه شده توسط هابل	۲۸-۲
۶۰	ترکیب قرار گرفتن نورون های V1 بر اساس جهت مورد نظر آنها	۲۹-۲
۶۱	میدان گیرندگی سلولهای ساده V1 به خوبی توسط فیلترهای گابور توصیف می شود	۳۰-۲
۶۱	مدلهای استاندارد نورونهای بینایی اولیه	۳۱-۲
۶۲	اثر حضور محرک ها در ناحیه دریافت گسترش یافته بر پاسخ نورون های V1	۳۲-۲
۶۳	ساختار اتصالات افقی در ناحیه V1 در قشر بینایی	۳۳-۲
۶۴	مجموعه محرک استفاده شده در برای آزمایش خصوصیات نورون های V2	۳۴-۲
۶۵	: مجموعه زوایای استفاده شده برای آزمایش خصوصیات نورون های V2	۳۵-۲
۶۶	مجموعه محرک استفاده شده برای آزمایش نورون های V4	۳۶-۲
۶۷	مجموعه محرک هایی که در نورون های IT بیشترین پاسخ را ایجاد کردند	۳۷-۲
۶۸	سه نمونه از نورون های حساس به دسته اشیا	۳۸-۲
۷۰	شدت روشنایی خطی صحنه های طبیعی به سمت مقادیر کوچک متمایل است	۳۹-۲
۷۳	طیف بزرگی تصاویر طبیعی متناسب با $1/f$ است	۴۰-۲
۷۴	فیلترهای وایتنینگ برای تصاویر طبیعی سازماندهی میدان گیرندگی حول مرکز را از خود نشان می دهد	۴۱-۲
۷۵	فیلترهای PCA تصاویر طبیعی محلی نیستند	۴۲-۲
۷۶	طیف فاز نسبت به طیف بزرگی در ادراک ما سهم بیشتری دارد	۴۳-۲
۷۷	آمارگان مرتبه دوم برای مشخص کردن داده های غیرگوسی کافی نیستند	۴۴-۲
۷۹	روش ICA فیلترهایی محلی و راستدار مشابه میدان گیرندگی سلولهای ساده V1 تولید می کند	۴۵-۲
۸۱	نتایج بازسازی با واحدهایی دارای ظرفیت محدود	۴۶-۲
۸۲	بازنمایی فراکامل پراکنده با واحدهای نويز دار که به تنظیم فرکانس مکانی زیادی دست می اید	۴۷-۲
۸۳	وابستگی پاسخ مؤلفه های مستقل به زمینه در تصویر	۴۸-۲

۸۴	توزیع احتمال توأم پاسخ فیلترهای گیر به مجموعه تصاویر از کلاس های مختلف.	۴۹-۲
۹۱	مدل VisNet	۱-۳
۹۱	یادگیری بر اساس تبدیلات پیوسته	۲-۳
۹۳	مدل نقشه های قشری مغز	۳-۳
۹۴	مدل سلسله مراتبی بر اساس مدل HMax	۴-۳
۹۵	جزئیات مدل ارائه شده بر اساس HMax	۵-۳
۹۶	مدل سلسله مراتبی نرمالسازی با تقسیم	۶-۳
۹۸	مجموعه عام تمام تصاویر طبیعی	۷-۳
۹۹	هیستوگرام تصاویر، اطلاعات بسیار کمی درباره محتوای تصویر دربر دارد.	۸-۳
۱۰۰	نمودار پراکندگی جفتهای پیکسل در سه جابجایی مکانی مختلف	۹-۳
۱۰۲	تخمین طیف توانی برای پنج تصویر نمونه	۱۰-۳
۱۰۳	تصویری نمونه که بطور تصادفی از مدل طیفی گوسی انتخاب شده است	۱۱-۳
۱۰۴	هیستوگرام لگاریتمی یک زیرباند موجک برای چهار تصویر	۱۲-۳
۱۰۶	: مثالی از توزیع داده های دوبعدی و محورهای تحلیل مؤلفه های اصلی	۱۳-۳
۱۰۷	نمودار احتمال پاسخ یک نورون بر اساس بزرگی پاسخ	۱۴-۳
۱۰۸	روش تحلیل مؤلفه های مستقل وقتی روی تکه تصویرهای طبیعی اعمال شود	۱۵-۳
۱۰۹	توابع پایه بدست آمده از بهینه سازی توابع حاشیه ای معیار کورتوسیس	۱۶-۳
۱۱۰	تصاویری که از یک مدل گوسی و یک مدل ICA با ضرایب پراکنده	۱۷-۳
۱۱۱	وابستگی بزرگی بین زوجهای ضرایب	۱۸-۳
۱۱۲	واریانسهای ضرایب ICA از یک محتوا تا محتوای دیگر تصاویر تغییر می کنند	۱۹-۳
۱۱۵	یک تصویر نمونه که از مدل حاشیه ای موجک استخراج شده است	۲۰-۳
۱۱۶	بزرگی ضرایب چندمقیاسه موجک برای تصویر انیشتین	۲۱-۳
۱۱۸	مدل ترکیب متغیرهای گوسی برای توجیه فتار وابستگی واریانس	۲۲-۳
۱۲۰	توزیع توأم تجربی ضرایب موجک که با جفتهای متفاوتی از توابع پایه آزموده شده اند	۲۳-۳
۱۲۳	مدل مولد برای یک GSM دوبعدی	۲۴-۳
۱۲۵	آمارگان توأم شرطی برای تکه های تصویری متفاوت و نویز سفید	۲۵-۳

۱۲۵	مقایسه آمارگان ضرابی که از یک زیرباند تصویر نمونه اخذ شده اند	۲۶-۳
۱۲۶	تخمین محلی و عمومی داه های تصویری	۲۷-۳
۱۲۷	مثالهایی از رفع نویز	۲۸-۳
۱۲۸	مثالی از زیرباند موجک	۲۹-۳
۱۲۹	نمودار ون رابطه بین مدل های توزیع	۳۰-۳
۱۳۰	خلاصه ای از مدل های گوسی	۳۱-۳
۱۳۲	یک بازنمایی گرافیکی از مدل سلسله مرانبی برای وابستگی واریانسی	۳۲-۳
۱۳۴	مجموعه ای از مؤلفه های واریانس پایه که از تصاویر طبیعی استخراج شده اند	۳۳-۳
۱۳۵	مدل انتساب احتمالی یک مدل GSM چند بعدی	۳۴-۳
۱۳۷	الگوهای آماری، نواحی مختلف صحنه های طبیعی را متمایز می کند.	۳۵-۳
۱۳۹	مدل کدگذاری توزیع	۳۶-۳
۱۴۱	هنگامی که پنجره های سینوسی به نورون مدل داده شود...	۳۷-۳
۱۴۲	تعمیم روی تنوع تصاویر طبیعی	۳۸-۳
۱۴۷	تصویر اصلی و تصاویر متناظر آن با دقت پایینتر	۳۹-۳
۱۴۷	یک چهره معمول استفاده شده در روشهای بالا به پایین مبتنی بر دانش	۴۰-۳
۱۴۸	تشخیص یک چهره با جستجو بدنبال دره های نمودار نما	۴۱-۳
۱۵۱	مدل چهره به صورت یک صفحه با شش ویژگی چهره همراستا	۴۲-۳
۱۵۸	یک کلیشه ۱۶×۱۵ برای مکانیابی تصویر	۴۳-۳
۱۶۱	خوشه های چهره و غیر چهره	۴۴-۳
۱۶۲	معیارهای فاصله. دو معیار فاصله بین الگوی تصویر ورودی و خوشه های نمونه ها حساب می شود	۴۵-۳
۱۶۳	تجزیه زیرفضای چهره به دو زیر فضای مؤلفه های اساسی F و فضای مکمل متعامد \bar{F}	۴۶-۳
۱۶۴	نمونه های هر کلاس چهره برای SOM کوهون که هر کدام مرکز یک خوشه هستند	۴۷-۳
۱۶۵	دیاگرام سیستم روش رولی	۴۸-۳
۱۶۸	مدل مخفی مارکف برای شناسایی موقعیت چهره	۴۹-۳
۱۷۸	نمای کلی مدل پیشنهاد شده	۱-۴
۱۸۰	توزیع محلی تصاویر در تصویر چهره نیز الگوهای همبستگی متفاوتی را نشان می دهد	۲-۴

۱۸۵	تصویری نمادین از اثر ضرایب کواریانس بر روی ساختار کواریانس داده ها	۳-۴
۱۹۱	مقادیر متغیرهای نهفته واقعی در مقایسه با تخمینهای $(MAP \hat{y})$ MAP و تخمینهای پیشخور آنها $(ff \hat{y})$	۴-۴
۱۹۲	تقریب پیشخور تخمین MAP در مقایسه با تقریبهای MAP	۵-۴
۱۹۷	هنگامی که مدل با تصاویر طبیعی تطبیق می یابد، بردارهای b_k که توسط مدل...	۶-۴
۱۹۸	مقایسه بردارهای خطی b_k در یک تنظیم کامل و فراکامل	۷-۴
۱۹۹	وزنهای یک نورون معمولی برای بردارهای b_k	۸-۴
۲۰۰	شمایی از اثر یک نورون بر توزیع کدگذاری شده	۹-۴
۲۰۱	تحلیلی از جمعیت واحدهایی که مدل یاد گرفته است	۱۰-۴
۲۰۴	بازنمایی مدل توانسته است روی تنوع نواحی چهره تعمیم یابد	۱۱-۴
۲۰۵	تکه های تصویری نمونه گرفته شده از مدلهای مولد گوناگون	۱۲-۴
۲۰۸	آزمون بازیابی تکه تصویرهای مشابه	۱۳-۴
۲۱۳	مراحل آشکارسازی چهره توسط مدل پیشنهاد شده	۱۴-۴
۲۱۴	: نمونه ای از مجموعه داده های BioID Face Database	۱۵-۴
۲۱۵	تأثیر تغییرات پارامترها روی عملکرد مدل برای تشخیص چهره در تصاویر ساده/چهره/غیر چهره	۱۶-۴
۲۱۶	افراد حاضر در مجموعه داده های Yale Face Database B	۱۷-۴
۲۱۷	نمونه ای از تصاویر یک فرد در مجموعه داده های Yale Face Database B	۱۸-۴
۲۱۸	نمونه ای از مجموعه داده های JAFFE	۱۹-۴
۲۲۹	بازسازی تصویر چهره توسط ترکیب خطی چهره های یگه	۱-۵
۲۳۰	چهره یکه	۲-۵
۲۳۱	مدل ساخت چهره با معماری ۱	۳-۵
۲۳۲	بازنمایی تصاویر پایه مستقل از ضرایب b برای ترکیب خطی تصاویر پایه مستقل u ...	۴-۵
۲۳۴	دو مدل برای اعمال ICA بر روی تصاویر	۵-۵
۲۳۵	مدل ساخت چهره با معماری ۲	۶-۵
۲۳۵	بازنمایی کد فاکتوریل از ضرایب مستقل u تشکیل می شود تا ترکیبی خطی از تصاویر پایه	۷-۵
۲۳۶	نمونه ای از مجموعه داده های AT&T (Olivetti) Database	۸-۵
۲۳۷	نتیجه هر چهار مدل برای آزمایش صحت آشکارسازی	۹-۵

[مقدمه]

۱. بینایی رایانه ای و مسئله آشکارسازی چهره
۲. اهداف و محدوده پروژه
۳. ساختار گزارش

۱-۱- بینایی رایانه ای و مسئله آشکارسازی چهره

در سالیان اخیر، آشکارسازی و شناسایی چهره به زمینه محبوبی از بینایی رایانه ای بدل شده است و یکی از موفقترین کاربردهای تحلیل و فهم تصاویر به شمار می رود. بدلیل ذات مسئله، علاوه بر پژوهشگران علوم کامپیوتر، عصب شناسان و فیزیولوژیستها نیز به مسئله علاقه نشان می دهند. به باور این پژوهشگران، پیشرفت در تحقیقات بینایی رایانه ای بینشی وسیع از چگونگی عملکرد مغز انسان بدست می دهد که البته بر عکس این موضوع، یعنی نقش چشمگیر شناخت مغز انسان در پیشرفت بینایی رایانه ای نیز صادق است.

مسائلی که برای روانشناسان و عصب شناسان در طراحی سیستمهای مربوط به چهره انسان مطرح است، با بر یک مقاله مروری بر مسئله (Yang, Kriegman et al. 2002)، عبارتند از پاسخ به اینکه آیا آشکار سازی و شناسایی چهره یک فرایند اختصاصی در مغز است؟ ادراک چهره نتیجه تحلیل یکپارچه چهره است یا تک تک ویژگیهای آن؟ رتبه بندی ویژگیهای چهره چگونه است؟ وجوه تمایز چهره ها در مغز به چه خصوصیات از آنها باز می گردد؟ نقش تحلیل فرکانس مکانی در این عملیات مغز چیست؟ آیا این مسئله نسبت به نقطه دید نامتغیر است؟ تأثیر تغییرات نورپردازی چگونه است؟ حالات چهره چقدر اطلاعات در بر دارد؟ و غیره.

پژوهشگران متعددی کوشیده اند که فرایند آشکارسازی چهره در مغز انسان را بر اساس یافته های عصب شناسی و دانسته های جامعه علمی از مغز مدل کنند. این روشها یا مستقیماً به مدلسازی پردازش مغز با استفاده از ترکیب واحدهای پردازشی آن یعنی نورنها پرداخته اند و تلاش کرده اند که رفتار مغز را با پیاده سازی خصوصیات شناخته شده از نورونها شبیه سازی کنند، یا با حفظ التزام به ساختار پردازشی مغز به شبیه سازی رفتار مغز پرداخته اند و سعی کرده اند شباهتهای روش خود را با یافته های عصب شناسی بسنجند، یا با الهام گرفتن از انواع پردازشهای درون مغز، مدلهایی ساخته اند که همان رفتار یا تقریبی از آنرا با رویکردی دیگر (مثلاً با استفاد از چارچوب علم آمار) انجام دهند.

چیزی که مسلم است اینست که مغز انسان با بهره گیری از پردازش نورونی و ساختار خاص خود، مسئله آشکارسازی و شناسایی چهره را بدون دشواری و در کمترین زمان ممکن انجام می دهد. رسیدن به این کارایی منوط به رسیدن به درجه ای از پیشرفت سخت افزاری است که بتواند قدرت پردازش موازی عظیم نهفته در مغز را مدل کند و از سوی دیگر برای تهیه نرم افزارهایی که بتواند در چنین بستر سخت افزاری، رفتار مغز را پیاده سازی کند به شناخت هر چه بیشتر مغز نیاز است. در هر حال اکنون بینایی انسان به عنوان یک آزمون انسان محور^{۱۶} برای سیستمهای آشکارسازی و شناسایی چهره رایانه ای استفاده می شود و تا رسیدن به دقت، سرعت، و نرخ اشتباهات پایین مغز در این زمینه راه زیادی در پیش است.

برای پیشبرد تحقیقات در این مقوله به دانش گسترده ای در مورد سیستم بینایی انسان، ساختار مغز انسان خصوصاً بخش بینایی آن، نحوه پردازش سیستم بینایی مغز و مدلهای آن، روشهای تشخیص چهره کنونی و نقاط قوت و ضعف آنها، و

¹⁶ subjective test

ردگیری ایده های استفاده شده در ادبیات پردازش چهره تا پیدا شدن اصول حاکم بر آن نیاز می شود. علاوه بر این نکات باید به نقش خلاقیت، استفاده از شهود، رها کردن مسائل پیچیده و یافتن اصول زیربنایی، و بررسی هم وقوعیهای مشکوک و پدیده های خارج از انتظار (که این پایان نامه به آنها بسیار بها می دهد) در این روال اشاره کرد.

۱-۱-۱- سیستم بینایی انسان به عنوان منشا الهام

سیستم بینایی انسان یک مکانیسم عالی و جالب توجه است. ما در محیطهایی پیچیده، نویزی و با تغییراتی سریع به ادراک، مسیریابی و تعامل می پردازیم در حالی که بینایی ما تنها از ورودیهای ساده که همان پرتوهای نور تابیده شده روی شبکه هستند تغذیه می شود. اطلاعاتی حیاتی که تصمیمات اساسی ما را هدایت می کند - مکان یک میوه روی درخت، سرعت یک اتومبیل نزدیک شونده، چهره یک فرد آشنا و ... - تماماً در این سیگنال ورودی مخفی شده اند. سوال اینجاست که درک جامع و پایدار ما از اشیا، سطوح، بافتها، سرعت، فاصله و ... چگونه در قشر خاکستری مغز ما شکل می گیرد؟ به بیانی دقیقتر، چه خرده اطلاعاتی توسط فعالیت هر نورون خاص منتقل می شود؟ چگونه یک جمعیت نورونی در کنار هم می توانند یک ساختار تصویر را بازنمایی کنند؟ چه اصل محاسباتی در مدارهای نورونی پیاده سازی می شوند؟

یکصد سال پژوهش تجربی نشان داده است که چه بخشهایی از مغز در پردازش اطلاعات تصویری دخیلند، ولی محاسبات دقیقی که زیربنای این پردازش را می سازد همچنان تا حد زیادی در پرده ابهام باقی مانده است. مستقیم ترین روشی که تا کنون از آن استفاده شده، اندازه گیری پاسخهای الکترو-فیزیولوژیکی نورونهای بینایی به ازای تصویر ارائه شده است. این روش نقشی اساسی در نقشه برداری از پردازش بخش نخستین سیستم بینایی است. در شبکه و عقده زانویی خارجی (نخستین مراحل مسیر بینایی، ر.ک. فصل ۲)، گامهای محاسباتی ابتدایی تا حد زیادی شناخته شده است (هر چند که پدیده های زیادی نیز وجود دارد که توجیهی برای آنها یافت نشده است). نورونهای این نواحی را می توان با توجه به پاسخ آنها به محرکهای ساده مشخص کرد و مدلهایی برای توجیه بسیاری از خصوصیاتهای پاسخها پدید آمده است (برای مثال ر.ک. به (Carandini 2004)). با اینحال در مراحل بعدی مسیر بینایی، در نواحی بینایی قشر خاکستری مخ، فعالیتهای نورونی نشان دهنده وجه های بسیار پیچیده تر و مجردتری از ورودی بینایی است. پاسخها گزینندگی بیشتری می یابند؛ یعنی سلولها به اغلب تصاویری که در آزمایش موجودند پاسخ ضعیفی می دهند و پاسخ به محرکهای ساده الزاماً به پردازش تصاویری طبیعی، پر از ساختار، و مرتبط از لحاظ اکولوژیکی ربطی ندارد. یک شخصیت گزاری مناسب برای نورونها محتاج ضبط کردن پاسخ آنها به تمام تصاویر ممکن است که اینکار برای یک نورون هم امکانپذیر نیست، چه برسد به تمام جمعیت میلیونی نورونها!

یک راه معمول برای کنار آمدن با این معضل، استفاده از مجموعه محدودی از تصاویر (نظیر پنجره ها، میله ها و نقطه ها) است به امید اینکه ابعاد مرتبط با پاسخهای نورونی را به خوبی نمونه گیری کند. بدیهی است که اینکار از قبل نیاز به دانش درباره نوع ساختارهای تصویر دارد که نورونها کدگذاری می کنند و اگر حدسهایی ما گمراه کننده باشد این روش به شکست منجر خواهد شد.

یک روش ساخت یافته تر استفاده از مدل‌های نظری است، که حول اهداف محاسباتی بنیادین شکل می‌گیرند تا پژوهش‌های تجربی را سمت و سو بدهند و به توجیه نتایج کمک کنند. هدف از این کار تشخیص اصول سازماندهنده ایست که پشت خصوصیات پاسخهای نورونی قرار دارد در حالی که فرضهای زیربنایی را صریحاً مشخص می‌کند. مدل‌های نظری در مورد اینکه کدام بُعد در فضای محرک برای پاسخ نورونی مهم است و باید بررسی شود، پیش بینی‌هایی انجام می‌دهند؛ چگونگی قرار گرفتن فعالیت هر نورون را در میان بازنمایی جمعیت نورونها توصیف می‌کنند؛ و مهمتر از همه نقشهای کارکردی را برای خصوصیات پاسخهای مشاهده شده پیشنهاد می‌کنند.

پردازش نورونی در قشر بینایی مغز در دهه‌های اخیر بسیار مورد مطالعه قرار گرفته است و خصوصیت‌های پیچیده‌ای در پاسخهای نورونی کشف شده است ولی هنوز راه زیادی تا پدیدار شدن یک تصویر کاملاً منطقی و توجیه پذیر از عملکرد نورونها باقی مانده است.

۱-۱-۲- مدل‌های آماری ساده برای توصیف پدیده‌های پیچیده

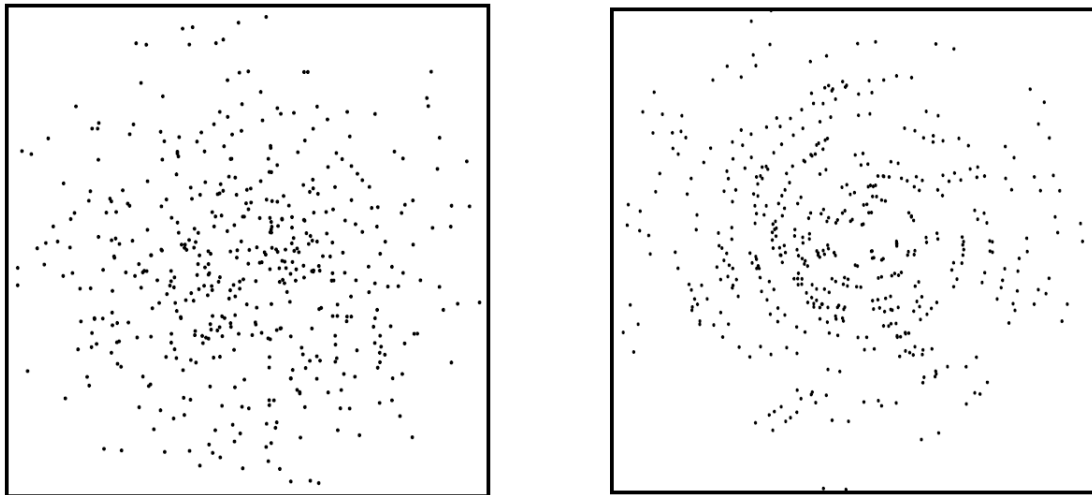
هوراس بارلو^{۱۷} معتقد است که افزونگی در بردارنده دانش است (Barlow 1989). با توجه به این نظریه، افزونگی آماری حاوی اطلاعاتی درباره ساختار محیط است. اطلاعات مهمی که یک سیستم ادراکی باید آنرا بیابد «هم وقوعیهای مشکوکی» است که قاعده مندیهایی آماری جدیدی در ورودیهایی حسی محیط است، که با آنچه سیستم با آن سازگار شده است متفاوت است. به عنوان مثال میله‌ها و لبه‌ها مکانهایی در ورودیهایی تصویری هستند که در آن بین مقیاسهای مختلف مکانی همراستایی فاز وجود دارد و می‌توان آنرا یک «هم وقوعی مشکوک» نامید (Barlow 1994). مکانیسمهای یادگیری‌ای که افزونگی که در ورودی محتمل است را کدگذاری می‌کنند و آنرا از خروجی حذف می‌کند، به سیستم امکان تشخیص مطمئن تر این قاعده مندیهایی جدید را می‌دهد.

تقلیل افزونگی به عنوان یک اصل عمومی سازماندهی برای یادگیری بدون ناظر معرفی شده است. هوراس بارلو نشان داده است (Barlow 1989) که افزونگی آماری حاوی اطلاعاتی درباره الگوها و قاعده مندیهایی محرک‌های حسی است. محرکهایی که هیچ افزونگی ندارند از نویز اتفاقی غیرقابل تمایزند و بارلو ادعا می‌کند که ادراک ساختارها نیز ناشی از وجود افزونگی است. مجموعه نقاطی که در سمت چپ شکل ۱-۱ قرار دارد به صورت اتفاقی از یک توزیع گوسی انتخاب شده است، و نیمی از نقاطی که در تصویر راست دیده می‌شوند با دوران دادن مجموعه‌ای از نقاط اولیه حول مرکز توزیع تولید شده است. این وابستگی ساده بین جفت نقطه‌ها این ظاهر دارای ساختار را بوجود آورده است.

مطابق نظریه بارلو، آنچه کشف آن برای سیستم حائز اهمیت است، قاعده مندیهایی جدیدی در داده‌های حسی است که با محیطی که سیستم با آن سازگار شده است، متفاوت باشد. بارلو‌ای وابستگیهای جدید را «هم وقوعیهای مشکوک»^{۱۸} نامیده است. به عنوان مثال میله‌ها و لبه‌ها مکانهایی در ورودیهایی تصویری هستند که در آن بین مقیاسهای مختلف مکانی همراستایی فاز وجود دارد و می‌توان آنرا یک هم وقوعی مشکوک نامید (Barlow 1994).

¹⁷Horace Barlow

¹⁸suspicious coincidents



شکل ۱-۱: مفهوم ساختار توسط وابستگیها پدید می آید (چپ) مجموعه ای از نقاط از یک توزیع گوسی (راست) نیمی از نقطه ها از توزیع گوسی انتخاب شده اند و نیم دیگر از دوران ۵ درجه ای حول مرکز توزیع پدید آمده اند.

مکانیسمهای یادگیری که افزودنی مورد انتظار در ورودی را کدگذاری می کنند و آنرا از خروجی حذف می کنند به سیستم امکان کشف مطمئن تر قاعده مندیهایی جدید را می دهد. یاد گرفتن چنین تبدیلی معادل مدلسازی دانش پیشین وابستگیهای آماری ورودیهاست (Barlow 1989). کدهای مستقل برای کدگذاری اشیای پیچیده که از ترکیب مرتبه بالای ویژگیها بدست آمده اند سود بیشتری دارند زیرا احتمال اولیه هر یک از ترکیبهای مرتبه بالا کم است. دادهای حسی ورودی بطور خودکار با یک فرضیه پوچ استقلال آماری مقایسه می شود و هم وقوعیهایی مشکوک که یک فاکتور علی جدید را نشان می دهند را با اطمینان بیشتری می توان پیدا کرد.

بارلو یک اصل سازماندهی برای یادگیری بدون ناظر بر پایه نظریه اطلاعات پیشنهاد کرده است. اطلاعاتی که توسط هر پاسخ بدست می آید را به صورت تعداد بیت‌های لازم برای مخابره یک واقعه که احتمال $p(\vec{x})$ تعریف کرد که فرستنده و گیرنده روی توزیع آن موافقت کرده اند:

$$I(\vec{x}) = -\log_2 p(\vec{x}) \quad (1-1)$$

اطلاعات با احتمال نسبت معکوس دارد و می توان آنرا به عنوان یک «غافلگیری» در نظر گرفت چرا که وقوع رویدادهایی که حامل اطلاعات بیشتری هستند کمتر خواهد بود. آنتروپی توزیع یک پاسخ، $H(\vec{x})$ ، برابر مقدار مورد انتظار اطلاعات می باشد:

$$H(\vec{x}) = -\sum p(\vec{x}) \log_2 p(\vec{x}) \quad (2-1)$$

آنتروپی در توزیع یکنواخت به بیشترین مقدار خود می رسد و در توزیعهایی با کورتوسیس بالا^{۱۹} (دارای قله های تیز) کمترین مقدار خود را می یابد. آنتروپی توأم بین دو متغیر اینگونه محاسبه می شود:

$$H(x_1, x_2) = H(x_1) + H(x_2) - I(x_1, x_2) \quad (3-1)$$

که در آن $I(x_1, x_2)$ اطلاعات متقابل بین x_1 و x_2 است که از رابطه ۱-۱ با استفاده از توزیع احتمال توأم $p(x_1, x_2)$ بدست می آید. همانطور که اشاره شد، بارلو کدگذاری کمینه آنتروپی را به دید یک استراتژی بازنمایی عمومی می نگریست. کدهای با آنتروپی کمینه و کورتوسیس بالا دارای اطلاعات متقابل اندکی بین عناصر خود هستند. دلیل این امر اینست که آنتروپی توأم یک کد چندبُعدی به صورت مجموع تک تک آنتروپی ها منهای اطلاعات متقابل بین عناصر آن تعریف می شود (رابطه ۱-۳). از آنجایی که آنتروپی توأم کد، با کمینه کردن جمع تک تک آنتروپیها ثابت باقی می ماند، اطلاعات متقابل نیز کمینه می شود. راه دیگری برای توجیه این قضیه جابجا کردن افزونگی بین عناصر به افزونگی بین توزیعهای تک تک عناصر است. توزیعهای تک تک عناصر با آنتروپی کمینه از این لحاظ افزونه است که آنها اغلب مقادیر مشابهی را اختیار می کنند.

بارلو به کاهش افزونگی در سطوح مختلفی از سیستم بینایی اشاره کرده است. افزونگی مرتبه اول همان میانگین روشنایی است. مکانیسمهای سازگاری از این ویژگی غیرتصادفی با استفاده از آن به عنوان امید ریاضی و توصیف سایر مقادیر نسبت به آن بهره می جویند (Barlow 1989). کنترل بهره کنتراست^{۲۰} که در V1 نیز شرح داده شده است (Heeger 1992) حساب تغییرات واریانس سیگنالهای ورودی را ننگه می دارد. تحلیل مؤلفه اصلی یک راه کدگذاری وابستگیهای مرتبه دوم در ورودی است که محورها را متناظر با راستاهای بیشینه کواریانس می چرخاند. تحلیل مؤلفه اصلی یک کد با ابعاد کاهش یافته تهیه می کند که همبستگی های بین ورودیها را تفکیک می کند.

اتیک و ردلیچ^{۲۱} (Atick and Redlich 1992) چنین سیستمهای از بین بردن همبستگی را به عنوان استراتژی کدگذاری عمومی برای سیستمهای تصویری معرفی می کنند. روش اتیک و ردلیچ از دیدگاه کدگذاری مؤثر به کاهش افزونگی پرداخته است. آنها دریافته اند که محرکهای طبیعی بسیار افزونه اند و در نتیجه نمونه های سیگنالهایی که توسط آرایه ای از دریافت کننده ها دریافت می شوند ناکارا هستند. اتیک مزایای تکاملی کدگذاری مؤثر مانند کنار آمدن با گلوگاههای اطلاعاتی را - که به دلیل محدودیت پهنای باند و بازه دینامیکی بوجود می آیند - شرح داده است (Atick 1992). وی اصل کارایی بازنمایی اطلاعات را به عنوان اساس طراحی کدگذاری حسگرها مطرح می کند و شواهدی طبیعی مانند شبکه مگس گوشت^{۲۲} و پستانداران برای آن می آورد. سلولهای بزرگ تک قطبی^{۲۳} در چشم مرکب مگس گوشت، ناکارایی را به کمک استفاده نامساوی از سطوح پاسخ اعصاب از بین می برد. مؤثرترین بهره پاسخ زمانی رخ می دهد که توزیع احتمال خروجیها برای تمام حالتها ی خروجی ثابت باشد. راه حل اینست که بهره تابع انتقال با تابع توزیع تجمعی ورودی منطبق باشد. لافلین

¹⁹highly kurtotic

²⁰contrast

²¹Atick and Redlich

²²Blowfly

²³large monopolarcells (LMC)

توزیع جمععی احتمال کنتراست را در محیط مگس اندازه گرفت و تطابق زیادی بین بهره نوروهای LMC و تابع توزیع جمععی احتمال یافت.

اتیک نتایج مشابهی درباره شبکیه تابع انتقال مدولاسیون²⁴ شبکیه پستانداران را بدست آورده است. چگالی جمععی طیف دامنه صحنه های طبیعی تقریباً I/f است (Field 1987). تابع MTF یک کدگذاری مؤثر را با یکسان سازی توزیع پاسخ خروجی روی فرکانس مکانی ایجاد می کند. اتیک نشان داده است که ضرب کردن مقادیر MTF مشاهده شده به صورت تجربی در I/f یک خروجی نسبتاً هموار برای فرکانسهای کمتر از 3cpd بدست می دهد. اتیک چنین توابع انتقالی را فیلترهای وایتنینگ²⁵ می خواند زیرا توزیع پاسخ خروجی را برای تمام فرکانسها برابر می سازد.

این تحلیلها ابزاری بدست دادند که به وسیله آن بتوان کارایی کدگذاری برای یک ورودی و خروجی را بیشینه کرد. تحلیل مؤلفه اصلی یک وسیله کاهش افزونگی بین چند خروجی است. روش PCA ورودیها را به دوران محورها ناهمبسته می سازد. این روش مجموعه ای از محورها برای کدگذاری ورودیها با ابعاد کمتر در عین کمترین میزان از دست دادن اطلاعات – البته از نظر خطای مربعی – فراهم می کند. روش PCA یک نمونه از استراتژیهای کدگذاری در فرمولاسیون بارلو است که همبستگیهایی که در ورودی انتظار آن می رود را کدگذاری می کند و از خروجی حذف می کند.

برخی از موفقترین الگوریتمهای آشکارسازی و شناسایی چهره بر پایه مکانیسمهایی یادگیر استوار است که به همبستگیهای درون تصویر چهره حساس هستند. به عنوان مثال، بازنمایی هایی مانند چهره یکه (Turk and Pentland 1991)، هولون ها (Cottrell and Metcalfe 1991) و تحلیل ویژگیهای محلی²⁶ (Penev and Atick 1996) بازنماییهای چهره ای داده گرا بر مبنای تحلیل مؤلفه اصلی است. تحلیل مؤلفه اصلی همبستگی ورودیها را تفکیک می کند ولی به وابستگیهای مرتبه های بالاتر – مثلاً رابطه بین سه پیکسل یا بیشتر – توجهی ندارد. لبه ها مثالی از یک وابستگی مرتبه بالا در یک تصویر هستند و از این قبیل وابستگی ها می توان به شکل و انحنا نیز اشاره کرد. در عملیاتی مانند شناسایی چهره، بسیاری از اطلاعات مهم ممکن است در روابط مرتبه بالای پیکسلها پنهان شده باشد. تحلیل مؤلفه مستقل تعمیمی بر تحلیل مؤلفه اصلی است که علاوه بر جداسازی وابستگیهای مرتبه دوم، وابستگیهای مرتبه بالا را نیز در ورودی جدا می سازد. بل و سجنوسکی یک الگوریتم برای تجزیه مؤلفه های مستقل آماری مجموعه داده ساخته اند. این الگوریتم یک قانون یادگیری بدون ناظر بر مبنای اصل انتقال بیشینه اطلاعات در نوروهای سیگموئیدی است. این الگوریتم اطلاعات متقابل بین ورودی و خروجی یک تابع انتقال را با بیشینه سازی آنتروپی توأم خروجی، بیشینه می کند. اینکار تحت شرایط خاصی خروجی های مستقل آماری می سازد. این الگوریتم برای تفکیک سیگنالهای شنیداری ترکیبی تصادفی (مسئله مهمانی کوکتیل، ر.ک. فصل ۲) موفق بوده است و برای سیگنالهای EEG، تصاویر fMRI و تصاویر صحنه های طبیعی به منظور آزمون فیلترهای تصویری که خروجی های مستقل می سازند، نیز به کار رفته است.

²⁴ modulation transfer function (MTF)

²⁵ Whitening filters

²⁶ local feature analysis = LFA

حال سوال اینجاست که چگونه یک سیستم ادراکی بدون اینکه به او گفته شود که چه ویژگی‌هایی را باید تحلیل کند، یا اینکه آیا تشخیص آن درست است، می‌تواند ساخته شود؟ هنگامی که هیچ معلمی برای تطبیق دادن وجود ندارد، هدف دیگری مورد نیاز است تا سیستم ادراکی بتواند ساختار زیرین را استخراج کند. یادگیری بدون ناظر با مفهوم گیبسون^{۲۷} درباره کشف نشانه‌ها در محیط، مرتبط است. چیزی که مسلم است اینست که ساختار و اطلاعات در محرک خارجی قرار دارد و وظیفه سیستم ادراکی اینست که این ساختار را کشف کند. یک رویه برای خودسازماندهی ایجاد مدل‌های مولدی است که احتمالاً داده‌های مشاهده شده را پدید آورده‌اند. پارامترهای این مدل‌های مولد طوری تنظیم می‌شود که درست‌نمایی آن با حفظ محدودیت‌هایی همانند مفروضات ابتدایی درباره معماری مدل، بیشینه گردد. دسته دوم از اهداف به حفظ اطلاعات و کاهش افزونگی مرتبط است.

مدل‌ها به عنوان مسئول محاسبات پردازش‌های عصبی، از لحاظ پیش‌بینی‌هایی که درباره کدهای نورونی و تفسیر خصوصیات مشاهده شده در سلول‌ها می‌کنند، بسیار سودمند هستند. بنابراین استفاده از آنها در بخش‌هایی از مغز که ناشناخته باقی مانده‌اند از اهمیت بسزایی برخوردار است ولی تاکنون مدل‌های اندکی درباره پردازش در نواحی بینایی پیشرفته‌تر مانند V2 یا V4 دست به پیش‌بینی زده‌اند یا به توجیه کارکردی پدیده‌های غیرخطی V1 مبادرت ورزیده‌اند. مدل‌های فعلی چنین توصیف متراکمی از ساختارهای مرتبه‌های بالا بدست نمی‌دهد و نیاز مبرمی به مدل‌های قویتری که بتواند وابستگی‌های مشاهده شده را در بر بگیرد، قاعده مندیهایی آماری مراتب بالاتر را توصیف کند، و بازنمایی سودمندی از ساختارهای پیچیده تصویری ارائه کند، احساس می‌شود. در این پژوهش به مرور مفصلی روی این مدل‌ها پرداخته می‌شود و نقاط ضعف و قدرت هر یک بررسی می‌شود.

تهیه یک مدل محاسباتی برای آشکارسازی چهره بسیار دشوار است زیرا چهره‌ها پیچیده، چندبعدی و محرک‌های بینایی معناداری هستند (Turk and Pentland 1991). آنها را می‌توان رسته‌ای طبیعی از اشیا دانست که در مقابل پنجره‌های سینوسی، «دنیای بلوک‌ها» و سایر محرک‌های ساختگی در پژوهش‌های انسانی و رایانه‌ای قرار دارند. بنابراین برخلاف بسیاری از توابع اولیه بینایی (مانند مدل‌های شبکه‌ای) که می‌توان مدلی دقیق از آنها ساخت، آشکارسازی و شناسایی چهره وظیفه‌ای بسیار سطح بالاست که در حال حاضر برای آن، روش‌های محاسباتی، تنها قیودی را برای فعالیت نورونها در حالتی خاص پیشنهاد می‌کنند. مدل‌های مبتنی بر نظریه اطلاعات در این میان تاکنون چندان به شکل سه بعدی یا هندسه دقیق چهره کاری نداشته‌اند. هدف این مدل‌ها تهیه روشی برای آشکارسازی و شناسایی چهره است که سریع، نسبتاً ساده، و دقیق باشد. در این پژوهش مدلی دیگر از این دست ارائه خواهد شد.

۱-۲- اهداف و محدوده پروژه

چگونه می‌توان ساختار آماری دنیای تصاویر را توصیف کرد؟ برای شروع، در این پژوهش به رنگ و حرکت^{۲۸} پرداخته نمی‌شود (که با وجود اینکه جنبه‌های مهمی از دنیای تصویری هستند ولی فراتر از محدوده این پژوهش می‌باشند) و روی تحلیل

²⁷Gibson's concept

²⁸motion

ساختار تصاویر ثابت - شکل و قالب، بافت، کانتورها، مرزها - تمرکز می شود. یک تصویر خاکستری فام یک آرایه از شدت روشنایی پیکسلهاست و می توان به راحتی آنرا به صورت یک بردار از مقادیر اسکالر دید. هر تصویر، به نوبه خود با یک نقطه در فضای برداری بازنمایی می شود و مجموعه ای از تصاویر یک جمع کلی (مثلاً مجموعه عکسهای صحنه های بیرون از منزل، یا یک آلبوم خانوادگی) توزیع مدل را شکل می دهند. تصاویر پر از ساختار هستند، ولی این ساختارها در توزیع خام شدت روشنایی پیکسلها به وضوح دیده نمی شوند. نخستین تحقیقات نشان داد که این آمارگان تصاویر طبیعی کاملاً با آمارگان تصاویر تصادفی متفاوتند و بسیار معنادار هستند و لذا برای مدلسازی در این پژوهش از آنها بهره برده خواهد شد.

در طراحی یک سیستم بینایی برگرفته از مدل سیستم بینایی انسان، آنرا به چند فاز تقسیم کرده ایم که هر کدام از آنها جداگانه مدل شده اند و با توالی تصویر در آنها چهره های موجود در آن آشکارسازی می شود:

- ۱- از چشم تا انتهای عصب بینایی که وارد قسمت V1 می شود؛
- ۲- ناحیه V1 که شامل دو لایه سلولهای ساده و پیچیده می شود،
- ۳- ناحیه V2 که شامل اتصالات جانبی آن نیز می شود؛
- ۴- ناحیه V4 که فعالیت هر نورون آن نشانگر وجود یک چهره در میدان گیرندگی آنست.

در این پایان نامه کوشیده شده است که سیستمی برای آشکارسازی چهره با الهام از مغز انسان تهیه شود. برای تشکیل فاز ۱ مدل، از مدل کارکردی سلولهای عقده ای شبکه و LGN استفاده می شود. مدل استفاده شده در این فاز، حاصل پژوهشهای عصب شناختی است و نویسنده این مقاله تنها از آن استفاده کرده است. در فاز ۲ مدلی برای سلولهای ساده و پیچیده ناحیه V1 قشر مخ پیشنهاد می شود. این روش ترکیبی از ایده های عصب شناسی و علوم کامپیوتر است و در یک چارچوب آماری مسئله ای بهینه سازی را مطرح می کند که این دو لایه سلول را برای تصاویر طبیعی از جمله چهره های انسانی به بهترین شکل تطبیق می دهد و رسته های متفاوت تکه های تصاویر را به کمک بازنمایی خود تقریباً بطور خطی جدا پذیر می نماید. در فاز ۳ با استفاده از شبکه های عصبی چند لایه پرسپترون، نورونها، خروجی های لایه پایینتر را پردازش کرده و تکه های تصویری مرتبط به چهره را جداسازی می کند. سپس با بهره گیری از اتصالات جانبی درون لایه ای شبکه عصبی، احتمال وجود تکه های تصویر چهره توسط تکه های اطراف تقویت یا تضعیف می شود. در فاز ۴ نیز با یک آستانه گیری برگرفته از مدل استاندارد نورونی، وجود یا عدم وجود چهره و مکان آنها مشخص می شود.

هدف از ارائه این سیستم، نشان دادن نقش نظریه کدگذاری بهینه در کارایی کد حاصل از قسمتهای اولیه بینایی قشر خاکستری مخ انسان است. این کدگذاری بر اساس وابستگی کواریانس فیلترهای خطی حاصله از تصاویر طبیعی است و نشان داده می شود که نه تنها اهداف کدگذاری خود را بخوبی بر آورده می سازد، بلکه برای آشکارسازی چهره نیز بستر مناسبی را فراهم می کند و کارایی آن اندکی بهتر از روشهای مشابه نظریه اطلاعاتی می باشد.

۱-۳- ساختار گزارش

در ادامه این پایان نامه در فصل دوم به مباحث نظری و علمی پیش زمینه ای پرداخته می شود. پس از ارائه مختصری در مورد اصول ریاضی مورد استفاده، به بررسی روشهای استنتاج در مدل‌های مولد پرداخته می شود. پس از آن روش تحلیل مؤلفه های مستقل که الهام بخش قسمت بزرگی از کارهای این پروژه است به طور کامل مورد بررسی قرار می گیرد و روند آن بررسی می شود تا راه حل استفاده شده در پروژه را روشن کند. در ادامه برای روشن شدن نقش مدل‌های عملکردی انواع روشهای بکار گرفته شده در علوم عصب شناسی معرفی می شود. در ادامه این فصل برخی از خصوصیات شبکه های عصبی و روشهایی که در بازنمایی کمک می کنند معرفی شده است. در بخش بعدی این فصل بینایی انسان مورد بررسی قرار گرفته است. این بررسی در دو قسمت اطلاعات بیولوژیکی موجود در مغز و مدل‌های رفتاری آنها آورده شده است. در پایان این فصل به ویژگیهای آماری تصاویر طبیعی پرداخته می شود.

فصل سوم به بررسی پیشینه موضوعات مطروحه در شرح این پایان نامه می پردازد. در بخش نخست، مدل‌های نظری عملکرد قشر بینایی مغز مورد مذاقه قرار می گیرد. در این بخش مروری بر نظریه کدگذاری بهینه، مدل‌های پایین با بالای قشر بینایی (که ایده اتصالات جانبی لایه V2 برخاسته از آنست)، و مرور بسیار جامعی بر مدل‌های بالا به پایین قشر بینایی (که شالوده روش استفاده شده در لایه V1 مدل پیشنهادی این مقاله را می سازد) صورت می گیرد که شامل سیر تکاملی آن، ضعفها و قوت‌های روشها، روند پیشرفت و مدل‌های سلسله مراتبی است. در بخش دوم و انتهایی این فصل نیز به روشهای بسیار متنوعی از ادبیات آشکارسازی چهره پرداخته شده است که بسیاری از آنها منشأ الهام در روشهای مبتنی بر مدل‌های مغز هستند (و روش استفاده شده در لایه V2 مدل پیشنهادی نیز از آنها ریشه گرفته است).

فصل چهارم، به طراحی و پیاده سازی مدل اختصاص یافته است. در این فصل ابتدا معماری مدل پیشنهادی مطرح می شود. سپس مدل وابستگیهای کواریانسی پیشنهادی آورده شده است. در این بخش ابتدا انگیزه استفاده از این مدل خاص مطرح می شود، سپس کارهای پیشین در این زمینه خاص بررسی می شود، مدل پیشنهادی کاملاً توصیف می شود، و رابطه آن با مدل‌های دیگر بیان می شود. سپس نتایج این مدل که اصلی ترین بخش سیستم پیشنهادی است، تحلیل می شود و استفاده از آنها در این سیستم توجیه می شود. در ادامه تشخیص دهنده تکه های چهره مدل شده، معیار بدست آمده برای آن توصیف شده و طریقه استفاده از این معیار برای آشکارسازی چهره بیان می شود. در پایان نتایج آزمایشان مختلف روی این سیستم آورده شده است.

در فصل پنجم، به جمع بندی کارهای انجام شده در این پروژه پرداخته می شود و ایده های استفاده شده در بخشهای مختلف مدل یکجا بیان می شود و راه حل عبور از بن بستهای پیاده سازی و تحلیلی بیان می گردد. سپس این مدل با روشهای مشابه خود که ریشه در نظریه اطلاعات دارند مقایسه می شود. در پایان نیز ضعفهای احتمالی مدل و رهنمودهایی برای ادامه کار در این زمینه ارائه خواهد شد.

[مباحث نظری]

۱. پایه ریاضی
۲. روش مطالعات عصب شناسی
۳. شبکه های عصبی برای بازنمایی
۴. مقدمه ای بر بینایی انسان
۵. ویژگیهای آماری تصاویرهای طبیعی

۲-۱- پایه های ریاضی

۲-۱-۱- حوزه فرکانس

۲-۱-۱-۲ فرکانس مکانی

فرکانس مکانی خاصیت هر ساختاری است که در فضای مکان تناوبی باشد. این فرکانس مکانی معیاری برای اندازه گیری اینست که ساختار در چه فواصلی تکرار می شود و آنرا بر مبنای واحد طول بیان می کند. فرکانسهای مکانی متفاوت مقدار متفاوتی از اطلاعات را درباره ظاهر یک محرک منتقل می کنند. فرکانسهای مکانی بالا تغییرات مکانی ناگهانی در تصویر مانند لبه ها، را نشان می دهند و معمولاً مختص اطلاعات پیکربندی و جزئیات ریز تصاویر هستند. فرکانسهای مکانی کم بر خلاف دسته اول، اطلاعاتی عمومی را درباره شکل نمایش می دهند که از آن قبیل می توان به راستا و تناسب اشاره کرد. در انسانهای بالغ آستانه تمیز فرکانسهای مکانی ۷٪ است.

بسیاری از سلولهای V1 دارای خاصیت گزینندگی^{۲۹} نسبت به فرکانسهای مکانی هستند (سلولهای ساده به فاز حساس هستند در حالی که سلولهای پیچیده اینگونه نیستند). بزرگی پاسخ یک سلول خاص به فرکانسهای مکانی متفاوت یک منحنی میزان را تشکیل می دهد که در یک فرکانس خاص دارای قله است و با زیاد یا کم شدن فرکانی از آن مقدار بهینه، پاسخ کمتری می دهد. این سلولها از دید فرکانسهای بهینه خود، یک دنباله را تشکیل می دهند که کرانههای آن از فرکانس مکانی کمی مانند ۲cpd تا فرکانسهای مکانی بالایی مانند ۱۰cpd متغیر است. پهنای باند یک دنباله فرکانسهای مکانی برابر است با تفاضل فرکانسی بین دو فرکانس هنگامی که بزرگی پاسخ آن برابر نصف مقدار بیشینه است (یکی با فرکانس بیشتر و یکی با فرکانسی کمتر از قله) (Wang 1999).

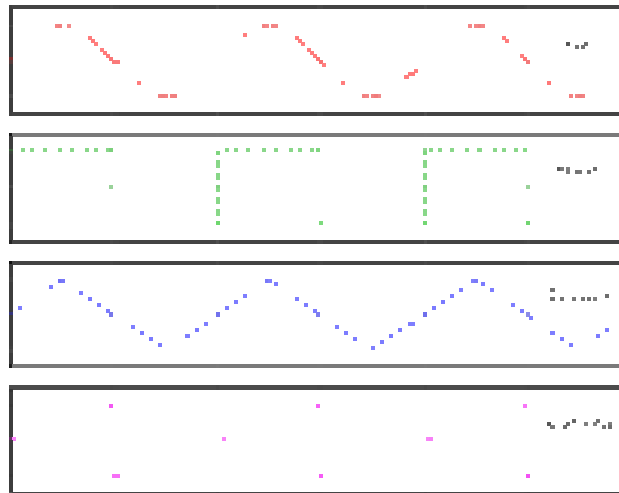
۲-۱-۱-۲ پنجره های سینوسی

یک پنجره، مجموعه ای دارای هندسه فضایی منظم از عناصر یک شکل، موازی و کشیده است. پنجره ها معمولاً دارای یک مجموعه از عناصر کشیده هستند ولی می توانند دارای دو مجموعه نیز باشند که در این حالت مجموعه دوم معمولاً بر مجموعه اول عمود است (که اگر این دو مجموعه عمود باشند می توان آن پنجره را مشبک نامید. یک پنجره می تواند یک تصویر باشد که خصوصیات یک پنجره را دارا باشد. برای مثال یک تصویر ممکن است مجموعه ای از میله های سیاه باشد که توسط میله های سفید هم اندازه جدا شده باشند. این نوع پنجره ها را می توان با یک نمودار نشان داد. روی محور عمودی نمودار شدت روشنایی قرار دارد که با لغزاندن نورسنج در راستای عمود بر راستای پنجره بدست می آید. محور افقی فاصله ای که نورسنج جابجا شده است را نشان می دهد. به عنوان مثال پنجره با موج مربعی دارای خطوط پایین است که با میله های سیاه منتظر است و گوشه هایی ناگهانی آنرا تا خطوط مسطح بالاتر که متناظر با میله های سفید رنگ است بالا می برد. یک تناوب (یا سیکل) چنین پنجره ای از یک میله سیاه و یک میله مجاور سفید تشکیل شده است. پنجره ها در جاهایی که میله

²⁹selectivity

³⁰cycle/degree

های سیاه عرضشان با میله های سفید برابر نیست مثلثی هستند و با تناوب کاری^{۳۱} توصیف می شوند (شکل ۲-۱). تناوب کاری برابرست با نسبت عرض میله سیاه به کل تناوب (یا گام^{۳۲}، جمع عرضهای میله سیاه و میله سفید).



شکل ۲-۱: تصویر پنجره های سینوسی، مربعی، مثلثی و دندان اره ای. محور عمودی شدت روشنایی را نشان می دهد و محور افقی نمایانگر فاصله است.

پنجره ها معمولاً با چهار پارامتر مشخص می شوند:

۱. فرکانس مکانی که برابر تعداد تناوبهایی است که در یک فاصله خاص قرار دارند (مثلاً ۱۰ خط یا تناوب بر میلیمتر)؛
۲. کنتراست که برابر تفاضل روشنایی بین قسمتهای روشن پنجره و قسمتهای خاموش آنست که معمولاً به صورت فاصله میچلسون^{۳۳} بیان می شود که برابرست با: بیشترین روشنایی منهای کمترین روشنایی تقسیم بر حاصلجمع بیشترین روشنایی و کمترین روشنایی؛

$$d = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (1-2)$$

۳. فاز که موقعیت نمودار را نسبت به یک موقعیت استاندارد نشان می دهد و معمولاً با درجه (بین ۰ تا ۳۶۰ برای یک تناوب کامل) یا رادیان (2π برای یک تناوب کامل) اندازه گیری می شود؛
۴. راستا که زاویه ایست که پنجره با یک راستای استاندارد می سازد (مانند محور افقی) و معمولاً با درجه یا رادیان اندازه گیری می شود.

عناصر پنجره می توانند شدت روشناییهایی بجز میله های لبه تیز داشته باشند. اگر نمودار پنجره ها سینوسی باشد، پنجره مانند میله های سیاه و سفید تازی به نظر می آید و پنجره سینوسی نامیده می شود. پنجره سینوسی در آپتیک

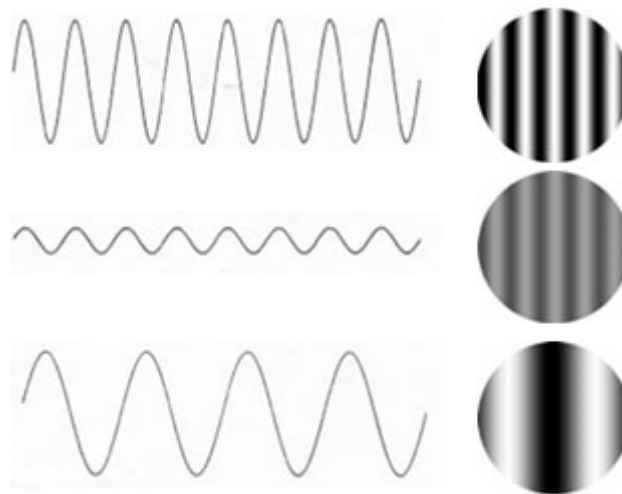
³¹duty cycle

³²pitch

³³Michelson distance

برای تعیین تابع انتقال لنزها کاربرد زیادی دارد. یک لنز تصویری از پنجره سینوسی می سازد که همچنان سینوسی است ولی با توجه به فرکانس مکانی و احتمالاً تغییر فاز، کنتراست آن کاهش می یابد. شاخه ای از ریاضیات که با این قسمت از علم اپتیک سر و کار دارد تحلیل فوریه است. پنجره های سینوسی در پژوهش روی ادراک تصویری نیز بطور گسترده ای مورد استفاده قرار می گیرند. کمپل و رابسون^{۳۴} استفاده از پنجره سینوسی را باب کردند. استدلال آنها این بود که بینایی انسان یک تحلیل فوریه روی تصاویر شبکه انجام می دهد (Campbell and Robson 1968).

در مطالعه ادراک تصویری، غالباً پنجره های سینوسی (تصاویری شامل نوارهایی سفید و سیاه با یک الگوی خاص، شکل ۲-۲) برای بررسی قابلیت های سیستم بینایی استفاده می شود. در این محرک فرکانس مکانی به صورت تعداد چرخه ها در واحد درجه (از زاویه بینایی) بیان می شود.



شکل ۲-۲: پنجره های سینوسی. تاثیر دامنه موج و فرکانس مکانی روی شکل پنجره

۲-۱-۲- استنتاج در مدل های مولد

یکی از روش های یادگیری بدون ناظر سعی می کند یک بازنمایی از داده ها ارائه دهد که در آن توزیع احتمال زیرین داده ها را آشکار کند. در این روش، یک مدل از پیش در نظر گرفته می شود که شکل کلی توزیع احتمال داده ها را محدود می کند. سپس پارامترهای این مدل با بیشینه کردن درستنمایی مدل که داده های مشاهده شده را تولید کرده است بدست می آید. پارامترهای مدل را می توان به عنوان وزن های شبکه نیز در نظر گرفت. مثلاً یک مدل مخلوط گوسی (ر.ک. فصل ۳) فرض می کند که هر نقطه داده از ترکیبی از مؤلفه ϕ_i بوجود آمده است که هر یک از این مؤلفه ها توزیعی گوسی دارای میانگین μ_i و واریانس σ_i^2 و احتمال اولیه π_i می باشد. در این چارچوب شبکه به صورت یک مدل مولد احتمالاتی داده ها دیده می شود. فرض می کنیم $\vec{x} = (x_1 \dots x_n)$ نمایانگر داده های مشاهده شده باشد که نمونه ها در آن مستقلند. احتمال داده ها با داشتن این مدل از رابطه زیر بدست می آید:

³⁴Campbel & Robson

$$p(\bar{x} | \Phi) = \sum_i p(\bar{x} | \phi_i) p(\phi_i) = \prod_j \sum_i p(x_j | \phi_i) p(\phi_i) \quad (2-2)$$

احتمال داده ها بر حسب احتمال اولیه هر یک از زیرمدلهای $p(\phi_i)$ و احتمال ثانویه داده ها با فرض داشتن زیرمدل $p(\bar{x} | \phi_i)$ تعریف می شود که به صورت (u_i, σ_i, π_i) تعریف می شود. پارامترهای هر یک از زیرمدلهای (u_i, σ_i, π_i) ، با شیب نوردی صعودی روی رابطه ۲-۲ پیدا می شود. احتمال لگاریتمی یا درستنمایی معمولاً برای تسهیل محاسبه مشتقهای جزئی رابطه ۲-۲ به ازای هر پارامتر محاسبه می شود. پارامترهای مدل در چارچوب یادگیری بدون ناظر به مثابه وزنه‌های شبکه در نظر گرفته می شود. چنین مدلهایی در دسته مدلهای مولد قرار می گیرند که در آن مدل طوری انتخاب می شود که داده های مشاهده شده به احتمال زیادتری توسط آن تولید شده باشند. مدلهای بیشینه سازی درستنمایی حالت خاصی از مدل استنتاج بیزی است. احتمال رخداد مدل با فرض داشتن داده ها از رابطه زیر بدست می آید:

$$p(\Phi | \bar{x}) = \frac{p(\bar{x} | \Phi) p(\Phi)}{p(\bar{x})} \quad (3-2)$$

تابع هزینه بیشینه درستنمایی، مقدار $p(\bar{x} | \Phi)$ را بیشینه می کند، که اگر فرض کنیم تابع احتمال اولیه مدل $p(\Phi)$ یکنواخت است آنگاه $p(\Phi | \bar{x})$ نیز بیشینه می شود چرا که $p(\bar{x})$ تنها یک فاکتور مقیاس گذاری است. یادگیری رقابتی بیشینه درستنمایی (Nowlan 1990) یک توسعه بر نوع یادگیری مدل مخلوط گوسی است (که در فصل سوم در مورد مدل مخلوط گوسیبیشتر گفته خواهد شد). همانند مدل مخلوط گوسی، احتمال ثانویه $p(x_i | \phi_j)$ توسط یک تابع گوسی به مرکز w_i داده می شود. این رقابت بدین صورت است که احتمالات اولیه زیر مدلهای $p(\phi_i)$ به عنوان جمع وزندها داده های ورودی، به یک عملگر نرمال سازی بیشینه گیری نرم^{۳۵} داده می شوند. این احتمالات اولیه نسبتهای ترکیب، π_i ، را بدست می دهند. شباهتهایی بین قانونهای بروزسانی که از این مشتق جزئی چنین توابع هدفی بدست می آید و سایر قوانین یادگیری بدون ناظر، مانند یادگیری هب^{۳۶} (که در قسمت ۲-۳-۱ شرح داده شده است) وجود دارد. به عنوان مثال قانون بروزسانی یادگیری رقابتی بیشینه درستنمایی از یک مؤلفه هب نرمال شده و یک بخش تنزل وزن بدست می آید.

یک محدودیت مدلهای مولد اینست که بجز برای مدلهای بسیار ساده، هر الگو می تواند از مقداردهی های زیادی به پارامترها بدست بیاید که تنظیم پارامترها برای این تعداد نمایی حالتها برای بیشینه کردن احتمال الگوهای مشاهده شده چندان راحت نیست. ماشین هلمهولتز^{۳۷} (Dayan, Hinton et al. 1995) با بیشینه کردن کران پایین برای احتمال مشاهدات که راحت محاسبه می شود، راه حلی برای غلبه بر این انفجار ترکیبیاتی ارائه کرده است. این روش می تواند به صورت یک یادگیری سلسله مراتبی خود نظارتی در نظر گرفته شود که مسیرهای قشری پیشخور و پسخور مغز را مرتبط سازد. اتصالات «شناسایی» پایین به بالا ورودیها را به بازنمایی هایی در لایه های مخفی بعدی تبدیل می کنند و اتصالات «مولد» بازنمایی یک لایه را برای بازنمایی لایه بالا بازسازی می کنند. این شبکه یک مدل معکوس «شناسایی» برای تخمین توزیع درست احتمال ثانویه داده های ورودی استفاده می کند.

³⁵ soft-maximum normalization

³⁶ Hebbian learning

³⁷ the Helmholtz machine

هینتون^{۳۸} الگوریتم «بیدار-خواب» را برای اصلاح وزنهای پیشخور(شناسایی) و پسخور(مولد) ماشین هلمهولتز پیشنهاد کرده است(Hinton, Dayan et al. 1995). الگوریتم «بیدار-خواب» از تابع هدف کمترین طول توصیف^{۳۹} بهره می برد. هدف یادگیری، کمینه کردن مجموع تعداد بیت‌هایی است که برای مخابره بردارهای ورودی مورد نیاز است که این مخابره ابتدا بازنمایی لایه های مخفی را می فرستد، سپس تفاضل بین بردار ورودی و بردار بازسازی شده از روی بازنمایی لایه های مخفی را ارسال می کند. کمینه کردن طول توصیف، شبکه را مجبور به یادگیری یک بازنمایی مقرون به صرفه می کند که قاعده مندیهایی ضمنی داده را تسخیر کرده است. یک تابع هزینه C به صورت مجموع تعداد کل بیت‌های لازم برای توصیف تمام وضعیت‌های مخفی در تمام لایه های مخفی، α ، به علاوه هزینه توصیف اطلاعات باقیمانده در بردار ورودی d با داشتن وضعیت‌های مخفی تعریف می شود.

$$C(\alpha, d) = C(\alpha)C(d | \alpha) \quad (۴-۲)$$

الگوریتم هزینه تخمینی روی تمام وضعیت‌های مخفی را کمینه می کند:

$$E(C(\alpha, d)) = \sum_{\alpha} Q(\alpha | d)C(\alpha, d) \quad (۵-۲)$$

برای محاسبه هزینه تخمینی باید توزیع احتمال شرطی روی بازنمایی واحدهای مخفی $Q(\alpha | d)$ ، تخمین زده شود. الگوریتم «بیدار-خواب» مقدار $Q(\alpha | d)$ را با انتقال فعالیت واحدهای مخفی از طریق اتصالات شناسایی از ورودی تخمین می زند. این اتصالات شناسایی به نوبه خود با فعالسازی واحدهای مخفی و تخمین توابع توزیع ورودی آموزش می بینند، که این تخمین با تولید «توهم‌هایی»^{۴۰} از ورودی از طریق اتصالات مولد تولید می شود. از آنجایی که واحدها تصادفی هستند، تکرار این فرایند ممکن است منجر به تولید توهم های متفاوتی شود. این توهمها یک نمونه ناریب از مدلی که شبکه از دنیا تهیه کرده است بدست می دهد. در طول فاز «بیدار»، نورونها توسط اتصالات شناسایی هدایت می شوند و مدل شناسایی برای تعریف تابع هدفی که برای یادگیری پارامترهای مدل مولد استفاده می شود، بکار می رود. اتصالات مولد به نحوی تطبیق می یابند که احتمال بازسازی درست بردار فعالیت در لایه پایین را افزایش دهند. در فاز «خواب»، نورونها توسط اتصالات مولد هدایت می شوند و مدل مولد برای تعریف تابع هدف یادگیری پارامترهای مدل شناختی استفاده می شود. اتصالات شناختی طوری آموزش می بینند که احتمال اینکه در لایه بالاتر بردار فعالیت درست را بسازند بیشتر شود. طول توصیف می تواند با فرض داشتن مدل مولد شبکه به صورت یک کران بالا روی احتمال لگاریتم منفی داده ها دیده شود و از همین رو این روش قرابت زیادی به روشهای پیشینه سازی درست‌نمایی برای تطبیق مدل با داده ها دارد(Hinton, Dayan et al. 1995).

می توان نشان داد که مدل‌های استنتاج بیزی هم ارز اصل طول توصیف کمینه هستند(Bartlett 1998). بنابراین همه مدل‌های مولدی که در این پژوهش بررسی خواهند شد تحت زیر مجموعه کدگذاری مؤثر دسته بندی خواهند شد. روش دیگری برای نیل به کدگذاری مؤثر کاهش صریح افزونگی بین واحدها و داده های ورودی است. با در نظر گرفتن چند قید روی

³⁸ Hinton

³⁹ minimum description length

⁴⁰ hallucinations

تعداد واحدهای کدگذار می توان افزودگی را کمینه کرد، همان طوری که در طول توصیف کمینه اینگونه بود، و یا می توان بدون فشردن سازی بازنمایی با بردن به فضایی با ابعاد بالاتر و نمایش به صورت کد پراکنده، افزودگی را تقلیل داد.

۲-۱-۳- تحلیل مؤلفه اصلی

تحلیل مؤلفه اصلی (PCA) یک مجموعه متعامد از محورهای بیشترین کواریانسهای داده ها می یابد. ماتریس X را مجموعه داده ای در نظر می گیریم که در آن هر ستون یک مشاهده و هر سطر یک اندازه گیری با میانگین صفر است. محورهای مؤلفه های اصلی عبارتند از بردارهای یکه ماتریس کواریانس اندازه گیریها، $\frac{1}{N}XX^T$ ، که در آن N تعداد مشاهدات را نشان می دهد. مقادیرهای ویژه متناظر آنها نسبت تنوع در داده هایی را نشان می دهد که هر بردار یکه بدان مربوط است. نخستین مؤلفه اصلی به راستای بیشترین تنوع اشاره می کند، دومین بردار یکه به راستای بیشترین تنوع عمود بر اولی اشاره می کند و الی آخر. داده ها با تصویر کردن بردار هر داده روی این محورهای جدید، به صورت جدیدی نمایش داده می شوند. فرض می کنیم P ماتریس حاوی بردارهای یکه مؤلفه اصلی در ستون خودش باشد. بازنمایی PCA برای هر مشاهده در سطرهای A با این رابطه بدست می آید:

$$A = X^T P \quad (۶-۲)$$

بردارهای یکه درون P می تواند به صورت یک مجموع وزن روی داده ها، X دیده شود که خروجی ها ضرایب درون ماتریس، A را می سازند. از آنجا که بردارهای یکه مؤلفه اصلی متعامدند، در واقع بردارهای پایه مجموعه داده نیز به شمار می روند. این امر می تواند اینگونه نشان داده شود که از آنجایی که P متقارن است و ستونهای آن نیز متعامدند، $PP^T = I$ حاصلضرب P^T در رابطه ۶-۲ از سمت راست برابر است با $AP^T = X$ بنابراین داده های اصلی می توانند با استفاده از بردارهای یکه P (اینبار به عنوان بردار پایه) از روی ضرایب A ساخته شوند. یک بازنمایی با ابعاد کمتر را می توان با انتخاب مجموعه ای از مؤلفه های اصلی که بیشترین مقادیر یکه را دارا هستند بدست آورد. نشان داده شده است که برای هر تعداد بُعد، بازنمایی مؤلفه های اصلی خطای میانگین مربعی بازسازی را کمینه می کند. از آنجایی که بردارهای یکه در فضای کواریانس به جهات عمود برهم اشاره می کنند، بازنمایی مؤلفه های اصلی ناهمبسته است یعنی اینکه ضرایب یکی از محورها را نمی توان به صورت خطی از روی ضرایب محور دیگری پیش بینی کرد. راه دیگری برای نگرش به بازنمایی مؤلفه اصلی از طریق مدل های تخمین چگالی است. روش PCA داده ها را به صورت یک مدل گوسی چند متغیره مدلسازی می کند که ماتریس کواریانس آن الزاماً قطری است. می توان نشان داد که یک مدل مؤلفه که در دستنمایی دادهایی با توزیع گوسی و کواریانس قطری را بیشینه می کند، در واقع خطای میانگین مربعی داده های تولید شده را کمینه می کند.

۲-۱-۴- تحلیل مؤلفه مستقل

یک مسئله زیربنایی در تحقیقات روی شبکه عصبی و برخی زمینه های دیگر، یافتن یک بازنمایی مناسب برای داده های چند متغیره یا همان بردارهای تصادفی است. برای دلایلی مانند سادگی محاسباتی و مفهومی، بازنمایی ها معمولاً داده های اصلی را به صورت تبدیلی خطی نشان می دهند. به عبارت دیگر، هر مؤلفه بازنمایی ترکیبی خطی از داده های اصلی است. تبدیلات

خطی معروف شامل تحلیل مؤلفه اصلی، تحلیل فاکتور، و پیگیری افکنش است. تحلیل مؤلفه مستقل (ICA) روشی است که هدف آن پیدا کردن ترکیبی خطی از داده های غیرگوسی است بطوری که مؤلفه ها از هم کاملاً (و یا حتی الامکان) استقلال آماری داشته باشند. چنین بازنمایی به نظر می رسد که ساختار اصلی داده در بسیاری از کاربردها مانند استخراج ویژگی و تفکیک سیگنال داشته باشد.

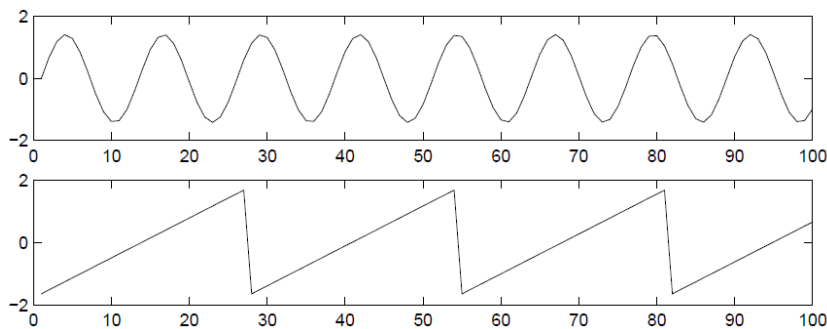
۲-۱-۴-۱- انگیزه

فرض کنید در یک اتاق هستید که دو نفر همزمان صحبت می کنند. شما دو میکروفون در اختیار دارید که در مکانهای متفاوتی از اتاق تعبیه شده اند. میکروفونها دو سیگنال زمانی ضبط شده می دهند که آنها را $x_1(t)$ و $x_2(t)$ می نامیم که x_1 و x_2 دامنه سیگنال و t اندیس زمان است. هر کدام از این سیگنالهای زمانی یک جمع وزندار از سیگنالهای گفتگویی است که از دو نفر، که با $s_1(t)$ و $s_2(t)$ می نامیم، منتشر می شود. می توان این جملات را با روابط زیر نشان داد:

$$x_1(t) = a_{11}s_1 + a_{12}s_2 \quad (۷-۲)$$

$$x_2(t) = a_{21}s_1 + a_{22}s_2$$

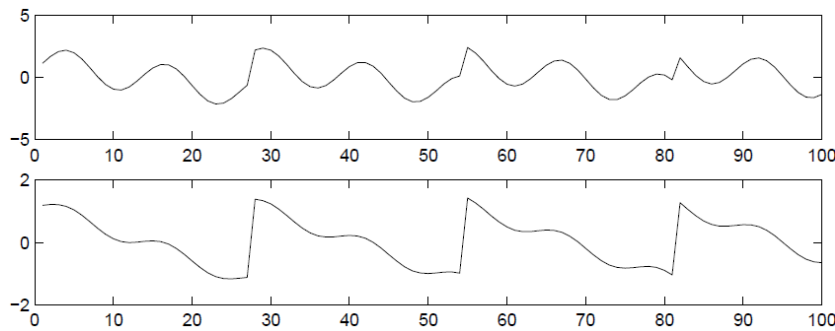
که در آن a_{11} ، a_{12} ، a_{21} و a_{22} پارامترهایی هستند که به فاصله میکروفون از گویندگان وابسته هستند. اگر بتوان تنها با استفاده از دو سیگنال ضبط شده $x_1(t)$ و $x_2(t)$ دو سیگنال اصلی گفتار $s_1(t)$ و $s_2(t)$ را تخمین زد، روش مورد استفاده، بسیار پرکاربرد خواهد بود. این مسئله به مسئله مهمانی کوکتیل^{۴۱} معروف است. در اینجا تاخیرهای زمانی و عوامل دخیل دیگر را در مدل تلفیقی ساده سازی شده خود نادیده می گیریم.



شکل ۳-۲: سیگنالهای اصلی

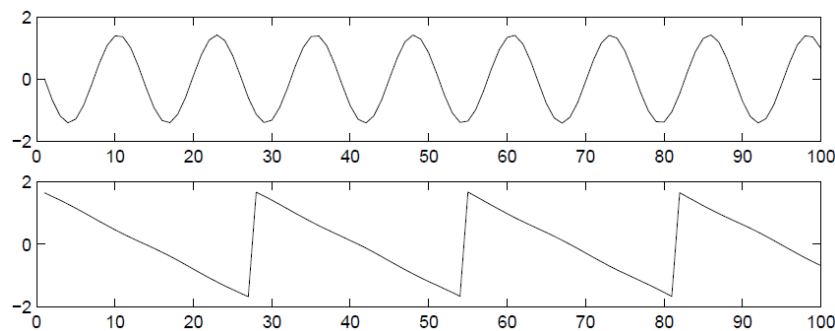
برای درک بهتر، شکل موجهای موجود در شکل ۳-۲ و شکل ۴-۲ را در نظر بگیرید. البته این سیگنالهای گفتار واقعی نیستند ولی برای روشن کردن موضوع کافی هستند. سیگنالهای گفتار اصلی ممکن چیزی مشابه موجهای شکل ۳-۲ باشند و سیگنالهای مخلوط ممکن است شبیه موجهای شکل ۴-۲ باشند. مسئله عبارتست از بازیابی داده های شکل ۳-۲ تنها با استفاده از داده های شکل ۴-۲. بدیهی است که اگر پارامترهای a_{ij} معلوم بود، رابطه ۷-۲ به راحتی از روشهای کلاسیک حل می شد. دقیقاً مشکل همینجاست که با معلوم نبودن a_{ij} حل مسئله بسیار دشوار خواهد بود.

⁴¹cocktail party problem



شکل ۲-۴: مخلوط مشاهده شده از سیگنالهای اصلی شکل ۲-۳

یک روش حل این مسئله استفاده از خواص آماری سیگنال $S_i(t)$ برای تخمین a_{ij} است. به صورت غیر منتظره ای تنها فرض استقلال آماری $S_1(t)$ و $S_2(t)$ در هر لحظه t برای حل مسئله کافیتست. این فرض در خیلی از موارد چندان هم غیر واقعی نیست و در عمل لزومی هم ندارد که کاملاً برقرار باشد. تکنیک ICA می تواند برای تخمین a_{ij} بر اساس اطلاعات استقلال آنها به کار رود که این امکان را می دهد که دو سیگنال گفتار اصلی تفکیک شود. شکل ۲-۵ دو سیگنال که توسط روش ICA بدست آمده اند را نشان می دهد. همانطور که مشخص است این سیگنالها بسیار به سیگنالهای اصلی منبع مشابهند (علامت آنها بر عکس است که تفاوت چندانی ندارد).

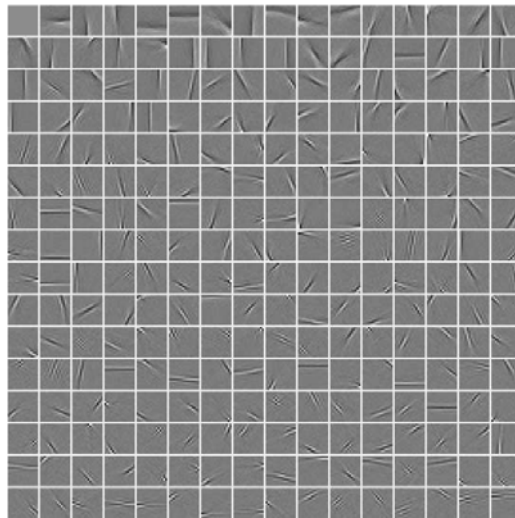


شکل ۲-۵: تخمینی از سیگنالهای اصلی که از روی سیگنالهای مخلوط مشاهده شده در شکل ۲-۴ بدست آمده است. سیگنالهای اصلی فارغ از علامت آنها بسیار دقیق تخمین زده شده اند.

تحلیل مؤلفه مستقل در اصل برای مسائلی طراحی شده است که بسیار شبیه مسئله مهمانی کوکتیل هستند. از کاربردهای گسترده آن می توان به تفکیک سیگنال EEG اشاره کرد. فعالیتهای مغزی که ضبط الکتریکی آنها توسط دستگاه الکتروانسفالوگرام انجام می شود، از ضبط پتانسیلهای الکتریکی بخشهای بسیار متفاوتی از مغز متشکل شده است که توسط الکترودها روی مجموعه جمع آوری می شوند. این پتانسیلهای الکتریکی احتمالاً از اختلاط چندین مورد فعالیت بخشهای متفاوت مغز تولید شده اند. این وضعیت بسیار مشابه مسئه مهمانی کوکتیل است: می خواهیم مؤلفه های اصلی فعالیت مغز را بیابیم ولی آنچه جمع آوری می شود تنها مخلوطی از این مؤلفه هاست. اطلاعات جالبی از اعمال روش ICA بر روی این سیگنالهای جمع آوری شده بدست می آید.

یک کاربرد کاملاً متفاوت ICA در انتخاب ویژگی است. یک مسئله زیربنایی در پردازش سیگنال رقمی پیدا کردن بازنمایی مناسب برای تصویر، صوت و سایر انواع داده برای کارهایی نظیر فشرده سازی و رفع نویز است. بازنمایی داده ها معمولاً بر مبنای تبدیلات خطی (گسسته) است. تبدیلات خطی استاندارد می شود که در پردازش تصویر استفاده می شوند عبارتند از فوریه، هار، تبدیل کسینوسی و غیره که هر کدام از خصوصیات ممتازی برخوردارند.

بهترین تبدیل خطی زمانی بدست می آید که تبدیل از خود داده ها تخمین زده شود، که در چنین حالتی تبدیل به طور ایده آلی به نوع داده ای که تحت پردازش است سازگار می شود. شکل ۲-۶ توابع پایه ای که از اعمال ICA بر روی تکه های تصویر طبیعی بدست می آیند را نشان می دهد. هر پنجره تصویر در مجموعه پنجره های تصویر قابل انطباق با یکی از این پنجره هاست بنابراین ضرایب این انطباق مستقلند.



شکل ۲-۶: توابع پایه ICA از تصاویر طبیعی. اندازه تکه تصاویر ورودی ۱۶×۱۶ پیکسل بوده است. توابع پایه را می توان ویژگیهای مستقل تصویر تلقی کرد.

تمامی کاربردهایی که در بالا به آن اشاره شد می توانند در چارچوب ریاضیاتی یکتای ICA فرموله شوند. این روش پردازش سیگنال و تحلیل داده ها، روشی همه منظوره است.

۲-۱-۴-۲- تعریف تحلیل مؤلفه مستقل

برای تعریف دقیق ICA می توان از مدل آماری «متغیرهای نهفته» استفاده کرد. فرض می کنیم که n مخلوط خطی x_1, \dots, x_n از n مؤلفه مستقل مشاهده کرده ایم (نیازی به اندیس زمانی t در مدل ICA نیست):

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n, \forall j \quad (۸-۲)$$

فرض می کنیم که هر مخلوط x_j و هر مؤلفه مستقل s_k یک متغیر تصادفی (و نه یک سیگنال وابسته به زمان) باشند. بنابراین مقادیر مشاهده شده $x_j(t)$ مثلاً سیگنالهای میکروفونها در مسئله مهمانی کوکتیل یک نمونه این متغیر تصادفی هستند. می توان بدون از دست دادن عمومیت فرض کرد که متغیرهای مخلوط و مؤلفه های مستقل دارای میانگین صفر

هستند: اگر این مسئله صادق نباشد متغیر مشاهده x_i را می توان با کم کردن میانگین از آن به مرکز برد که اینکار آنرا دارای میانگین صفر خواهد کرد.

برای راحتی بجای حاصلجمع هایی مانند رابطه قبلی از نشان گذاری ماتریسی استفاده می کنیم. لذا \bar{x} را برداری تصادفی با عناصر x_1, \dots, x_n و به طور مشابه \bar{s} را برداری تصادفی با عناصر s_1, \dots, s_n در نظر می گیریم. همچنین A را ماتریسی با درایه های a_{ij} فرض می کنیم. تمامی بردارها ستونی هستند و بالطبع ترانهاده آن \bar{x}^T برداری سطری است. با استفاده از این نشان گذاری مدل تلفیقی بالا اینگونه نوشته می شود:

$$\bar{x} = A\bar{s} \quad (9-2)$$

گاهی به ستونهای ماتریس A نیاز داریم و آنها را اینگونه \bar{a}_j نشان می دهیم. بنابراین مدل را می توان اینگونه نیز نوشت:

$$\bar{x} = \sum_{i=1}^n \bar{a}_i s_i \quad (10-2)$$

مدل آماری در رابطه ۲-۱۰ به مدل تحلیل مؤلفه مستقل یا مدل ICA معروف است. مدل ICA یک مدل مؤلفه است که بدین معنی است که این مدل چگونگی تولید داده های مشاهده شده از روی مؤلفه های s_i را توصیف می کند. مؤلفه های مستقل، متغیرهایی نهفته هستند، بدین معنی که نمی توان مستقیماً آنها را مشاهده کرد. همچنین ماتریس تلفیق مجهول فرض شده است. مشاهدات محدود به بردار تصادفی \bar{x} است و هم A و هم \bar{s} باید تخمین زده شود و اینکار تا جای ممکن باید کلی انجام شود.

نقطه شروع ICA فرض بسیار ساده ای مبنی بر اسقلال آماری s_i هاست. استقلال آماری به طور دقیق در ادامه همین بخش تعریف خواهد شد. همچنین در ادامه توجیه خواهد شد که مؤلفه های مستقل باید توزیعی غیر گوسی داشته باشند. با اینحال در این مدل ابتدایی فرضی نمی کنیم که این توزیعا را می دانیم (اگر این توزیعا معلوم باشند مسئله به طور چشمگیری ساده خواهد شد). برای سادگی همچنین فرض می کنیم که ماتریس مجهول تلفیق یک ماتریس مربعی باشد (که البته این فرض گاهی می تواند در نظر گرفته نشود). پس از تخمین ماتریس A می توان معکوس آنرا، که به آن W می گویند، بدست آورد و مؤلفه های مستقل را بسادگی از رابطه زیر حساب کرد:

$$\bar{s} = W\bar{x} \quad (11-2)$$

روش ICA ارتباط تنگاتنگی با روشی موسوم به تفکیک کورکورانه منابع^{۴۲} (BSS) یا تفکیک کورکورانه سیگنال دارد. یک منبع در اینجا همان سیگنال اصلی و در واقع مؤلفه مستقل است مانند یک گوینده در مسئله مهمانی کوکتیل. کورکورانه بدین معناست که درباره ماتریس تلفیق تقریباً هیچ دانشی در دسترس نیست و فرضیات اندکی روی سیگنالهای منبع صورت می گیرد. روش ICA یکی از روشهای تفکیک کورکورانه منابع و احتمالاً رایجترین آنهاست.

⁴²blind source separation

در بسیاری از کاربردها افزودن مقداری نویز به اندازه گیریها مسئله را واقعی تر جلوه می دهد. برای سادگی در این مدل ابتدایی جمله مربوط به نویز حذف شده است چرا که تخمین مدل بدون نویز، خود به اندازه کافی دشوار هست و برای بسیاری از کاربردها هم کافی می نماید.

۲-۱-۴-۳- ابهامات تحلیل مؤلفه مستقل

مدل ICA در رابطه ۲-۱۰ برای نشان دادن این ابهامات کافی است:

۱- نمی توان واریانس (انرژی) مؤلفه های مستقل را تعیین کرد. علت این امر اینست که هر دوی A و \bar{A} مجهولند و ضریب اسکالر هر یک از منابع s_i همیشه می تواند با تقسیم شدن ستون متناظر \bar{a}_j از A بر همان اسکالر خنثی شود، رابطه ۲-۱۰ مشاهده گردد. متعاقباً می توان بزرگی مؤلفه های مستقل را ثابت نگه داشت زیرا متغیرهای تصادفی هستند و یک راه طبیعی برای اینکار اینست که فرض کنیم واریانس هر کدام برابر واحد است ($E\{s_i^2\} = 1$). آنگاه در ICA ماتریس A طوری سازگار می شود که این محدودیت را جبران کند. باید توجه داشت که همچنان ابهام در زمینه علامت را باقی می گذارد: می توان بدون تاثیر خاصی در مدل مؤلفه مستقل را در ۱- ضرب کرد. خوشبختانه این محدودیت در اغلب کاربردها چندان مهم نیست.

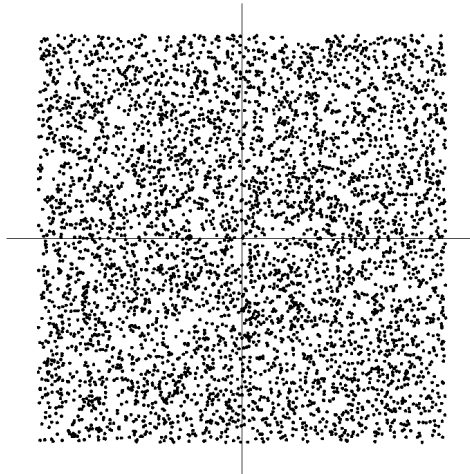
۲- ترتیب مؤلفه های مستقل مشخص نیست. علت آن بازهم به مجهول بودن هر دوی A و \bar{A} باز می گردد. ترتیب جملات حاصلجمع در رابطه ۲-۱۱ دلخواه است و هر مؤلفه مستقلی می تواند اول باشد. به طور رسمی یک ماتریس جایگشت P و معکوس آن اینگونه در مدل گنجانده می شود $\bar{x} = AP^{-1}P\bar{s}$. عناصر $P\bar{s}$ مؤلفه های مستقل اصلی s_i هستند، البته در یک ترتیب دیگر. ماتریس AP^{-1} نیز یک ماتریس مجهول جدید است که الگوریتم ICA باید آنرا حل کند.

برای روشن کردن مدل ICA با جملات آماری، دو مؤلفه مستقل که توزیع یکنواخت زیر را دارا هستند در نظر

بگیرید:

$$p(s_i) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{if } |s_i| \leq \sqrt{3} \\ 0 & \text{otherwise} \end{cases} \quad (2-12)$$

بازه مقادیر در این توزیع طوری انتخاب شده است که میانگین آن صفر و واریانس واحدی داشته باشد. چگالی توأم s_1 و s_2 بنابراین روی یک مربع یکنواخت خواهد بود. این امر از تعریف ابتدایی چگالی توأم دو متغیر مستقل ناشی می شود که طبق رابطه ۲-۱۵ برابر حاصلضرب چگالی جانبی آنهاست. این چگالی توأم در شکل ۲-۷ با نشان دادن نقاط داده ای که بطور اتفاقی از توزیع آورده شده اند، نشان داده شده است.

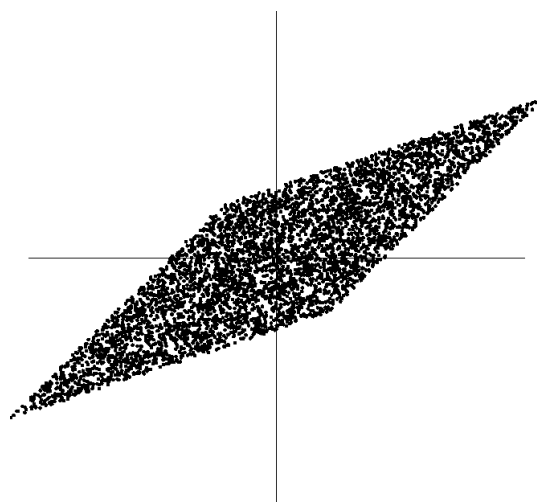


شکل ۲-۷: توزیع توام مؤلفه های S_1 (محور افقی) و S_2 (محور عمودی) با توزیع یکنواخت

حال این دو مؤلفه مستقل را مخلوط می کنیم. فرض می کنیم این ماتریس تلفیق استفاده شود:

$$A_0 = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \quad (۲-۱۳)$$

این رابطه دو متغیر مخلوط x_1 و x_2 دارد. به راحتی می توان دید که داده های مخلوط نیز روی متوازی الاضلاع شکل ۲-۸ توزیع یکنواخت دارند. باید توجه داشت که متغیرهای تصادفی x_1 و x_2 دیگر مستقل نیستند. یک راه ساده برای نشان دادن این امر اینست که ببینیم آیا می توان مقدار یکی از آنها، مثلاً x_2 را از روی مقدار دیگری بدست آورد. واضح است که اگر x_1 به یکی از مقادیر کمینه یا بیشینه خود برود، مقدار متغیر دیگر کاملاً مشخص است و لذا ایندو مستقل نیستند. برای متغیرهای S_1 و S_2 شرایط متفاوت است. از شکل ۲-۶ می توان دید دانستن مقدار S_1 اصلاً کمکی به حدس زدن مقدار S_2 نمی کند.



شکل ۲-۸: توزیع توام مخلوطهای مشاهده شده x_1 (محور افقی) و x_2 (محور عمودی)

حال مسئله تخمین مدل داده های ICA به تخمین ماتریس تلفیق A_0 تنها با استفاده از دانش موجود در مخلوطهای x_1 و x_2 تبدیل شده است. با توجه به شکل ۲-۸ می توان راهی برای تخمین A دریافت: ضلع متوازی الاضلاع در راستاهای ستونهای A قرار دارد. این بدان معناست که می توان مدل ICA را ابتدا با تخمین چگالی توأم x_1 و x_2 و سپس یافتن لبه های آن تخمین زد. بنابراین این مسئله یک راه حل دارد. البته این راه حل بسیار ضعیفی است زیرا تنها برای متغیرهایی کار می کند که توزیع یکنواخت دارند و علاوه بر آن پیچیدگی محاسباتی بسیار زیادی دارد. باید روشی پیدا کرد که برای هر توزیع مؤلفه های مستقل با سرعت و مطمئن کار کند.

۲-۴-۱-۴- تعریف و خصوصیات زیربنایی استقلال

برای تعریف مفهوم استقلال، دو متغیر تصادفی y_1 و y_2 را در نظر بگیرید. به زبان ساده، دو متغیر y_1 و y_2 را مستقل می گویند اگر اطلاعاتی درباره مقدار y_1 هیچ اطلاعاتی درباره مقدار y_2 ندهد و بالعکس. در بالا هم همانطور که اشاره شد این شرط برای S_1 و S_2 برقرار است ولی برای x_1 و x_2 برقرار نیست.

از دید ریاضی، استقلال توسط چگالی احتمال تعریف می شود. فرض می کنیم $p(y_1, y_2)$ تابع توزیع احتمال (PDF) توأم y_1 و y_2 باشد و $p_1(y_1)$ تابع توزیع احتمال حاشیه ای y_1 باشد (یعنی زمانی که y_2 تنها در نظر گرفته شود):

$$p_1(y_1) = \int p(y_1, y_2) dy_2 \quad (۲-۱۴)$$

و بطور مشابه برای y_2 . آنگاه x_1 و x_2 را مستقل می نامیم اگر و تنها اگر چگالی احتمال توأم آنها به صورت زیر قابل فاکتورگیری باشد:

$$p(y_1, y_2) = p_1(y_1)p_2(y_2) \quad (۲-۱۵)$$

این تعریف به سادگی به n متغیر تصادفی نیز تعمیم داده می شود که در آن تابع چگالی باید حاصلضرب n جمله باشد. این تعریف می تواند برای بدست آوردن یکی از مهمترین خصوصیات متغیرهای تصادفی استفاده شود. با داشتن دو تابع h_1 و h_2 همواره داریم:

$$E\{h_1(y_1)h_2(y_2)\} = E\{h_1(y_1)\}E\{h_2(y_2)\} \quad (۲-۱۶)$$

۲-۴-۱-۵- متغیرهای ناهمبسته جزئی مستقلند

فرم ضعیفتری از استقلال، ناهمبستگی است. دو متغیر تصادفی y_1 و y_2 ناهمبسته گفته می شوند اگر کواریانس آنها صفر باشد:

$$E\{y_1 y_2\} - E\{y_1\}E\{y_2\} = 0 \quad (۲-۱۷)$$

اگر متغیرها مستقل باشند، ناهمبسته هم هستند که در رابطه ۱۶-۲ اگر $h(y_1)=y_1$ و $h(y_2)=y_2$ کاملاً دلیل آن مشخص می شود. از سوی دیگر ناهمبستگی الزاماً به معنای استقلال نیست. به عنوان مثال فرض کنید (y_1, y_2) دارای مقادیر گسسته ای است و از توزیع زیر پیروی می کنند که هر زوج آن، $(0, 1)$ و $(0, -1)$ و $(1, 0)$ و $(-1, 0)$ ، دارای احتمال 0.25 برابر است. طبق رابطه y_1 و y_2 ناهمبسته هستند. از سوی دیگر $E\{y_1^2\}E\{y_2^2\} = 0 \neq 0.25 = E\{y_1^2 y_2^2\}$ است، بنابراین شرایط رابطه ۱۶-۲ برقرار نیست و نمی توان آنها را مستقل دانست.

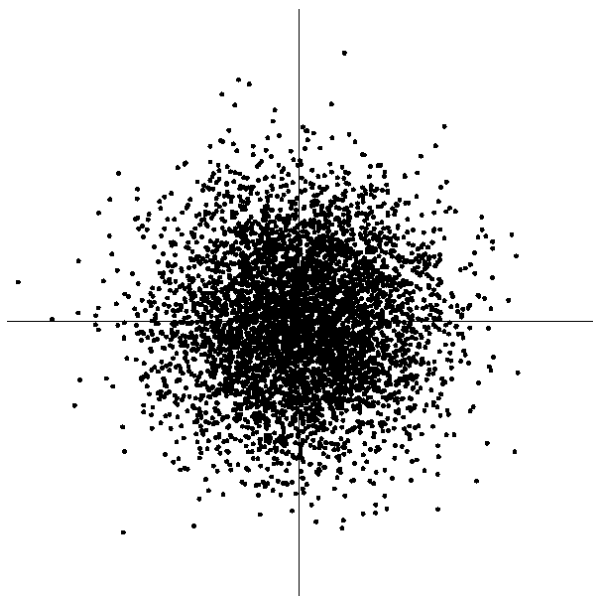
از آنجایی که استقلال به طور ضمنی ناهمبستگی را در بر دارد، بسیاری از روشهای ICA فرایند تخمین خود را طوری تحدید می کنند که در هر مرحله تخمین مؤلفه های مستقل ناهمبسته باشند. اینکار تعداد پارامترهای آزاد را کاهش می دهد و مسئله را ساده سازی می کند.

۲-۱-۴-۶- چرا متغیرهای گوسی ممنوع هستند؟

محدودیت اساسی روش ICA اینست که مؤلفه های مستقل ICA باید غیرگوسی باشند تا روش کار بکند. برای آنکه روشن شود که چرا متغیرهای گوسی حل شدن ICA را غیرممکن می کنند، فرض می کنیم که ماتریس تلفیق متعامد است و s_i گوسی است. پس x_1 و x_2 نیز گوسی، ناهمبسته، و دارای واریانس واحد هستند. چگالی توأم آنها با رابطه زیر بدست می آید:

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \quad (2-18)$$

این توزیع در شکل ۲-۹ به تصویر کشیده شده است. این تصویر نشان می دهد که تابع چگالی کاملاً متقارن است. بنابراین حاوی هیچ اطلاعاتی در مورد راستای ستونهای ماتریس تلفیق A نیست و به همین دلیل نمی توان A را تخمین زد.



شکل ۲-۹: توزیع چند متغیره دو متغیر مستقل گوسی

به عبارت دقیقتر، می توان اثبات کرد که توزیع هر تبدیل متعامدی از متغیرهای گوسی (x_1, x_2) دارای توزیعی کاملاً مشابه (x_1, x_2) هستند و x_1 و x_2 مستقلند. بنابراین در حالتی که متغیرها گوسی باشند مدل ICA را تنها می توان تا تبدیل متعامد تخمین زد. یعنی برای مؤلفه های مستقل گوسی، ماتریس A قابل تشخیص نیست. البته اگر فقط یکی از مؤلفه ها گوسی باشد باز هم می توان مدل ICA را تخمین زد.

۲-۱-۴-۷- مؤلفه های غیر گوسی مستقلند

کلید تخمین زدن مدل ICA غیر گوسی بودن است. در واقع همانطور که اشاره شد، بدون غیر گوسی بودن تخمین اصلاً ممکن نیست. این امر باعث اختلال در پیشرفت سریع تحقیقات روی ICA نیز شده است چون در بسیاری از نظریه های کلاسیک آمار، فرض می شود که متغیرهای تصادفی دارای توزیع گوسی هستند.

قضیه حد مرکزی، یک نتیجه کلاسیک از نظریه احتمالات، بیان می دارد که توزیع جمع متغیرهای تصادفی مستقل تحت شرایط خاصی به سمت توزیع گوسی میل می کند. بنابراین جمع دو متغیر تصادفی مستقل معمولاً نسبت به دو تک تک این متغیرها، دارای توزیعی نزدیکتر به گوسی می باشد.

اکنون فرض می کنیم که بردار داده های \vec{x} مطابق داده های ICA در رابطه ۲-۹ توزیع شده است، بدان معنی که مخلوطی از دو متغیر مستقل است. برای سادگی در این زیربخش فرض می کنیم که تمام مؤلفه های مستقل دارای توزیع مشابهی هستند. برای تخمین یکی از مؤلفه های مستقل ترکیبی خطی از x_i را فرض می کنیم (رابطه ۲-۱۱) و آنرا با $y = \vec{w}^T \vec{x} = \sum_i w_i x_i$ نشان می دهیم که در آن برداری است که باید تعیین شود. اگر \vec{w} یکی از سطرهاى معکوس A باشد، این ترکیب خطی در واقع برابر یکی از مؤلفه های مستقل هستند. اکنون پرسش اینست که چگونه از قضیه حد مرکزی برای تعیین \vec{w} استفاده شود که برابر یکی از سطرهاى معکوس A شود؟ در عمل، نمی توان چنین \vec{w} ای را دقیقاً پیدا کرد چون هیچ دانشی درباره ماتریس A موجود نیست ولی می توان تخمین زن خوبی را پیدا کرد که تقریب خوبی بدست بدهد.

برای اینکه مشخص باشد که چگونه این به اصل ابتدایی تخمین ICA منجر می شود، تغییر کوچکی در متغیرها می دهیم و $\vec{z} = A^T \vec{w}$ را اینگونه تعریف می کنیم. در نتیجه داریم $\vec{y} = \vec{w}^T \vec{x} = \vec{w}^T A \vec{s} = \vec{z}^T \vec{s}$. بنابراین y ترکیبی خطی از s_i با وزنهایی که z_i مشخص می کند است. از آنجایی که حاصل جمع دو متغیر تصادفی مستقل از متغیرهای اصلی گوسی تر است، $\vec{z}^T \vec{s}$ از هر کدام از s_i ها گوسی تر است و زمانی کمترین مقدار گوسی بودن را دارد که برابر یکی از s_i ها باشد. در این حالت بدیهی است که یکی از z_i های \vec{z} غیر صفر است (و فرض اینست که s_i ها توزیع یکسانی داشته باشند).

بنابراین \vec{w} برداری در نظر گرفته می شود که غیر گوسی بودن $\vec{w}^T \vec{x}$ را بیشینه کند. چنین برداری (در دستگاه مختصاتی تبدیل شده) الزاماً متناظر با یک \vec{z} است که تنها یک مؤلفه غیر صفر دارد. این بدین معناست که $\vec{w}^T \vec{x} = \vec{z}^T \vec{s}$ برابر یکی از مؤلفه های مستقل است!

بنابراین بیشینه کردن غیرگوسی بودن $\vec{w}^T \vec{x}$ یکی از مؤلفه های مستقل را به ما می دهد و در واقع نمودار بهینه سازی غیرگوسی بودن فضای n بُعدی \vec{w} دارای $2n$ بیشینه محلی است، دو تا به ازای هر مؤلفه مستقل که با s_i و $-s_i$ متناظر است (باید به یاد داشت که مؤلفه های مستقل دارای علامت خاصی نبودند) برای یافتن چند مؤلفه مستقل، باید تمام این بیشینه های محلی پیدا شوند. اینکار دشوار نیست زیرا مؤلفه های مستقل مختلف به هم ناهمبسته اند و همواره می توان فضا را به دنبال تخمینی که با تخمینهای قبلی ناهمبسته است جستجو کرد. این کار متناظر با متعامد کردن^{۴۳} در فضایی است که به طور مناسبی تبدیل یافته است (تبدیل وایتینگ^{۴۴}). باید توجه داشت که این روش مطرح شده یک هیوریستیک نیست و اثبات آن موجود است و در ادامه خواهد آمد.

۲-۱-۴-۸- معیارهای غیرگوسی بودن

برای استفاده از غیرگوسی بودن در تخمین ICA، باید معیاری کمی از غیرگوسی بودن متغیری تصادفی مانند γ داشته باشیم. برای سادگی فرض می کنیم که γ دارای میانگین صفر است و واریانس آن واحد است. یکی از کاربردهای پیش پردازشهایی نظیر به مرکز آوردن، وایتینگ، و تبدیلات روی سربهای زمانی برای تامین پیش فرضهای این ساده سازیهاست.

کلاسیکترین معیار غیرگوسی بودن کورتوسیس یا گشتاور مرتبه چهارم است. معمولاً کورتوسیس از رابطه زیر محاسبه می شود:

$$\tilde{\kappa}(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (۱۹-۲)$$

از آنجایی که فرض می کنیم که γ واریانس واحد دارد، سمت راست این رابطه به شکل $3 - E\{y^4\}$ ساده می شود. این نشان می دهد که کورتوسیس یک نسخه نرمال شده از مرتبه چهارم $E\{y^4\}$ است. برای متغیر گوسی γ ، گشتاور مرتبه چهارم برابر $3(E\{y^2\})^2$ است. بنابراین مقدار کورتوسیس برای متغیرهای تصادفی گوسی برابر صفر است. برای اغلب متغیرهای تصادفی غیرگوسی (و نه تمام آنها) مقدار کورتوسیس غیرصفر است. مقدار کورتوسیس می تواند مثبت و یا منفی باشد. متغیرهای تصادفی که مقدار کورتوسیس منفی دارند زیرگوسی^{۴۵} نامیده می شوند و آنها که کورتوسیس مثبت دارند زیرگوسی^{۴۶} نامیده می شود. در ادبیات علم آمار، به ترتیب لغات پلّتی کورتیک^{۴۷} و لپتوکورتیک^{۴۸} نیز استفاده می شود. متغیرهای تصادفی زیرگوسی معمولاً دارای تابع توزیع احتمالاتی نوک تیز^{۴۹} و دنباله های سنگین^{۵۰} هستند، یعنی این تابع توزیع در نزدیکیهای صفر بسیار بزرگند و در مقادیر بزرگ متغیر نیز مقدار زیادی دارند ولی در مقادیر میانی کوچک هستند. یک مثال رایج برای آن توزیع لاپلاس است که تابع توزیع آن (وقتی که واریانس آن به یک نرمالیزه شود) به رابطه زیر بیان می شود:

⁴³ orthogonalization

⁴⁴ whitening transformation

⁴⁵ subgaussian

⁴⁶ supergaussian

⁴⁷ platykurtic

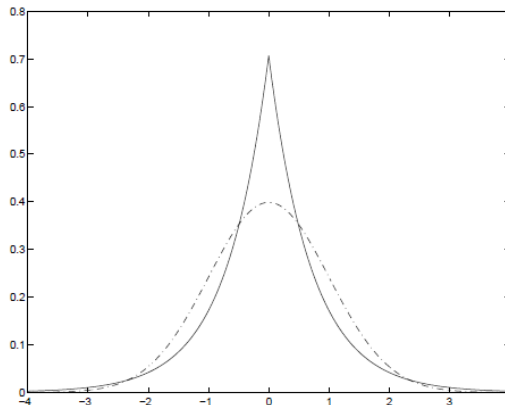
⁴⁸ leptokurtic

⁴⁹ spiky, peaky

⁵⁰ heavy tails

$$p(y) = \frac{1}{\sqrt{2}} \exp(\sqrt{2}|y|) \quad (20-2)$$

این تابع توزیع احتمال در شکل ۲-۱۰ نشان داده شده است. از سوی دیگر، متغیرهای تصادفی زیرگوسی معمولاً دارای توزیع مسطحی هستند که بیشتر در اطراف صفر مقدار ثابتی دارد و به ازای مقادیر بزرگتر متغیر، مقادیر نسبتاً کوچکی دارند. مثالی از آنها توزیع یکنواخت تابع ۲-۱۲ است.



شکل ۲-۱۰: تابع توزیع توزیع لاپلاسی که یک توزیع زیرگوسی رایج است و برای مقایسه توزیع گوسی نیز به طور خط چین در آن رسم شده است.

معمولاً غیرگوسی بودن با مقدار مطلق کورتوسیس اندازه گیری می شود. مربع این کورتوسیس نیز می تواند استفاده شود. این مقدار برای متغیر گوسی صفر است و برای بسیاری از متغیرهای تصادفی غیرگوسی بزرگتر از صفر است. با اینکه تعداد معدودی از متغیرهای تصادفی غیرگوسی دارای کورتوسیس صفر هستند ولی می توان از آنها چشم پوشی کرد.

کورتوسیس یا مقدار مطلق آن به طور گسترده ای در اندازه گیری غیرگوسی بودن در ICA و سایر روشهای مرتبط استفاده می شود. علت اصلی آن سادگی، چه از لحاظ عملی و چه از نظر محاسباتی. از لحاظ محاسباتی کورتوسیس را می توان به استفاده از گشتاور مرتبه چهار داده های نمونه تخمین زد. تحلیل نظری نیز به دلیل خاصیت خطی زیر ساده سازی می شود: اگر x_1 و x_2 دو متغیر تصادفی مستقل باشند، روابط زیر برقرار است:

$$\tilde{\kappa}(x_1 + x_2) = \tilde{\kappa}(x_1) + \tilde{\kappa}(x_2) \quad (21-2)$$

و

$$\tilde{\kappa}(\alpha x_1) = \alpha^4 \tilde{\kappa}(x_1) \quad (22-2)$$

که در آن α کمیتی اسکالر است. این خاصیتها با استفاده از تعریف به راحتی اثبات می شوند.

برای اینکه بتوان نشان داد که نمودار بهینه سازی کورتوسیس چگونه خواهد بود، و چگونه مؤلفه های مستقل می توانند با کمینه سازی و یا بیشینه سازی کورتوسیس بدست آیند، باید به یک مدل دوبعدی $\bar{x} = A\bar{s}$ توجه کرد. فرض کنیم

که مؤلفه های مستقل s_1 و s_2 دارای مقادیر کورتوسیس $\tilde{K}(s_1)$ و $\tilde{K}(s_2)$ هستند که هر دو غیرصفرند. باید به یاد داشته باشید که واریانس هر دو واحد فرض شده اند. اکنون بدنبال مؤلفه های مستقلی مانند $y = \bar{W}^T \bar{x}$ می گردیم.

یکبار دیگر از $\bar{z} = A^T \bar{w}$ بهره گرفته می شود. خواهیم داشت $\bar{z}^T \bar{s} = \bar{z}^T A \bar{s} = \bar{w}^T \bar{x} = y$. حال بر اساس خصوصیات کورتوسیس داریم $\tilde{K}(y) = \tilde{K}(z_1 s_1) + \tilde{K}(z_2 s_2) = z_1^4 \tilde{K}(s_1) + z_2^4 \tilde{K}(s_2)$. از سوی دیگر شرط همانطور که بر s_1 و s_2 اثر می کند، این را نیز دیکته می کند که واریانس y برابر ۱ باشد. این امر به طور ضمنی به این قید روی \bar{z} منتهی می شود که: $E\{y^2\} = z_1^2 + z_2^2 = I$ از نظر هندسی این بدین معناست که بردار \bar{z} به دایره ای با شعاع واحد روی صفحه ای دوبعدی محدود است. حال مسئله بهینه سازی عبارتست از یافتن بیشینه های تابع $|\tilde{K}(y)| = |z_1^4 \tilde{K}(s_1) + z_2^4 \tilde{K}(s_2)|$ روی دایره واحد. برای سادگی می توان فرض کرد که کورتوسیس ها علامت مشابهی داشته باشند که در این صورت می توان از علامت قدرمطلق صرفنظر کرد. رسم این تابع همان دورنمای بهینه سازی^{۵۱} برای مسئله است.

نشان داده شده است که بهینه ها نقاطی هستند که در آنها دقیقاً یکی از عناصر بردار \bar{z} برابر صفر و مابقی غیرصفر باشند. به دلیل مقید بودن به دایره واحد، عناصر غیرصفر باید ۱ یا -۱ باشند. ولی این نقاط دقیقاً آنها نیست که y برابر یکی از مؤلفه های مستقل $\pm s_i$ است، و مسئله حل می شود.

در عمل ما از یک بردار وزن \bar{w} شروع می کنیم، راستایی را که در آن $y = \bar{w}^T \bar{x}$ قویترین رشد را می کند (اگر کورتوسیس مثبت باشد) و یا قویترین کاهش را دارد (اگر کورتوسیس منفی باشد) را بر اساس نمونه های موجود $\bar{x}(1), \dots, \bar{x}(T)$ از بردار مخلوط \bar{x} محاسبه می کنیم، و از یک روش شیب نوردی یا توسعه ای از آن برای پیدا کردن بردار جدید \bar{w} استفاده می کنیم. مثالی که زده شد می تواند به هر چند بُعدی که دلخواه است تعمیم داده شود و نشان داده شود که کورتوسیس می تواند به عنوان یک معیار برای بهینه سازی مسئله ICA استفاده شود. بهر حال کورتوسیس در عمل وقتی که مقدار آن باید از روی نمونه ها تخمین زده شود، دارای ضعفهایی نیز هست. مشکل اصلی حساسیت زیاد این معیار به اندازه گیریهای دور از مرکز^{۵۲} است چون مقدار آن ممکن است بر اساس تنها چند اندازه گیری از دنباله توزیع تخمین زده شود که می توانند مشاهداتی دارای خطا و یا بی ربط باشند. به عبارت دیگر کورتوسیس معیار مقاومی برای غیرگوسی بودن نیست. بنابراین گاهی اوقات معیارهای دیگری برای محاسبه غیرگوسی بودن مناسبترند. در ادامه معیار نگانتروپی^{۵۳} معرفی می شود که خواصی تقریباً متضاد با خواص کورتوسیس دارد و راههایی برای تخمین آن ارائه می شود.

دومین معیار مهم غیرگوسی بودن نگانتروپی است. این معیار بر اساس کمیت آنتروپی (تفاضلی) از علم نظریه اطلاعات بوجود آمده است. در واقع آنتروپی اصل ابتدایی نظریه اطلاعات است. آنتروپی یک متغیر تصادفی را می توان به عنوان درجه اطلاعاتی که آن مشاهده بدست می دهد دانست. هر چقدر متغیر تصادفی تر، یعنی غیرقابل پیش بینی تر و بی

⁵¹ optimization landscape

⁵² outliers

⁵³ negentropy

ساختارتر باشد، مقدار آنتروپی آن بالاتر است. آنتروپی قرابت زیادی با طول کد متغیر تصادفی در یک کدگذاری بهینه دارد. برای متغیر تصادفی Y مقدار آنتروپی H به صورت

$$H(Y) = -\sum_i P(Y = a_i) \log P(Y = a_i) \quad (23-2)$$

تعریف می شود که در آن a_i تمام مقادیر ممکن Y است. این تعریف بسیار معروف آنتروپی می تواند برای متغیرها و بردارهای تصادفی با مقادیر پیوسته نیز تعمیم یابد که در این حالت آنتروپی تفاضلی نام می گیرد. آنتروپی تفاضلی H یک متغیر تصادفی مانند \bar{y} با توزیع $f(\bar{y})$ به صورت زیر تعریف می شود:

$$H(\bar{y}) = -\int f(\bar{y}) \log f(\bar{y}) d\bar{y} \quad (24-2)$$

یک نتیجه زیربنایی از نظریه اطلاعات اینست که متغیرهای تفاضلی بین تمام متغیرهای تصادفی با واریانس یکسان، دارای بالاترین آنتروپی هستند. پس می توان از آنتروپی به عنوان معیار غیرگوسی بودن استفاده کرد. در واقع این دستاورد علمی نشان می دهد که توزیع تصادفی در بین تمام توزیعها تصادفی ترین و یا بدون ساختارترین توزیع است. آنتروپی برای توزیعهایی که در مقادیر خاصی متمرکزند اندک است و چنین توزیعهایی یا کاملاً خوشه بندی شده اند یا تابع توزیع تجمعی نوک تیزی دارند. برای بدست آوردن یک معیار غیرگوسی بودن که برای متغیرهای تصادفی صفر و برای دیگران مثبت باشد، می توان از یک نسخه کمی تغییر کرده آنتروپی تفاضلی به نام نگانترویی استفاده کرد. نگانترویی J به صورت زیر تعریف می شود:

$$J(\bar{y}) = H(\bar{y}_{gauss}) - H(\bar{y}) \quad (25-2)$$

که در آن \bar{y}_{gauss} یک متغیر تصادفی گوسی با کواریانس برابر با کواریانس \bar{y} است. به دلیل خواصی که اشاره شد، نگانترویی همیشه مثبت است و زمانی صفر است که \bar{y} دارای توزیع گوسی باشد. نگانترویی ویژگی جالب دیگری نیز دارد که در مقابل تبدیلات خطی قابل بازگشت⁵⁴، نامتغیر باقی می ماند.

مزیت استفاده از نگانترویی به عنوان معیار غیرگوسی بودن اینست که با نظریه آمار به خوبی توجیه می شوند. در حقیقت نگانترویی از برخی لحاظ، تا جایی که به ویژگیهای آماری مربوط است، بهترین تخمین زن غیرگوسی بودن است. مشکل استفاده از نگانترویی اینست که محاسبات آن بسیار دشوار است. تخمین زدن نگانترویی با استفاده از تعریف نیاز به تخمینی (احتمالاً بدون پارامتر) از چگالی دارد.

تخمین زدن نگانترویی چنان که اشاره شد، دشوار است و در عمل باید از برخی از تقریبها استفاده کرد. در اینجا تنها به تقریبهایی بسنده می شود که خاصیتهای مناسبی دارند و در تولید یک روش کارا برای ICA استفاده می شوند. روشهای کلاسیک برای محاسبه نگانترویی استفاده از گشتاورهای مرتبه بالاست که نمونه ای از آن آورده شده است:

⁵⁴invertible

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} \tilde{\kappa}(y)^2 \quad (26-2)$$

متغیر تصادفی \tilde{y} فرض می شود که دارای میانگین صفر و واریانس واحد است. با اینحال اعتبار چنین تقریبهایی ممکن است محدود باشد و این روش در قیاس با کورتوسیس از پایداری کمتری برخوردار باشد. برای اجتناب از مشکلاتی که این تقریبهای نگانترویی ممکن است ایجاد کند، تقریبهای دیگری نیز تهیه شده است که بر مبنای اصل آنتروپی بیشینه استوار شده است. به طور کلی این تقریب بدست می آید:

$$J(y) \approx \sum_{i=1}^p k_i [E\{G_i(y)\} - E\{G_i(v)\}]^2 \quad (27-2)$$

که در آن k_i یک ضریب مثبت است و v یک متغیر گوسی با میانگین صفر و واریانس واحد (گوسی استاندارد) فرض شده است و G_i یک تابع غیر درجه دوم در نظر گرفته می شود. توجه شود که حتی در چنین شرایطی هم تقریب خیلی دقیق نیست ولی می تواند معیاری برای غیر گوسی بودن را فراهم کند که از لحاظ غیر منفی بودن و برابری با صفر برای توزیعهای گوسی نظر ریاضیدانان را تأمین کند. در این حالت با داشتن تنها یک تابع غیر درجه دوم G تقریب برابر خواهد بود با:

$$J(y) \propto [E\{G_i(y)\} - E\{G_i(v)\}]^2 \quad (28-2)$$

که برای تقریباً همه توابع درجه دوم G برقرار است. این رابطه به وضوح یک تقریب مبتنی بر گشتاور رابطه ۲۷-۲ است اگر y متقارن باشد. در اصل با در نظر گرفتن $G(y)=y^4$ این رابطه مبتنی بر کورتوسیس دقیقاً بدست می آید.

البته با انتخاب هوشمندانه G در اینجا می توان تقریبهایی بسیار بهتر از ۲۷-۲ برای نگانترویی بدست آورد. با انتخاب تابعی که خیلی سریع رشد نکند می توان تخمین زندهای مقاومتری را بدست آورد. مثلاً انتخابهای زیر برای G بسیار سودمند هستند:

$$G_1(u) = \frac{1}{a_1} \log(\cosh(a_1 u)) \quad (29-2)$$

$$G_2(u) = -\exp\left(-\frac{u^2}{2}\right)$$

که $1 \leq a_1 \leq 2$ انتخاب مناسبی است. بنابراین تقریبهایی از نگانترویی بدست آمد که مصالحه خوبی را بین خصوصیات دو معیار کلاسیک غیرگوسی بودن یعنی کورتوسیس و نگانترویی برقرار می کنند. این رابطه ساده، از لحاظ محاسباتی سریع، و در عین حال دارای ویژگیهای ممتاز آماری بخصوص مقاوم بودن هستند.

۲-۱-۴-۹- کمینه کردن اطلاعات متقابل

روش دیگری برای تخمین ICA از نظریه اطلاعات الهام گرفته می شود و کمینه کردن اطلاعات متقابل است. این روش نیز مانند روش قبل به همان اصل یافتن جهت‌های دارای بیشترین مقدار غیرگوسی بودن ختم می شود.

با بهره گیری از مفهوم آنتروپی تفاضلی، می توان اطلاعات متقابل I بین m متغیر تصادفی اسکالر y_i که $i=1, \dots, m$ است را اینگونه تعریف کرد:

$$I(y_1, \dots, y_m) = \sum_{i=1}^m H(y_i) - H(\bar{y}) \quad (30-2)$$

اطلاعات متقابل یک معیار طبیعی برای تعیین میزان وابستگی بین متغیرهای تصادفی است و در واقع معادل افتراق کولبک-لیبلر⁵⁵ بین توزیع توأم $f(\bar{y})$ و حاصلضرب توزیعهای حاشیه ای است که معیاری بدیهی برای محاسبه استقلال آماری می باشد. این معیار همیشه غیر منفی است و تنها در حالتی صفر است که متغیرها مستقل آماری باشند. بنابراین اطلاعات متقابل تمام ساختار وابستگی را در نظر می گیرد و برخلاف PCA و روشهای مشابه تنها به کواریانس بسنده نمی کند. اطلاعات متقابل را با توجه به تفسیر آنتروپی می توان طول کد تفسیر کرد. عبارت $H(y_i)$ طول کد را هنگامی که y_i ها جدا جدا کد می شوند نشان می دهد و $H(\bar{y})$ طول کد را وقتی که \bar{y} به صورت یک بردار تصادفی کد می شود (یعنی تمام مؤلفه ها توسط یک کد، کدگذاری شوند) بدست می دهد. بنابراین اطلاعات متقابل نشان می دهد که چقدر کاهش طول کد در اثر کد کردن تمام بردار بجای کد کردن مجزای مؤلفه ها بدست می آید. کلاً با استفاده از کدگذاری تمام بردار کدهای بهتری بدست می آید. با اینحال اگر y_i ها مستقل باشند هیچ اطلاعاتی از یکدیگر بدست نمی دهند و کد کردن جداجدای مؤلفه ها از نظر طول فرقی با کد کردن یکجای آنها ندارد. یک خاصیت مهم اطلاعات متقابل اینست که برای هر تبدیل بازگشت پذیر $\bar{y} = W\bar{x}$ داریم:

$$I(y_1, \dots, y_n) = \sum_i H(y_i) - H(\bar{x}) - \log |\det W| \quad (31-2)$$

حال اگر y_i ها محدود شوند به طوری که باید ناهمبسته و دارای کواریانس واحد باشند، $E\{yy^T\} = WE\{xx^T\}W^T = I$ است که به طور ضمنی اشاره دارد که:

$$\det I = 1 = (\det W \cdot E\{\bar{x}\bar{x}^T\} \cdot W^T) = (\det W)(\det E\{\bar{x}\bar{x}^T\})(\det W^T) \quad (32-2)$$

و این نکته بدین معناست که $\det W$ باید ثابت باشد. افزون بر این برای y_i که دارای واریانس واحد است، آنتروپی و نگانترویی تنها در یک مقدار ثابت و یک علامت تفاوت دارند و لذا داریم:

$$I(y_1, \dots, y_n) = C - \sum_i J(y_i) \quad (33-2)$$

که در آن C یک ثابت است و به W بستگی ندارد. این فرمول رابطه بین اطلاعات متقابل و نگانترویی را نشان می دهد.

از آنجایی که اطلاعات متقابل یک معیار طبیعی نظریه اطلاعات برای استقلال متغیرهای تصادفی است، می توان از آن به عنوان معیاری برای یافتن تبدیل ICA استفاده کرد. در این روش که جایگزینی برای روش تقریب مدل است، ICA یک بردار تصادفی \bar{x} به صورت یک تبدیل برگشت پذیر مانند معادله $\bar{s} = W\bar{x}$ تعریف می شود که در آن ماتریس W طوری تعریف می شود که اطلاعات متقابل مؤلفه های تبدیل شده s_i کمینه شود. از رابطه ۳۳-۲ مشهود است که یافتن تبدیل

⁵⁵ Kullback-Leibler divergence

معکوسی مانند W که اطلاعات متقابل را کمینه کند چیزی مشابه یافتن جهت‌هایی است که در آنها نگانترویی بیشینه می‌شود. و یا به عبارت دیگر معادل یافتن زیرفضایی یک بُعدی است که افکنشها در آن بیشترین نگانترویی را داشته باشند. رابطه ۲-۳۳ نشان می‌دهد که تخمین ICA با کمینه کردن اطلاعات متقابل معادل بیشینه کردن جمع غیرگوسی بودن تخمین‌هاست وقتی که تخمینها محدودند که ناهمبسته باشند. در واقع قید ناهمبستگی الزامی نیستولی تا حد زیادی محاسبات را ساده می‌کند که این ساده شدن از رابطه ۲-۳۱ تا رابطه ۲-۳۳ مشهودست. بنابراین دیده می‌شود که فرمولاسیون ICA به صورت کمینه کردن اطلاعات متقابل اثبات دیگری بر ایده ذهنی یافتن غیرگوسی ترین جهت‌هاست.

۲-۱-۴-۱-۲- تخمین درست‌نمایی بیشینه

یک روش بسیار متداول برای تخمین مدل ICA تخمین درست‌نمایی بیشینه است که قرابت زیادی با اصل اینفومکس^{۵۶} دارد. در اینجا این روش شرح داده می‌شود و نشان داده می‌شود که الزاماً با کمینه کردن اطلاعات متقابل معادل است. فرموله کردن مستقیم درست‌نمایی یک مدل ICA بدون نویز در (Pham and Garat 1992) انجام شده است و سپس مدل توسط یک بیشینه سازی درست‌نمایی تخمین زده می‌شود. اگر ماتریس A^{-1} با نام $W = (\vec{w}_1, \dots, \vec{w}_n)$ نشان داده شود، درست‌نمایی لگاریتمی شکل زیر را می‌گیرد (Pham and Garat 1992):

$$L = \sum_{t=1}^T \sum_{i=1}^n \log f_i(\vec{w}_i^T \vec{x}(t)) + T \log |\det W| \quad (۲-۳۴)$$

که در آن f_i تابع توزیع s_i است (که در اینجا دانسته فرض شده است). عبارت $\log |\det W|$ در درست‌نمایی از قانون کلاسیک تبدیل (خطی) متغیرهای تصادفی و توزیعهای آنها می‌آید که طبق آن هر بردار تصادفی \vec{x} با توزیع p_x و هر ماتریس دلخواه W ، توزیع $\vec{y} = W\vec{x}$ برابرست با $p_x(W\vec{x})|\det W|$

یک تابع هدف دیگر نیز از دیدگاه شبکه عصبی مشتق شده است (Bell and Sejnowski 1997; Nadal and Parga 1997). این تابع بر مبنای بیشینه سازی آنتروپی خروجی (یا جریان اطلاعات) یک شبکه عصبی با خروجی های غیرخطی است. فرض می‌شود که \vec{x} ورودی شبکه عصبی است و خروجی آن به شکل $\phi_i(\vec{w}_i^T \vec{x})$ است که ϕ_i یک تابع اسکالر غیرخطی است و \vec{w}_i بردار وزنه‌های نورون‌هاست. برای بیشینه کردن آنتروپی خروجی داریم:

$$L_2 = H(\phi_1(\vec{w}_1^T \vec{x}), \dots, \phi_n(\vec{w}_n^T \vec{x})) \quad (۲-۳۵)$$

اگر ϕ_i خوب انتخاب شود این چارچوب می‌تواند ICA را نیز تخمین بزند. در واقع پژوهشگران (Cardoso 1989) نتایج شگفت‌انگیزی را منتشر کرده اند که ثابت می‌کند اصل بیشینه سازی آنتروپی شبکه^{۵۷} یا «اینفومکس» با تخمین درست‌نمایی بیشینه هم ارز است. این هم ارزی در صورتی برقرار است که توابع غیرخطی ϕ_i که در شبکه عصبی استفاده می‌شوند برابر تابع توزیع تجمعی توزیعهای f_i باشند، یعنی $\phi_i'(\cdot) = f_i(\cdot)$.

⁵⁶ Infomax principle

⁵⁷ network entropy maximization principle

برای دیدن ارتباط بین درست‌نمایی و اطلاعات متقابل، امید ریاضی درست‌نمایی لگاریتمی گرفته می‌شود:

$$\frac{1}{T} E\{L\} = \sum_{i=1}^n \log f_i(\bar{w}_i^T \bar{x}) + \log |\det W| \quad (۳۶-۲)$$

در واقع اگر f_i برابر توزیع واقعی $\bar{w}_i^T \bar{x}$ بود جمله اول عبارت برابر $-\sum_i H(\bar{w}_i^T \bar{x})$ می‌شد. بنابراین درست‌نمایی با صرف‌نظر از یک عدد ثابت برابر با قرینه اطلاعات متقابل می‌شد که در رابطه ۲-۳۱ آورده شده بود. در عمل این رابطه قویتر از آنچه دیده شد نیز هست زیرا توزیعهای مؤلفه‌های مستقل مجهول است. کاری که می‌توان کرد تخمین زدن توزیع $\bar{w}_i^T \bar{x}$ طی روش بیشینه سازی درست‌نمایی و سپس استفاده از آن برای تقریب توزیع s_i است. در این حالت درست‌نمایی و اطلاعات متقابل برابرند.

در هر حال تفاوت کوچکی وجود دارد که ممکن است در عمل خیلی به چشم بیاید. مشکل تخمین درست‌نمایی بیشینه اینست که توزیعهای f_i نباید درست تخمین زده شده باشد، البته دقت بالای آنها مهم نیست بلکه کافیت که زیرگوسی یا زبرگوسی بودن آنها درست تخمین زده شود (Cardoso 1989; Lee, Mumford et al. 1998; Hyvärinen and Oja 2000). در بسیاری از حالات دانش پیشین کافی از مؤلفه‌های مستقل موجود است و نیازی به تخمین حالت ذاتی آنها از داده‌ها وجود ندارد. در هر حال اگر اطلاعاتی که در مورد ذات مؤلفه‌های مستقل در دست است نادرست باشد، تخمین ML نتایج کاملاً غلطی بدست می‌دهد. بالعکس اگر معیار غیرگوسی بودن مناسبی استفاده شود این مشکل معمولاً رخ نمی‌دهد.

۲-۱-۴-۱۱-ارتباط ICA و تعقیب افکنش^{۵۸}

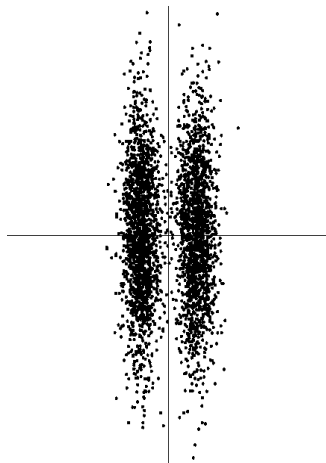
تعقیب افکنش (Friedman and Tukey 1971; Huber 1985; Friedman 1987) تکنیکی آماری است که افکنشهای «جالب» برای داده‌های چندبُعدی را می‌یابد. چنین افکنشهایی می‌توانند برای مصور سازی بهینه داده‌ها، تخمین توزیع و رگرسیون استفاده شود. در نسخه یک بعدی تعقیب افکنش، این روش به دنبال جهت‌هایی می‌گردد که افکنش روی آنها دارای توزیع جالبی باشد، یعنی ساختار خاصی را نشان دهد. بحث شده است (Huber 1985) که جالبترین جهت‌ها آنهایی هستند که کمترین خاصیت گوسی را از خود نشان می‌دهند که دقیقاً همان نکته ایست که برای تخمین ICA استفاده می‌شود.

سودمندی این یافته در شکل ۲-۱۱ به تصویر کشیده شده است که افکنش روی جهتی که تعقیب افکنش یافته است (افقی) کاملاً ساختار خوشه بندی شده داده را نشان می‌دهد. از سوی دیگر افکنش روی نخستین مؤلفه اصلی (عمودی) نمی‌تواند این ساختار را نشان دهد.

بنابراین در حالت کلی ICA را می‌توان نسخه‌ای از تعقیب افکنش در نظر گرفت. تمام معیارهای غیرگوسی بودن و الگوریتمهای ICA مربوطه را می‌توان به نوعی راهکارهای تعقیب افکنش دانست. البته تعقیب افکنش این اجازه را می‌دهد که وقتی مؤلفه‌های مستقل s_i کمتری از متغیرهای اصلی x_i وجود دارد نیز کار را ادامه داد. برای این منظور این روش فرض می‌کند که ابعادی از فضا که توسط مؤلفه‌های مستقل پوشش داده نشده اند پر از نویز گوسی هستند. بدین طریق با محاسبه

⁵⁸ Projection Pursuit

جهتهای تعقیب افکنش غیرگوسی می توان به طور کارایی مؤلفه های مستقل را تخمین زد. وقتی تمام جهتهای غیرگوسی پیدا شوند تمام مؤلفه های مستقل نیز تخمین زده می شوند. چنین روشی را می توان تلفیقی از تعقیب افکنش و ICA برشمرد. با اینحال باید در نظر داشت که در این فرمولاسیون هیچ مدل داده یا فرضی در مورد مؤلفه های مستقل وجود ندارد. اگر مدل ICA برقرار باشد، بهینه سازی معیارهای غیرگوسی بودن ICA مؤلفه های مستقل را می سازد ولی اگر مدل برقرار نباشد، آنگاه چیزی که بدست می آید جهتهای تعقیب افکنش است (Hyvärinen and Oja 2000).



شکل ۲-۱۱: نمایشی از روش تعقیب افکنش و جالب بودن افکنشهای غیرگوسی. داده هایی که در شک دیده می شود آشکارا به دو خوشه تقسیم می شود. با اینحال مؤلفه اصلی یعنی راستای واریانس بیشینه عمودی است که تمایزی بین خوشه ها قائل نمی شود. از سوی دیگر جهت شدیداً غیرگوسی تعقیب افکنش افقی است که بهترین جداکنندگی را برای خوشه ها ایجاد می کند.

۲-۱-۴-۱۲- نتیجه گیری

روش ICA یک تکنیک همه منظوره آماری است که در آن داده های تصادفی مشاهده شده با یک تبدیل خطی به مؤلفه هایی تبدیل می شوند که تا حد امکان مستقلند و توزیع جالبی دارند. می توان ICA را به عنوان یک مدل تخمین متغیرهای نهفته فرمول کرد. راهکار شهودی استفاده از بیشترین غیرگوسی بودن برای تولید توابع های هدف گوناگونی استفاده شود که بهینه سازی آنها به تخمین مدل ICA می انجامد. از سوی دیگر با مدد از روشهای کلاسیکتری مانند تخمین درستنمایی بیشینه یا کمینه کردن اطلاعات متقابل می توان ICA را تخمین زد. شگفت آنکه این روشهای تقریباً هم ارزند. یک روش کارای محاسباتی برای محاسبه ICA به نام FastICA پیشنهاد شده است (Hyvärinen and Oja 2000) که از این اصول بهره می برد. کاربردهای متعددی برای ICA موجود است که از آن قبیل پردازش صوت، پردازش سیگنالهای زیستی، مخابرات، اقتصاد و پردازش تصویر را می توان نام برد. در این پژوهش از روند مطالعات، ایده های مطرح شده، و راهکارهای حل مسائل ICA در زمینه های دیگر برای تکوین مدلی برای پردازش تصویر استفاده شده است که به آنها مجدداً اشاره خواهد شد.

۲-۲- روش مطالعات عصب شناسی

۲-۲-۱- مطالعات عملکرد گروهی نوروپها

از دهه ۱۹۸۰ میلادی، پزشکان از اسکنرهای MRI^{۵۹} برای تشخیص بیماریهای بافت نرم استفاده می کردند. این ماشینها با قرار دادن بخشی از بدن درون یک میدان مغناطیسی قوی و اسکن آن کار می کنند. سپس میدانهای ضعیفتری با زاویه های خاصی نسبت به میدان اصلی به بدن اعمال می شوند که باعث می شوند که هسته هیدروژن موجود در بافتهای بدن (که عمدتاً در مولکولهای آب و چربی یافت می شوند) دچار نوسان شوند و سیگنالهای ضعیف الکترومغناطیسی از خود ساطع کنند. مشخصات دقیق این سیگنالها نه تنها به مکان هسته بلکه به محیط فیزیکی و شیمیایی اطراف هسته ساطع کننده بستگی دارد. داده های جمع آوری شده توسط اسکنرهای MRI در تصاویری سه بعدی گرد هم آورده می شوند که به پزشکان امکان تشخیص ناهنجاریهای بسیاری را بدون هیچ عملیات تخریبی می دهند.

دانشمندان بسیار زود متوجه شدند که فناوری MRI می تواند راهی را نیز برای تعیین فعالیت ارگانهای خاص فراهم کند. ایده اصلی اینکار اسکن سریعتر ولی کم دقت تر است که عکسهایی بسیار فوری از کارکرد پویای بدن بدست می دهد. این متدولوژی به عنوان MRI کارکردی^{۶۰} یا اختصاراً fMRI شناخته می شود.

یکی از چیزهایی که توسط fMRI قابل پیگیری است اینست که خون در یک ناحیه خاص حاوی چه میزان اکسیژن است. علت این مسئله خاصیت مغناطیسی هموگلوبین، مولکول انتقال اکسیژن در خون، است که به میزان اکسیژنی که منتقل می کند وابسته است. از آنجاییکه سلولهای عصبی مغز هنگامی که مشغول پردازش اطلاعات هستند نیازمند میزان بیشتری خون اکسیژن دار هستند، یک دستگاه اسکن مغزی fMRI می توان مکان هندسی فعالیت ذهنی را ردیابی کند.

این تکنیک در اوایل دهه ۱۹۹۰ ابداع شد و از آن زمان فیزیولوژیستهای بسیاری علاقمند بودند تا ببینند که این سیستم قادر به نشان دادن چه چیزهایی است. یکی از کارهایی که با این دستگاه می توان انجام داد یافتن ارتباط بخشهایی از مغز که فعالیت بیشتری دارند با نوع محرکهایی است که به فرد نشان داده می شود. به عنوان مثال در پژوهشی که در سال ۱۹۹۹ توسط کانویشر^{۶۱} و همکاران توسط fMRI صورت گرفت مشخص شد که بخشی موسوم به FFA^{۶۲} در مغز انسان به چهره های انسانی (و تا حدی کمتر به سر انسان) واکنش نشان می دهد و هر چقدر هم این ویژگیها در تصویر کوچک باشد باز هم این فعالیت مقدار زیادی را دارد. یکی دیگر از روشهایی که بسیار سودمند است، تفریق شناختی^{۶۳} نامیده می شود و مبتنی بر این ایده است که دو محرک که کمی با هم متفاوتند (معمولاً یکی از آنها دارا یک ویژگی افزونه نسبت به دیگری است) به فرد تحت نظر نشان داده می شود و بخشهای جدیدی از مغز نیز علاوه بر بخشهای قبلی بکار می افتند که تاثیر آن ویژگی را نشان می دهند.

⁵⁹ magnetic resonance imaging

⁶⁰ functional MRI

⁶¹ Kanwisher et al.

⁶² fusiform face area

⁶³ cognitive subtraction

این روش به دلیل ماهیت ماکروسکوپی و ساختاری قادر به نشان دادن عملکرد دقیق نورونها نیست و برای تعیین ساختارها و عملکرد گروهی نورونها مناسب است (Harris 2010).

۲-۲-۲- مطالعات تک نورونها

برخی از پژوهشگران معتقدند که برای فهم کارکرد سیستم بینایی باید ابتدا به عملکرد اجزای سازنده آن یعنی نورونها پرداخت. بدین منظور به بررسی تک تک نورونهای بخشهای مختلف مغز می پردازند و کارکرد آنها به ازای محرکهای مختلف می سنجند که به طور کلی آنها را ضبط تک نورون^{۶۴} می نامند. به عنوان مثال برای این بررسی این نورونها باید در موجود زنده (اغلب گربه یا میمونهای ماکاکو^{۶۵}) یک پروب را دقیقاً در سلول مورد نظر قرار داد. برای اینکار از میکروسکوپیهای لیزری استفاده می شود تا از ورود پروب به همان نورون اطمینان حاصل شود. سپس دامنه فعالیت الکتریکی نورون به ازای محرکهای گوناگون روی الکتروسکوپ نمایش داده می شود و تابعی که نورون از آن تبعیت می کند کشف می شود. محرکهای استفاده شده معمولاً شامل پنجره های ساده برای بررسی عملکرد سلولهای ساده و خاصیت پاسخها، گوشه ها و زاویه ها، شکلهای ساده و گهگاه شکلهای رسته هایی خاص از اشیا است.

۲-۲-۳- مطالعات آسیب شناسی

بعضی از این مطالعات شامل ایجاد آسیبهایی عمدی^{۶۶} در هر بخشهایی از مغز میمون ها مثلاً در بخشهای مختلف مسیر بینایی آنهاست. بقیه هم شامل مطالعه بر روی انسان است که آسیب های مغزی در انسان را که در بخش خاصی از مغز بوده است نشان می دهد. برای مثال دو نمونه از کشفیاتی که از این طریق حاصل شده است در پایین آورده می شود:

نورونهایی در ناحیه IT یافت می شود که اختصاصاً به تصاویر چهره انسانی پاسخ می دهند. این کشف ابتدا از طریق ضبط بین سلولی^{۶۷} در میمونها حاصل شده است و سپس وجود آنها توسط مطالعات fMRI روی انسان تایید شده است. این کشف برای نوروفیزیولوژیستها جالب بود زیرا مدت‌های مدیدی به دنبال علت بیماری پروسپانگوزیا^{۶۸} بودند که در آن بیمار با وجود اینکه دارای بینایی سالمی در بقیه قسمت‌هاست، در تشخیص چهره دچار مشکل می شود. این عارضه در پی آسیبهایی به بخش فوق مخطط قشر خاکستری بینایی بوجود می آید که تا حد زیادی بیانگر اینست که نورونهای IT در آن دخیل هستند.

ویژوال اگنوزیا^{۶۹} که معمولاً در پی آسیبهایی دو طرفه پس سری-گیجگاهی بوجود می آید، عارضه ایست که فرد را در تمایز شکل اشیا ناتوان می سازد. حتی گاهی بیمار قدرت تمیز خوبی برای رنگها و بافتها دارد. مبتلایان به ویژوال اگنوزیا در تشخیص شکل هندسی اشیا، حروف الفبا و نقاشی های سیاه و سفید ضعیف عمل می کنند ولی در حرکت دادن دستشان به سمت یک شیء، چرخاندن مچشان برای وارد کردن یک شیء به یک محفظه، یا قرار دادن درست انگشتانشان برای بلند

⁶⁴single cell recording

⁶⁵Macaque Monkeys

⁶⁶selective lesions

⁶⁷intracellular

⁶⁸prosopagnosia

⁶⁹Visual agnosia

کردن یک شیء مشکلی ندارند. بنابراین ویژوال اگنوزیا قدرت هشیاری آگاهانه بینایی را از بیمار سلب می کند ولی قدرت تعامل با اشیا را بدون صدمه باقی می گذارد. برعکس، در بیماری آپتیکال آتاکسیا^{۷۰} افرادی که بخش آهیانه ای مغز آنها آسیب دیده است می توانند اشیا را تشخیص دهند ولی نمی توانند بدرستی آنها را بردارند و یا از آنها استفاده کنند. مشاهده چنین افرادی به پژوهشگران کمک کرده است که از نقش مسیرهای بطنی و خلفی پرده بردارند.

۲-۲-۴- استفاده از مدل برای پیش بینی خواص قشر بینایی مغز

مدلهای اولیه نوروهای قشر خاکستری مخ بر مبنای آزمایشاتی بنا شد که سعی داشتند خصوصیات پاسخ مشاهده شده به مجموعه محدودی از محرکها - میله ها، پنجره ها، و شکلهای ساده- را به نوروهای فعال شده نسبت دهند و هرچا که نتایج آزمایش با نظر محقق سازگار بود، به بیان نقش کارکردی این سلولها بپردازند. بعدها روشهایی برای کشف خودکار زیرحوزه های خطی و تشخیص نوع مناسب غیرخطی بودن تهیه شد که برای اینکار از نمونه برداری تصادفی فضای محرک برای تعیین توابع پاسخ نوروها استفاده می شد (De Boer and Kuyper 1968; Jones and Palmer 1987; Van Steveninck and Bialek 1988; Schwartz, Pillow et al. 2006). محققان درباره شکل ویژگیهای کد شده تصویر و نوع غیرخطی بودن چند فرض می کردند تا پاسخ نورو را تعریف کنند. این مدلها میدانهای گیرندگی^{۷۱} سلولهای ساده که با توصیفهای قبل از خود سازگار بوده است را پوشش داده اند (Ringach 2002) و شروع به کشف جزئیات ساختار مؤلفه های غیرخطی نوروهای مراتب بالاتر کردند (Rust, Schwartz et al. 2005; Touryan, Felsen et al. 2005). بدلیل محدودیتهای محاسباتی این روش تنها وقتی مناسب است که نوروها دارای خصوصیات پاسخ نسبتاً ساده ای باشند. به علاوه توزیع محرک باید تصادفی باشد و از خصوصیات خاصی تبعیت کند (مثلاً باید دارای تقارن کروی باشد) (Paninski 2003). این مسئله از آنرو اهمیت دارد که استفاده از این روش را هنگام مطالعه پاسخها به تصاویر طبیعی مسدود می کند (یا دست کم بایاس زیادی را در آن ایجاد می کند). این پاسخها بسیار غیرگوسی هستند و آمارگان آنها چندان شناخته شده نیست (Rust and Movshon 2005). از سوی دیگر محرک تصادفی که برای نمونه برداری متعادل فضای ورودی طراحی شده است و دارای آمارگان منظمی است (مثلاً یک تصویر نویز سفید یا الگوهای شطرنجی چشمکزن) نمی تواند به اندازه کافی از نوروهای لایه های بعد سلسله مراتب بینایی پاسخ استخراج کنند چون آنها در انتخاب تصویری که فعالشان کند سختگیر و سختگیرتر می شوند.

این فرایند به وضوح توسط چند عامل مختل شده است. نمونه گیری کاملاً تصادفی از فضای ورودی یک راه مناسب نیست، انتخاب دستی آنها وقتی که هدف محاسباتی ناحیه تحت بررسی نامعلوم است دشوار می باشد، و افزون بر اینها انتخاب یک مجموعه خاص از این محرکها تفسیری از کدگذاری نرونی را به دنبال خواهد داشت که اجباراً مناسب نیست (مثلاً نسبت دادن تحلیل فوریه، لبه یابی، کدگذاری شکل به بخشهایی از سیستم بینایی). اگر نواحی بینایی خصوصیات مجردتری را از تصویر کدگذاری می کنند، فعالیت نرونی برای همه محرکهایی که یک خاصیت خاص را دارند بدون تغییر خواهد بود. آزمودن با یک محرک نمی تواند این نامتغیری را آشکار کند و یا خاصیت مجرد کدگذاری شده را کشف کند. در همان حال، تعامل بین نوروها می تواند الگوهای پاسخ پیچیده ای تولید کند که با ضبط محدود تک نوروها بسادگی قابل شرح نباشد.

⁷⁰optical ataxia

⁷¹receptive field

در تحلیل های آماری از تصاویر، معمولاً مدلی در نظر گرفته می شود که نحوه ای ایجاد تصاویر را مشخص می کند. سپس پارامترهای این مدل با استفاده از داده های موجود تخمین زده می شوند تا مدل به داده ها منطبق شود. میزان شایستگی روش های مختلف با توجه به مدل در نظر گرفته شده برای تصاویر سنجیده می شود. روش هایی که مدل آنها تا حد امکان کلی باشد و شرط های کمی برای آن تعیین کرده باشند، مدل های مناسبی هستند. به عنوان نمونه می توان تحلیل مؤلفه های مستقل^{۷۲} از تصاویر طبیعی را نام برد (Bell and Sejnowski 1997). مؤلفه های مستقل برای یک مجموعه از داده ها، مؤلفه هایی هستند ابعاد داده های نگاشت شده به آنها از یکدیگر مستقل می باشند. این تحقیق توانست با تجزیه تصاویر به مؤلفه های مستقل، خصوصیات نورون های ساده در V1 را شبیه سازی کند.

همانگونه که بیان شد مدلها به عنوان مسئول محاسبات پردازشهای عصبی، از لحاظ پیش بینی هایی که درباره کدهای نورونی و تفسیر خصوصیات مشاهده شده در سلولها بسیار سودمند هستند. بنابراین استفاده از آنها در بخشهایی از مغز که ناشناخته باقی مانده اند از اهمیت بسزایی برخوردار است ولی تاکنون مدل‌های اندکی درباره پردازش در نواحی بینایی بالاتر مانند V2 یا V4 دست به پیش بینی زده اند یا به توجیه کارکردی پدیده های غیرخطی V1 مبادرت ورزیده اند. تمام اینها بیانگر اینست که نظریه های محاسباتی، که حول محور فرضیه های خاصی از عملکرد نورون بنا شده اند، برای هدایت جستجو و تفسیر کدگذاری نورونی ورودیهای تصویری پیچیده لازم هستند. این پژوهش در راستای این نیاز قرار گرفته است و از مدل‌های نظری برای پیش بینی ویژگیهای تک تک سلولها و نیز جمعیت نورونها استفاده می کند.

۲-۳- شبکه های عصبی برای بازنمایی

۲-۳-۱- قوانین یادگیری

۲-۳-۲-۱- قانون یادگیری هب

یادگیری هب یک قانون یادگیری بدون ناظر است که به عنوان مدلی برای اصلاح قدرت سیناپس بین نورونها که با فعالیت آنها تغییر می کند پیشنهاد شده است. این قانون یادگیری قدرت سیناپسی را متناسب با فعالیت نورونهای پیش سیناپسی و پس سیناپسی تغییر می دهد. از آنجا که ورودیهایی که همزمان فعال هستند در به وجود آمدن فعالیت در نورون خروجی همکاری دارند، یادگیری هب ساختار همبستگی ورودی را پیدا می کند. برای یک واحد با تنها یک خروجی می توان نشان داد که یادگیری هب واریانس فعالیت خروجی را بیشینه می کند که این امر با در نظر گرفتن محدودیتهای اشباع روی هر وزن است و قدرت کلی اتصال به نورون بعدی (خروجی) را محدود می کند. از آنجایی که نخستین مؤلفه اصلی به بردار وزنی مربوط است که واریانس خروجی را بیشینه می کند، آنگاه قانون هب با لحاظ کردن اینکه بردار وزن طول واحدی دارد، هم ارز با یافتن نخستین مؤلفه اصلی ورودی است (Bartlett 1998).

⁷² independent component analysis

برای یک واحد با تنها یک خروجی، y ، که فعالیت y در آن برابر جمع وزندار ورودیهاست، $y = \sum_i w_i x_i$ ، الگوریتم یادگیری هب ساده

$$\Delta w_i = \alpha x_i y \quad (37-2)$$

با نرخ یادگیری α بردار وزنها $\vec{w} = (w_1, \dots, w_n)$ را به سمت نخستین مؤلفه اصلی داده ها \vec{x} سوق می دهد. در این الگوریتم ساده یادگیری، اندازه w محدود نشده است. اوجاً^{۷۳} این الگوریتم را طوری تغییر داد که اندازه w پس از هر گام نرمالیزه شود. با داشتن α کوچک به قدر کافی، یادگیری هب با نرمال کردن اندازه توسط رابطه زیر تخمین زده می شود:

$$\Delta w = \alpha y(x - wy) \quad (38-2)$$

این قانون یادگیری به مؤلفه اصلی با اندازه واحد همگرا می شود. عبارت $-wy^2$ در این رابطه در صورت بزرگ شدن بیش از حد w ، اندازه را کوچک نگه می دارد و اگر w خیلی کوچک هم بود اجازه بزرگ شدن اندازه را می دهد. درباره N واحد خروجی که همگی برای فعال شدن رقابت می کنند، یادگیری هب می تواند فضای N مؤلفه اصلی اول ورودی را معین کند. با داشتن فرم مناسبی از رقابت، قانون هب صریحاً هر N مؤلفه اصلی فعالیت‌های لایه خروجی را بازنمایی می کند. قانون یادگیری برای وزن w_j به واحد خروجی y_j که صریحاً نخستین N مؤلفه اصلی داده را پیدا می کند اینگونه است:

$$\Delta w_j = \alpha y_j (x - w_j y_j + 2 \sum_{k=1}^{j-1}) \quad (39-2)$$

این الگوریتم خروجی های بعدی را مجبور می کند که مؤلفه های اصلی بعدی داده را یاد بگیرند و اینکار را با کم کردن تخمین مؤلفه های قبلی از ورودی پیش از آنکه اتصالات به یک خروجی بروز شوند انجام می دهد. لینکسر^{۷۴} نشان داده است که برای حالتی که واحد یک خروجی دارد، یادگیری هب انتقال اطلاعات بین ورودی و خروجی را بیشینه می کند. نرخ انتقال اطلاعات شانون^{۷۵}

$$R = I(x, y) = H(y) - H(y|x) \quad (40-2)$$

مقدار اطلاعاتی را که دانستن خروجی y درباره ورودی x منتقل می کند بدست می دهد و معادل اطلاعات متقابل بین آنهاست، $I(x, y)$. برای یک خروجی تنها با توزیع گوسی، رابطه $2-40$ با بیشینه کردن واریانس خروجی بیشینه می شود [124]. بیشینه کردن واریانس خروجی با فید توزیع گوسی، توزیع پاسخی را ایجاد می کند که تا جای ممکن هموار و صاف است (آنتروپی بالایی دارد). بیشینه کردن آنتروپی خروجی با توجه به یک وزن w رابطه $2-40$ را بیشینه می کند، زیرا دومین جمله، $H(y|x)$ ، نویز است و به w بستگی ندارد.

⁷³Oja

⁷⁴Linkser

⁷⁵Shannon

لینکسر معتقد است که حفظ بیشینه اطلاعات یک اصل سازماندهی برای سیستم ادراکی چند لایه است. لایه های بالاتر نیازی ندارند که داده های خام را از روی خلاصه ای که از لایه های پایینتر دریافت می کنند بازسازی کنند. هدف اینست که اطلاعات تا جای ممکن حفظ شوند تا لایه های بالاتر بتوانند از اطلاعات محیطی برای تمایز مقادیر مربوط به عملهای متفاوت استفاده کنند. لینکسر نشان داد که چگونه میدان گیرندگی ساختار یافته با خصوصیات تحلیل ویژگی مربوط به میدانهای گیرندگی مشاهده شده در شبکیه، LGN، و قشر بینایی مخ می توانند از اصل حفظ بیشینه اطلاعاتی نتیجه گرفته شوند که در قوانین یادگیری محلی پیاده سازی شده اند (البته با در نظر گرفتن قیدهای مربوطه) (Linkser 1988). بیشینه سازی اطلاعات می تواند به چند خروجی نیز تعمیم یابد (Linkser 1988).

۲-۲-۳-۲- قوانین یادگیری برای کشف صریح وابستگیهای آماری

یک سیستم ادراکی می تواند حول سیگنالهای آموزش درونی سیستم سازماندهی شوند. این سیگنالها از مفروضات بخشهای درونی ورودیهای ادراکی که ریشه های مشترکی در دنیای بیرون دارند تولید می شوند. یکی از مفروضات اینست که ورودی بینایی از منابع فیزیکی نشأت می گیرد که تقریباً در کل فضا ثابت هستند. به عنوان مثال عمق معمولاً در بسیاری از ورودیهای بینایی تغییر نسبتاً کمی دارد مگر در مرز اشیاء. الگوریتمهای یادگیری که صریحاً وابستگیهای آماری ورودی را کدگذاری می کنند می کوشند که این پایداریها^{۷۶} را بیابند. خروجی واقعی چنین آشکارگرهای نامتغیرها نمایانگر میزانی است که ورودی فعلی از مدل شبکه درباره قاعده مندی دنیای پیرامون تخطی کرده است. مکانیسم یادگیری هب که قبلاً شرح داده شد یک وسیله برای کدگذاری وابستگی (همبستگی) مرتبه دوم در ورودی است (Bartlett 1998).

الگوریتم GMAX یک قانون یادگیری برای چندین ورودی به یک واحد خروجی است که هدف آن کاهش افزونگی است. این الگوریتم توزیع پاسخ P در واحد خروجی را با توزیع پاسخ Q مقایسه می کند که این توزیع دوم زمانی مورد انتظار است که ورودیها کاملاً مستقل باشند. الگوریتم یادگیری باعث می شود که واحد وابستگیهای آماری ورودی را با بیشینه کردن تفاوت بین P و Q کشف کند. عامل P توسط پاسخ همه مجموعه داده با این تنظیمات فعلی وزنها تعریف می شود و به طور صریح می تواند با نمونه گیری تمام 2^n حالت ممکن Q واحد ورودی محاسبه شود. قانون یادگیری GMAX محدود به یادگیری یک واحد تک خروجی و واحدهای باینری احتمالاتی است.

بکر^{۷۷} مدل GMAX را به ورودیهای پیوسته با توزیع گوسی تعمیم داد. حاصل این تعمیم قانون یادگیری ایست که نسبت واریانس خروجی را به واریانس مورد انتظار در حالتی که ورودیها مستقل باشند، کمینه می کند. این قانون یادگیری وابستگیهای آماری ورودیها را کشف می کند و به نوعی یک آشکارگر نامتغیر بودن است. اگر فرض کنیم که خصوصیات ورودیهای بینایی از منابع ثابت فیزیکی مشتق می شوند، آنگاه یک قانون یادگیری که واریانس خروجی را کمینه می کند حتماً چیزی درباره آن منبع فیزیکی به ما خواهد گفت. بکر این الگوریتم را برای چند واحد خروجی نیز تعمیم داد. این واحدهای خروجی یک مدل تلفیقی از خصوصیات نامتغیر گوناگونی از الگوهای ورودی بدست می دهد. بکر و هینتون نسخه چند واحدی این قانون یادگیری را استفاده کردند تا نشان دهند که چگونه سیگنالهای آموزش هدایت شده از درون سیستم ادراکی می

⁷⁶ constancy

⁷⁷ Becker

تواند از مفروضات درون قسمتهای گوناگون ورودی ادراکی که علت‌های مشترکی در دنیای بیرون دارد نشأت بگیرد. در طرح یادگیر آنها، مدل‌های کوچکی که به بخش‌های جداگانه ولی مرتبط ورودی ادراکی نگاه می‌کنند، این علت‌های مشترک را با تلاش برای تولید خروجی‌هایی که با هم هماهنگی داشته باشد، کشف می‌کند.

واحدهای مختلف مانند بینایی و لامسه دارای ممکن است کیفیت‌های متفاوتی به نظر برسند، یا کیفیت مشابهی در زمان‌های متفاوتی داشته باشند که می‌شود صحنه‌های متوالی دو بعدی از یک شیء سه بعدی چرخان، و یا قسمتهای مجاور - از لحاظ مکانی - را مثال زد. قانون یادگیری، که به نام IMAX معروف شده است، اطلاعات متقابل بین جفت واحدهای خروجی، y_a و y_b را بیشینه می‌کند. با فرض اینکه دو واحد خروجی توسط یک سیگنال زیرین مشترک دارای نویز مستقل گوسی بوجود آمده باشد، اطلاعات مشترک بین سیگنال زیرین و میانگین y_a و y_b از رابطه زیر بدست می‌آید:

$$I = \frac{1}{2} \log \frac{V(y_a + y_b)}{V(y_a - y_b)} \quad (۴۱-۲)$$

در این رابطه V تابع واریانس روی حالت‌های آموزشی است. بیشینه کردن I مربع تفاضل بین خروجی دو ماژول را بسته به اینکه چقدر دو ماژول با تغییر ورودی تغییر کرده اند بیشینه می‌کنند. این الگوریتم را می‌توان اینگونه دید: یک راه ساده برای اینکه خروجی دو واحد مطابق کنند اینست که مربع تفاضل بین خروجی دو ماژول را به عنوان تابع هزینه استفاده کرد (مخرج کسر ۲-۴۱). استفاده از تابع هزینه کمترین مربع تفاضل به تنهایی هر دو ماژول را مجبور می‌کند که یک خروجی ثابت را بسازند که به ورودی هم ربطی ندارد و لذا هیچ اطلاعاتی درباره ورودی نیز در بر ندارد. صورت کسر تابع هزینه را طوری اصلاح کرده است که تابع هزینه کمترین مربع تفاضل را با توجه به میزان تغییر خروجی‌ها به ازای تغییر ورودی کمینه کند. این جمله باعث می‌شود که هر دو ماژول به چیزی که در ورودی هر دو مشترک است پاسخ دهند. بکر و هینتون نشان داده اند که بیشینه کردن اطلاعات متقابل بین قسمتهای مجاور از لحاظ مکانی تصویر می‌تواند عمق را در نقاط اتفافی سطوح خمیده در استریوگرام‌ها بیابد. این شبیه سازی از یک جفت شبکه دو لایه که هر یک تنها یک خروجی دارند تشکیل شده است، که نواحی مکانی متمایز فضای بینایی را به عنوان ورودی می‌گیرد. این ورودی از نقاط تصادفی استریوگرام با میزان عدم تطابقی که اندک اندک تغییر می‌کند تشکیل شده است. پس از آموزش، با وجود اینکه هیچ دانش پیشین در مورد بُعد سوم وجود ندارد، خروجی‌های مدل با عمق متناسبند. این مدل برای تولید جمعیتی از کدها برای میزان عدم تطابق استریو و مدلی برای نقاط ناپیوستگی در عمق توسعه یافته است.

شراودولف و سجنوسکی^{۷۸} الگوریتمی برای یادگیری نامتغیری‌ها پیشنهاد کردند که به روش کمینه کردن مقید واریانس بکر و هینتون بسیار نزدیک بود (Schraudolph and Sejnowski 1992). آنها یک جمله کمینه کننده واریانس ضد-هبی به ترکیب افزوده اند که در آن قدرت اتصالات به نسبت فعالیت نورونهای پیش سیناپسی و پس سیناپسی کاهش می‌یابد، و نیز جمله‌ای افزوده ان که از همگرایی تمام جملات به صفر جلوگیری می‌کند. آنها نشان دادند که مجموعه‌ای از واحدهای رقابت کننده می‌توانند جمعیتی از کدها را برای میزان عدم تطابق استریو در نقاط تصادفی استریوگرام کشف کنند.

⁷⁸Schraudolph and Sejnowski

زمل و هینتون⁷⁹ الگوریتم IMAX را در مسأله یادگیری بازنمایی پارامترهای نمای اشیای ساده، مانند اندازه شیء، مکان و مقیاس آن به کار برده اند. این الگوریتم سعی می کند چند ویژگی از یک تکه محلی تصویر که با هم پیوستگی ندارند را یاد بگیرد و در عین حال بتواند بردار ویژگی که از ورودیهایی که از نظر مکانی مجاورند استخراج می شود را پیش بینی کند. این الگوریتم بالقوه از روشهای ناهمبسته سازی -مانند تحلیل مؤلفه اصلی- قویتر است زیرا هدف آن نه تنها ناهمبسته سازی بردار ویژگی است بلکه در دامنه مکانی دنبال عوامل مشترک نیز می گردد. توسعه این الگوریتم به واحدهایی پیچیده تر از اشیای دو بُعدی ساختگی دچار محدودیت است زیرا محاسبه دترمینان ماتریسهای بد ترکیب⁸⁰ دشوار است.

۲-۳-۲- خواص مغز

یکی از شگفت انگیزترین خصوصیات سازماندهی مغز اینست که بسیاری از سطوح گیرندگی پیرامونی⁸¹ (مانند مسیر شبکه چشم، مجرای حلزونی گوش میانی، و سطح پوست) در مرحله های متوالی پردازشی مغز، بازنمایی توپوگرافیکی دارند. گروههای سلولی همسایه در شبکه به گروههای سلولی همسایه در بخش بینایی تالاموس مرتبط است که آن هم به نوبه خود با ناحیه های همسایه ای در قشر بینایی اتصال دارد. بدین ترتیب یک نقشه نورونی منظم از اطلاعات اخذ شده در گیرنده های سطحی در مراحل متوالی مغز حفظ می شود. چنین نقشه های نورونی نه تنها موقعیت گیرنده های را منعکس می کند، بلکه به کمک چگالی اتصالات میزان حساسیت حسگر به محرکها را نیز مشخص می کند. بطور مثال بخش مرکزی شبکه، فوویا⁸²، دارای بالاترین چگالی گیرندگی در میان بخشهای شبکه است و از اینرو از بیشترین دقت و تیزبینی نیز برخوردار است. متناظر با این نکته بخشی از قشر خاکستری مخ که به اطلاعات رسیده از فوویا اختصاص دارد از ناحیه ای که بخش جانبی شبکه را بازنمایی می کند بزرگتر است که در ناحیه متناظر با بخش جانبی چگالی گیرنده ها و نتیجتاً دقت بینایی پایینتر است (Kandel, Schwartz et al. 2001).

در بیشتر سیستمهای مغز، پردازش اطلاعات دارای سازمانی سلسله مراتبی و هرمی است که اطلاعات حسگر در قاعده هرم دریافت می شود و پیچیدگی اطلاعات پردازش شده هر چه به سمت راس هرم پیش برود بیشتر است. مثلاً در سیستم بینایی هر نورون در عقده زانویی خارجی⁸³ (LGN) که در تالاموس قرار دارد به لکه نوری دیده شده در بخش خاصی از میدان دید پاسخ می دهد. آکسونهای چند نورون همسایه تالاموس روی سلولهای بخش اولیه بینایی قشر خاکستری مخ همگرا می شوند که در آن هر نورون تنها زمانی آتش می کند که چپش خاصی از سلولهای پیش سینوسی فعال باشند. برای مثال یک سلول قشر خاکستری مخ ممکن است تنها وقتی آتش کند که ورودی دارای یک میله روشن در راستایی خاص باشد. سلولهای بخش بینایی اولیه قشر خاکستری مخ نیز به نوبه خود روی سلولهایی در قشر انجمنی همگرا می شوند. پاسخ این سلولها دارای گزینندگی بیشتری روی پاسخهاست و مثلاً به میله ای روشن که در جهت خاصی حرکت می کند پاسخ می دهد. در مراحل بسیار پیشرفته پردازش اطلاعات بینایی در قشر خاکستری مخ، هر نورون به اطلاعات بسیار پیچیده ای مانند شکل و چهره انسان پاسخ می دهد (Kandel, Schwartz et al. 2001).

⁷⁹Zemel and Hinton

⁸⁰ill-conditioned matrices

⁸¹peripheral

⁸²fovea

⁸³lateral geniculate nucleus

اطلاعات در بیش از ۳۵ ناحیه بینایی شناخته شده در قشر خاکستری مخ که به پردازش اطلاعات بینایی اختصاص دارند، هم به طور سری و هم به طور موازی حرکت می کند (Kandel, Schwartz et al. 2001). علاوه بر مسیر سری که سیگنالهای بینایی در مراحل سلسله مراتبی سیستم بینایی می پیمایند، کانالهای متعددی وجود دارد که اطلاعات را به طور موازی پردازش می کند و شبکه ای از مدارهای بسیار پیچیده می سازد. این پیچیدگی را تا حدی می توان به دلیل حلقه های بازخورد متعددی دانست که هر کدام از این نواحی با اتصالات بازگشتی به ناحیه هایی که به آنها اتصال فرستاده اند بوجود آورده اند. جنبه دیگری از این پیچیدگی نیز از ارتباطات با LGN و سوپریور کولیکولوس نشأت می گیرد (McGill-University 2006).

۲-۴- مقدمه ای بر بینایی انسان

توانایی انسان در پردازش داده های حسی و تشخیص الگو موضوعی است که توجه محققان عصب شناسی، هوش مصنوعی و رباتیک و ... را به خود جلب کرده است. انسان در تشخیص چهره ها، اشیا و صداها بلادرنگ عمل می کند و صحت این تشخیص تقریباً از تمام برنامه های کامپیوتری موجود بیشتر است. این خصوصیات سبب شده اند که مغز انسان به عنوان مدلی عالی برای مکانیزم های تشخیص الگو مطرح شود. در این قسمت، به معرفی قشر بینایی در مغز پرداخته می شود و خصوصیات ساختاری، عملکرد نورون ها و نحوه اتصال آنها شرح داده می شود. آشنایی با قشر بینایی مغز پیش نیاز پیاده سازی مدلی از روی مغز برای بینایی رایانه ای می باشد. در این بخش، به معرفی قسمت های مختلف قشر بینایی و خصوصیات نورونی آنها پرداخته می شود.

۲-۴-۱- فرایند بینایی

۲-۴-۱-۱- چشم

همانطور که در برش عرضی چشم دیده می شود (شکل ۲-۱۲) اولین ساختاری که نور از آن عبور می کند تا به چشم برسد رابط^{۸۴} نام دارد. یک غشای شفاف نازک که جلوی چشم را پوشش داده و سپس به درون پلک چشم ختم می شود. بعد از آن قبل از رسیدن نور به بخش های مختلف شبکیه^{۸۵} نور باید بخش های دیگر نوری چشم را بپیماید که شامل قرنیه^{۸۶}، مایع زلالیه^{۸۷}، مردمک^{۸۸} (در مرکز عنیبه^{۸۹})، عدسی و در آخر مایع زجاجیه^{۹۰} (مایعغلیظتر از مایع زلالیه) می باشد (McGill-University 2006).

⁸⁴conjunctive

⁸⁵retina

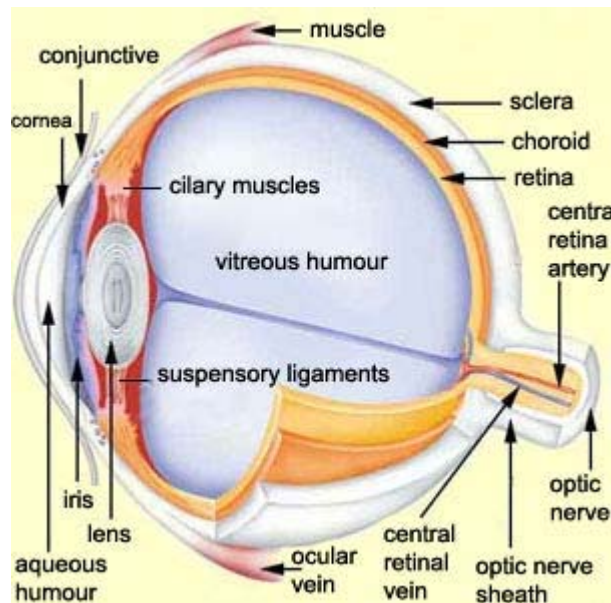
⁸⁶cornea

⁸⁷aqueous humor

⁸⁸pupil

⁸⁹iris

⁹⁰vitreous humor

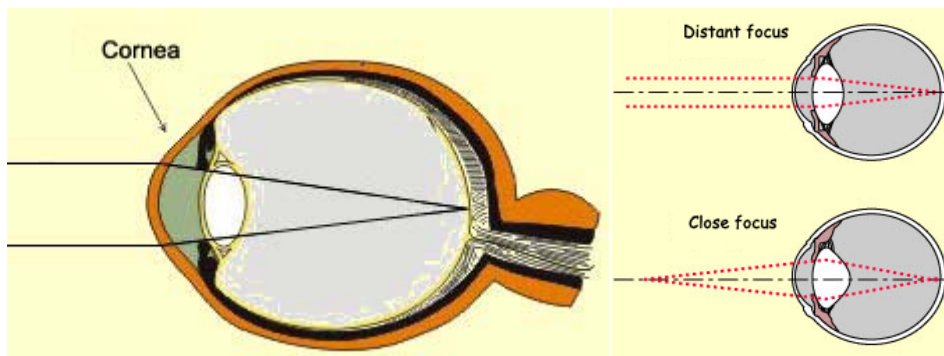


شکل ۲-۱۲: برش عرضی چشم

قرنیه تا سفیدی چشم یا صلبیه^{۹۱} ادامه پیدا می کند که دیواره سخت خارجی چشم را تشکیل می دهد. سه جفت عضله در صلبیه وجود دارد. این عضلات چشمی^{۹۲} کره چشم را قادر می سازد تا در کاسه چشمی در جمجمه حرکت کند. بین صلبیه و شبکیه، مشیمیه^{۹۳} قرار دارد که لایه ای پرعروق است و کار انتقال مواد غذایی به عنبیه و شبکیه را به عهده دارد. مشیمیه شامل رنگدانه های^{۹۴} تیره بی شماری است که درون چشم را همانطور که از درون مردمک دیده می شود، سیاه نشان می دهد. مایع زلالیه و مایع زجاجیه نقش مهمی در تمرکز تصویر روی شبکیه ایفا می کنند، چرا که سرعت نور در این دو محیط نسبت به هم و نیز نسبت به هوا متفاوت می باشد. این دو مایع تمام پرتوهای نوری بجز پرتو عمودی که وارد چشم می شوند را منحرف می کند. بنابراین این پرتوها به مکان مناسبی روی شبکیه برخورد می کنند. این فرایند شکست نور^{۹۵} نامیده می شود. انحنای قرنیه نیز باعث شکست پرتوهای نوری موازی می شود که از اشیاء بسیار دور به چشم می رسند. بعضی از این پرتوها به مرکز قرنیه برخورد می کنند و چون عمود بر قرنیه هستند، زاویه شان تغییر نمی کند و مستقیم به مرکز شبکیه امتداد پیدا می کنند. اما بقیه پرتوها از همین فاصله دور به بخش های منحنی شکل قرنیه برخورد می کنند. این پرتوها انحراف درونی پیدا کرده و دقیقاً در نقطه مرکزی روی شبکیه می افتند و یک تصویر متمرکز شده را تشکیل می دهند. وقتی اشیاء در فاصله دور باشند، قرنیه بیشترین شکست پرتوهای نوری را انجام می دهد تا آنها را در یک نقطه واحد روی شبکیه همگرا کند. عدسی نیز در این عمل شرکت می کند ولی شکست ناشی از آن در حد کمتری است. برای اشیاء نزدیکتر (از فاصله ۹ متر و کمتر از آن) عدسی نقش فعالتری را در تمرکز تصویر روی شبکیه ایفا می کند. هر چقدر پرتوهای نوری رسیده به چشم از اشیاء نزدیکتر باشد بیشتر از هم دور می شوند، بنابراین باید بیشتر شکست پیدا کنند تا بتوانند روی پرده شبکیه به هم

⁹¹ sclera⁹² ocular muscle⁹³ choroid⁹⁴ pigment⁹⁵ refraction

برسند. در این حالت عدسی به شکل دیگری تغییر می یابد تا این شکست نوری افزایش یابد. عدسی از طریق رباطهای معلق^{۹۶} به ماهیچه های مژگانی^{۹۷} وصل شده است که آنها نیز در حقیقت به صلبیه وصل شده اند. این عضلات یک حلقه را درون چشم تشکیل داده اند، که وقتی در حالت استراحت هستند کشش روی رباطهای معلق بیشتر است در نتیجه عدسی پهنتر است و شکست ناشی از آن به میزان کمتری است. از سوی دیگر وقتی عضلات مژگانی فشرده می شوند کشش روی عدسی کمتر می شود در نتیجه عدسی تمایل دارد که به وضعیت طبیعی خودش (یعنی شکل همگرایی بیشتر که پرتوهای نوری بیشتری شکست پیدا می کنند و بیشتر همگرا می شوند) برگردد. این افزایش قدرت شکست که عدسی بدست می آورد به آن اجازه می دهد که یک تصویر واضح از اشیای نزدیک روی پرده شبکیه تشکیل دهد که به این پدیده «تطابق»^{۹۸} اطلاق می شود (شکل ۲-۱۳).



شکل ۲-۱۳: پدیده شکست نور در چشم (چپ)، پدیده تطابق (راست)

رگ های خونی موجود در سطح شبکیه با یک وسیله به نام افتالموسکوپ^{۹۹} قابل دیدن است (وسیله ای که پزشکان توسط آن ته چشم را از طریق مردمک معاینه می کنند). این رگ های خونی از طریق ناحیه ای بی رنگ در شبکیه به نام صفحه بینایی^{۱۰۰} وارد چشم می شوند. صفحه بینایی ابتدای عصب بینایی است که در این نقطه آکسون های سلول های عقده ای شبکیه به هم می رسند و از چشم خارج می شوند. به دلیل این که هیچ گیرنده نوری در صفحه بینایی وجود ندارد این ناحیه از شبکیه نسبت به نور حساس نیست و در این ناحیه بزرگترین رگهای خونی عبور می کند. به همین دلیل به آن نقطه کور^{۱۰۱} نیز گفته می شود، اما هیچ اختلال میدان دید در نقطه کور چشم وجود ندارد زیرا مغز در این نقطه راهی برای پر کردن ادراک بینایی دارد.

در مرکز هر شبکیه ناحیه تاریکتری به نام ماکولا^{۱۰۲} یا لکه وجود دارد که برای بهینه کردن بینایی مرکزی (بر خلاف بینایی پیرامونی)، عملاً خالی از رگ های خونی می باشد. در مرکز لکه، یک گودی کوچک نقطه سیاهی به قطر ۲ میلیمتر

⁹⁶ suspensory ligaments

⁹⁷ ciliary muscle

⁹⁸ accommodation

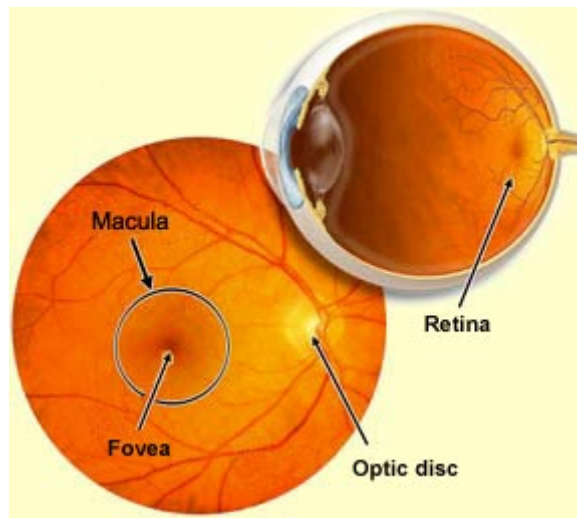
⁹⁹ ophthalmoscope

¹⁰⁰ optic disk

¹⁰¹ blind spot

¹⁰² macula

بوجود می آورد. این گودی فوویا است، همان بخشی از شبکیه که به خصوص از سلولهای مخروطی تشکیل شده و در آن بیشترین تیزبینی^{۱۰۳} و میزان دقت دید تجربه می شود (شکل ۲-۱۴).



شکل ۲-۱۴: جایگاه ماکولا، صفحه نوری و فوویا روی شبکیه

۲-۱-۴-۲- قسمت انتهایی عصب بینایی

آکسون های سلولهای عقده ای شبکیه^{۱۰۴} از میان عصب بینایی، کیاسمای بینایی^{۱۰۵} و مجرای بینایی^{۱۰۶} عبور می کنند. این آکسون ها دور مغز میانی پیچیده و صفحه میانی نیمه گیجگاهی را قطع می کنند و ۸۰٪ آنها به یک تقویت کننده سیناپسی که عقده زانویی خارجی (LGN) نامیده می شود منتهی می شود که در قسمت خلفی تالاموس قرار گرفته است. مقصد اصلی برای مجرای بینایی LGN می باشد (شکل ۲-۱۵).

در برش عرضی، LGN های سمت راست و چپ در یک ساختار ۶ لایه از انواع متفاوت سلولی سازماندهی شده اند (شکل ۲-۱۶). در بررسی سه بعدی، این ۶ لایه شبیه دسته ای از دیسکهای متحدالمرکز به نظر می آیند که اطراف راه بینایی پیچیده شده اند و شباهت زیادی به مفصل زانو دارند (که نام آن نیز از معادل لاتین زانو آمده است). پخش شدن نورون های LGN در لایه های گوناگون نشان می دهد که بعضی جنبه های متمایز اطلاعات بینایی شبکیه ممکن است به صورت جداگانه ای در این رله سیناپسی پردازش شوند و این دقیقاً همان چیزی است که به صورت تجربی ثابت شده است؛ در LGN سمت راست تمام سیگنال های بینایی مربوط به میدان دید سمت چپ پردازش می شود و بالعکس، LGN سمت راست آکسون های شبکیه بینی^{۱۰۷} چپ (طرف مخالف) و شبکیه گیجگاهی^{۱۰۸} راست (طرف موافق) را دریافت می کند.

¹⁰³ visual acuity

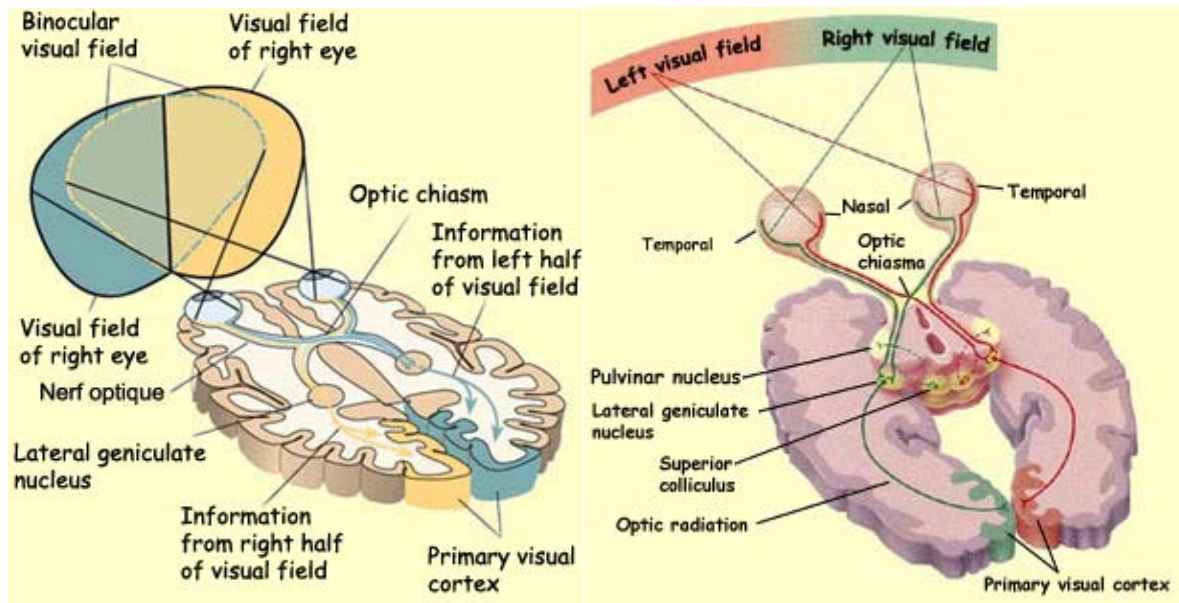
¹⁰⁴ retina ganglion cells

¹⁰⁵ optic chiasm

¹⁰⁶ optic tractus

¹⁰⁷ Nasal retina

¹⁰⁸ Temporal retina



شکل ۲-۱۵: میدان دید یا میدان گیرندگی چشم ناحیه ای از فضا است که وقتی چشم روی نقطه دوری متمرکز است، روی شبکیه می افتد. میدان دید دو چشم تا حد زیادی همپوشانی دارند ولی میدان راست به راست کشیده تر است و بالعکس. ناحیه میانی که در آن میدان دید دو چشم همپوشانی دارند، میدان دید دو چشمی خوانده می شود. تصویر سمت چپ همچنین نشان می دهد که چگونه میدان بینایی چشم راست در قشر بینایی سمت چپ تحلیل می شود و بالعکس. تصویر سمت راست نواحی عبور مسیرهای جداگانه سیگنال بینایی را نشان می دهد.

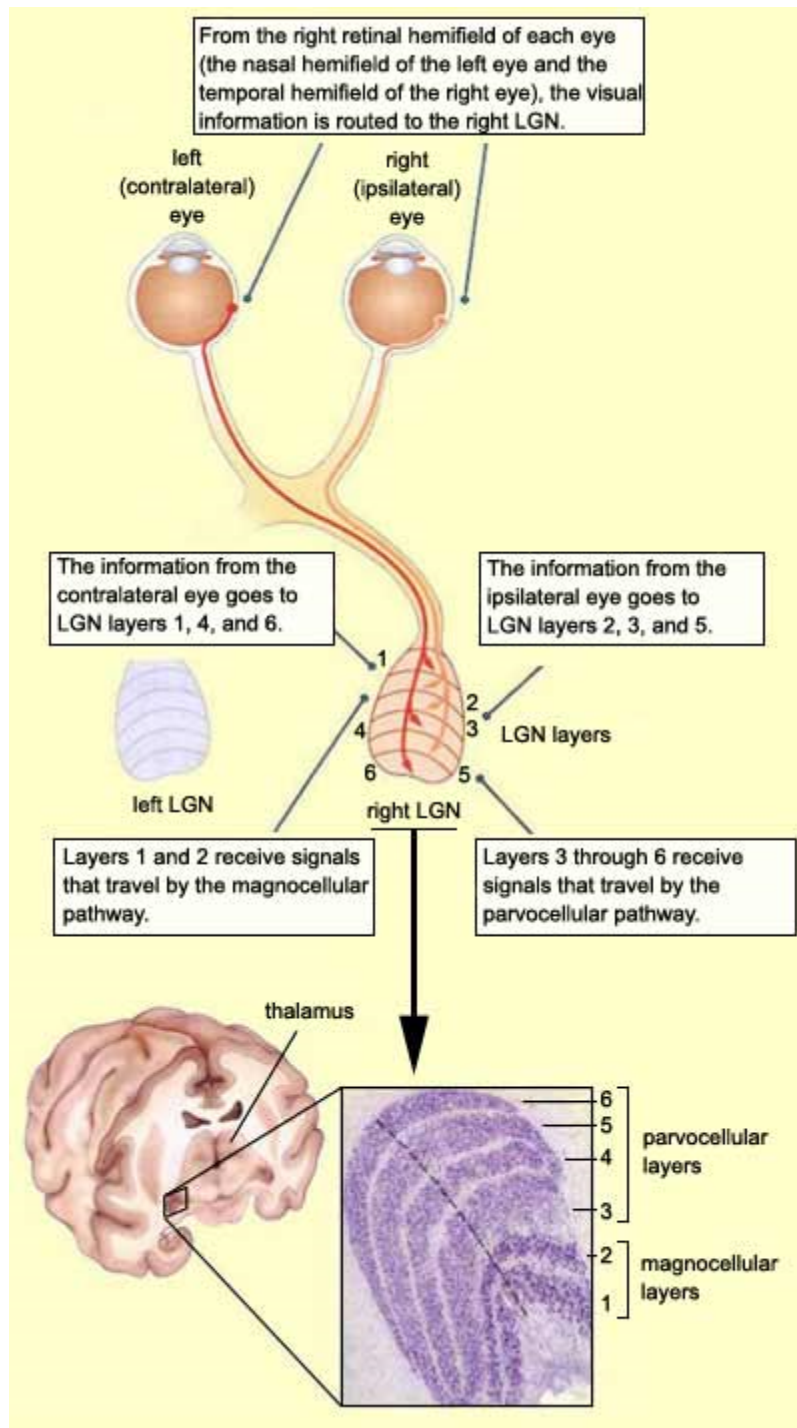
تحقیقات نشان می دهد که آکسون های سمت چشم راست روی لایه های ۲ و ۳ و ۵ در LGN سیناپس دارند در حالیکه آنهایی که از چشم سمت چپ هستند به لایه های ۱ و ۴ و ۶ می روند. این بررسی ها همچنان نشان می دهد که لایه های درونی LGN (۱ و ۲) شامل نورون های بزرگتری نسبت به لایه های خلفی (۳ و ۴ و ۵ و ۶) هستند که به ۲ لایه درونی، لایه های درشت سلول^{۱۰۹} (M) گفته می شود در حالیکه چهار لایه دیگر، لایه های ریزسلول^{۱۱۰} (P) نامیده می شود که در واقع سلول های عقده ای نوع M به لایه های درشت سلولی LGN می روند و سلول های عقده ای نوع P به لایه های ریزسلول می روند.

آکسون های برخی از سلول های عقده ای شبکیه از راه بینایی به ساختارهایی به غیر از LGN (که رابط اصلی بین شبکیه و قشر بینایی است) انشعاب می یابد (شکل ۲-۱۷). یکی از این ساختارها، هیپوتالاموس و به ویژه هسته فوق کیاسما^{۱۱۱}ی آن است که تعداد مشخصی از آکسون های شبکیه را دریافت می کند. هسته فوق کیاسما بخش اصلی ساعت بیولوژیکی داخلی بدن بشمار می آید. سیگنال های بینایی که از آکسون های شبکیه می رسند پیوسته درباره تاریکی و روشنایی محیط اطراف اطلاعاتی را گردآوری می کنند، بنابراین این امکان فراهم می شود که طیف وسیعی از ریتم های بیولوژیکی مانند شامل خواب و بیداری که به چرخه شبانه روز مرتبط است با هم هماهنگ شوند.

¹⁰⁹ magnocellular

¹¹⁰ parvocellular

¹¹¹ superchiasmatic nucleus

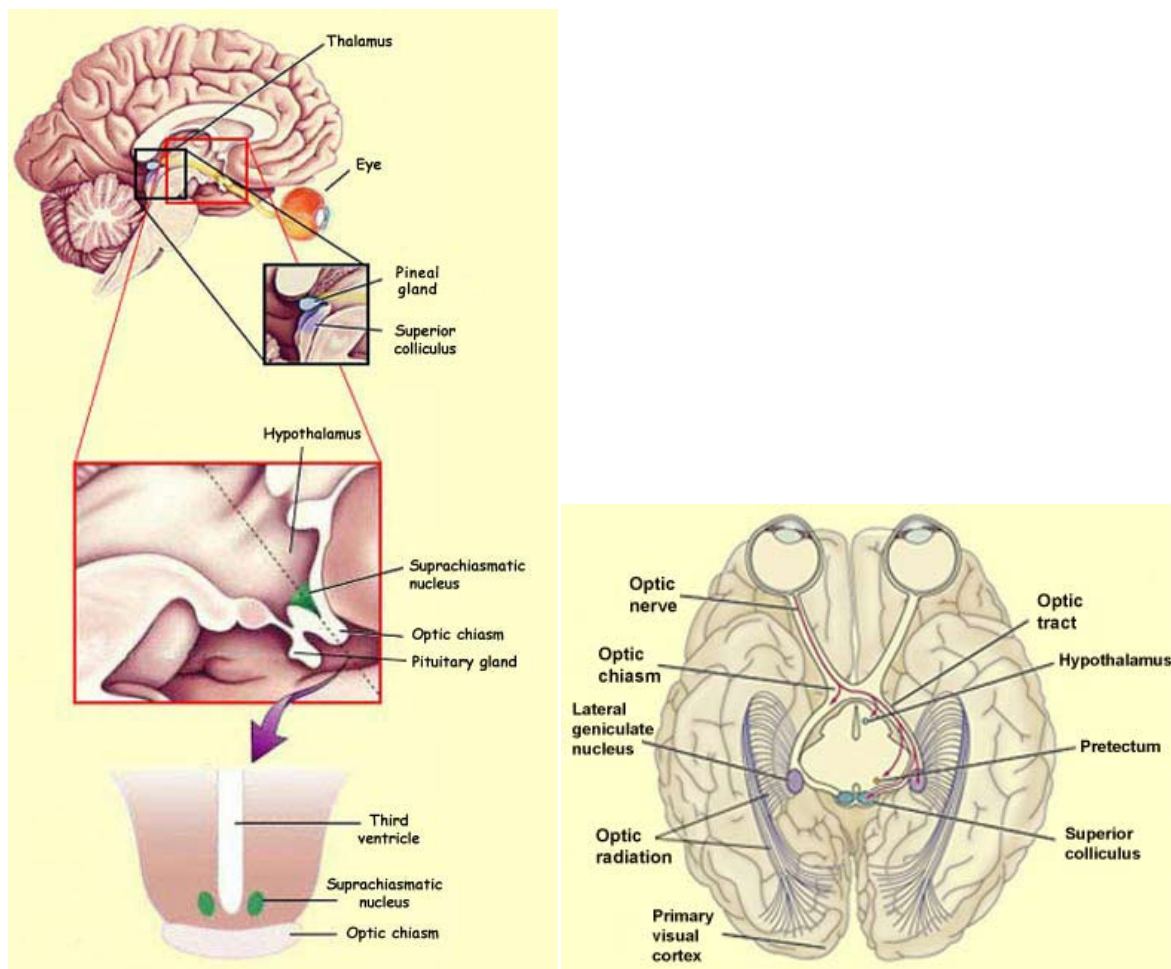


شکل ۲-۱۶: ساختار کلی LGN

آکسون هایی که از برخی دیگر از سلول های عقده ای شبکیه می آیند به پرتکتوم^{۱۱۲} می روند. پرتکتوم شی از مغز میانی است که میزان بازشدن مردمک و حرکت های خاصی از چشم را کنترل می کند. سرانجام حدود ۱۰٪ از آکسون هایی

¹¹² pretectum

که از شبکیه می آیند به یک بخش ریشه ای در مغز میانی که کالیکولوس فوقانی نامیده می شود می روند. این مسیر نسبتاً وسیع است و شامل ۱۵۰,۰۰۰ آکسون می شود که معادل تعداد کل سلول های عقده ای شبکیه یک گربه است.



شکل ۲-۱۷: مکان عقده های سر راه سیگنالهای بینایی

در حقیقت کالیکولوس فوقانی با مجرای بینایی در تمام موجودات مهره دار غیرپستاندار متناظر است که در آنها این برآمدگی شبکیه ای مسیر رگ بیرون بر اصلی از سمت شبکیه است. به دلیل نحوه ای که میدان های گیرندگی شبکیه بر روی همدیگر می افتند، یک پرتوی نوری تابیده شده بر روی شبکیه، تعداد زیادی از نورونهای کالیکولوس فوقانی را فعال می کند. این نرون ها حرکت های چشم و سر را از طریق نورون های حرکتی ساقه مغز کنترل می کنند تا تصویر پرتوی نوری را به فوویا بیاورند. بنابراین مسیر شبکیه ای^{۱۱۳} شامل مکان چشم برای محرک می شود که در آغاز در میدان دید محیطی ظاهر می شود. کالیکولوس فوقانی مانند LGN ارتباطاتی از قشر بینایی اولیه دریافت می کند. نورون های کالیکولوس فوقانی به نوبه خود آکسون هایشان به ساختارهای زیر قشری مانند ساختار مشبک^{۱۱۴}، کالیکولوس تحتانی^{۱۱۵} و نخاع می رود. همچنین نورون های کالیکولوس فوقانی دو ساختار دیگر را تحت تاثیر قرار می دهند: LGN و هسته بالشتک^{۱۱۶}.

¹¹³ retinotectal pathway

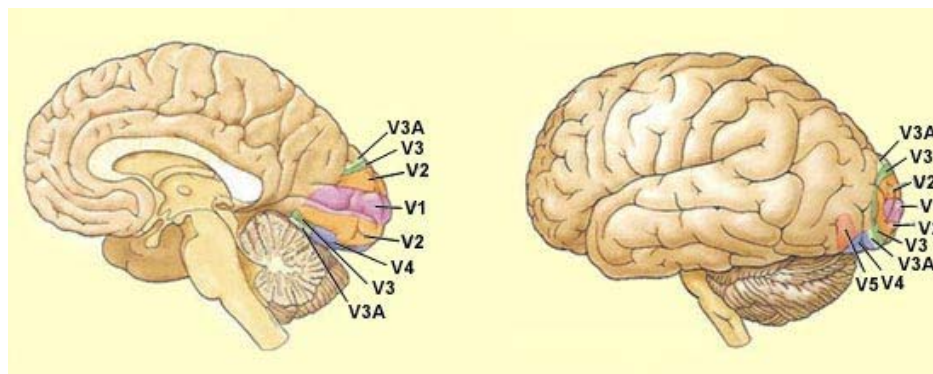
¹¹⁴ reticular formation

هسته بالشتک، یک هسته در بخش پشتی تالاموس است که تحت تاثیر مجرای بینایی از طریق LGN می باشد (از طریق آکسون های متوازی). مثل خیلی از هسته های دیگر تالاموس، هسته بالشتک برای مدت طولانی به عنوان تقویت کننده ای منفعل برای مسیر اطلاعاتی به قشر مغز تلقی می شد، اما چنین دیدگاهی با جمع آوری داده هایی که نشان می دهد نورون های آن پاسخ های بینایی پیچیده ای را که قبلاً فقط گمان می رفت که قشر مغز قادر به انجام آن باشد می تواند انجام دهد، شدیداً تغییر یافت. امروزه هسته بالشتک به عنوان یک مرکز تفسیر تصویر در نظر گرفته می شود که نقش مهمی در توجه^{۱۱۷} و درک حرکتی ایفا می کند. به عنوان مثال هسته بالشتک به پایداری محیط بینایی کمک می کند و به جبران تاثیراتی که حرکات بدنی روی موقعیت تصاویر در شبکیه دارد می پردازد. بنابراین خیلی عجیب نیست که نورون های هسته بالشتک در قشر بینایی ثانویه یافت می شوند که در کشف حرکت بکار گرفته می شوند.

۲-۴-۱-۳- قسمتهای گوناگون بینایی قشر خاکستری مخ

ارتباطات سلسله مراتبی بین قشر اولیه و قشر ثانویه در چندین بخش از مغز وجود دارد که شامل قشر حسی، حرکتی و قشرهای شنیداری هستند. نواحی ثانویه قشر مخ روی آنچه که ناحیه انجمنی^{۱۱۸} نامیده می شود همگرا می شوند تا پردازش اطلاعات با توجه به برخی ورودیهای دیگر انجام گیرد. این ناحیه ها سیگنال های سایر واحدهای حسی را تدریجاً تخصیص می دهد تا یک بازنمایی ترکیبی و چندحسی از دنیا را خلق کند.

تاکنون محققان بیش از ۳۰ ناحیه قشری مختلف را کشف کرده اند که در ادراک بینایی نقش دارند. ناحیه V1 و V2 با تعداد زیادی ناحیه بینایی انجمنی شامل V3، V4، V5، (MT)، PO و ... احاطه شده اند (شکل ۲-۱۸). به هر صورت از روی همه این پیچیدگی ها می توان یک الگوی کلی استخراج کرد.



شکل ۲-۱۸: نواحی دخیل قشر مخ در ادراک بینایی انسان

به دنبال مطالعات انقلابی منتشر شده توسط آنرلیدر^{۱۱۹} و میشکین^{۱۲۰} در سال ۱۹۸۲، دانشمندان دو روش برای پردازش های قشری اطلاعات بینایی مشخص کرده اند: مسیر بینایی بطنی برای شناسایی اشیاء و مسیر بینایی خلفی برای

¹¹⁵inferior colliculus

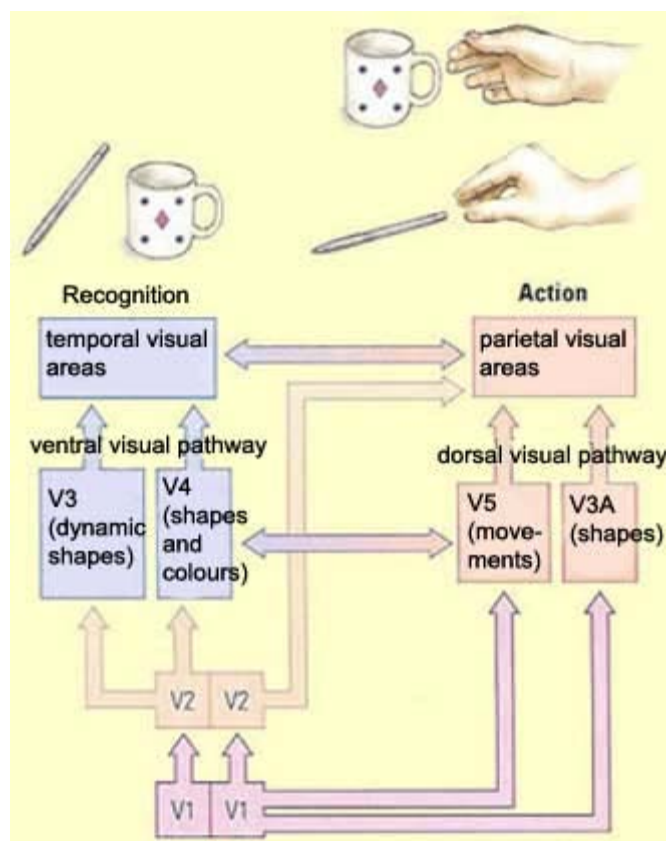
¹¹⁶pulvinar

¹¹⁷attention

¹¹⁸associative areas

¹¹⁹Leslie Ungerleider

تعیین موقعیت آنها در فضا. در واقع دو سیستم قشری اصلی برای پردازش اطلاعات بینایی وجود دارد: مسیر بینایی بطنی^{۱۲۱} که تا بخش گیجگاهی گسترش یافته و یک مسیر بینایی خلفی^{۱۲۲} که به بخش آهیانه ای می رود. عملکرد اصلی مسیر بینایی بطنی ادراک آگاهانه، تشخیص و شناسایی اشیاء به کمک پردازش خاصیت های بینایی ذاتی آنها مانند شکل و رنگی باشد. عملکرد مسیر بینایی خلفی نیز ظاهراً کنترل حرکت بینایی روی اشیاء با پردازش خاصیت های خارجی آنهاست. خاصیت های این نظیر اندازه، موقعیت، و راستای آنها در فضا از این قبیل می باشد که برای تعامل با محیط اطراف حیاتی است. در نواحی قشری که در سیستم بطنی نقش دارد، بازنمایی پیچیده و مخصوصی از دنیای خارج شرح داده می شود. به هر حال مطالعات بعدی، سوالات زیادی در مورد این دو مسیر^{۱۲۳} به وجود آورده است. بعضی از این مطالعات شامل ایجاد آسیب های عمدی در هر کدام از مسیرها در میمون هاست. بقیه هم شامل مطالعه بر روی انسان است که آسیب های مغزی در انسان را که در یکی از این دو مسیر بوده است نشان می دهد.



شکل ۲-۱۹: نواحی قشری موجود در مسیرهای بینایی و ارتباطات آنها

امروزه این گمان وجود دارد که عملکرد اصلی مسیر بینایی خلفی، هدایت اعمال بلادرنگی است که ما در اشیاء اطرافمان می بینیم. بیشتر پردازش های انجام شده در این مسیر، ناآگاهانه می باشد. بنابراین به مسیر خلفی، «مسیر

¹²⁰Mortimer Mishkin

¹²¹ ventral pathway

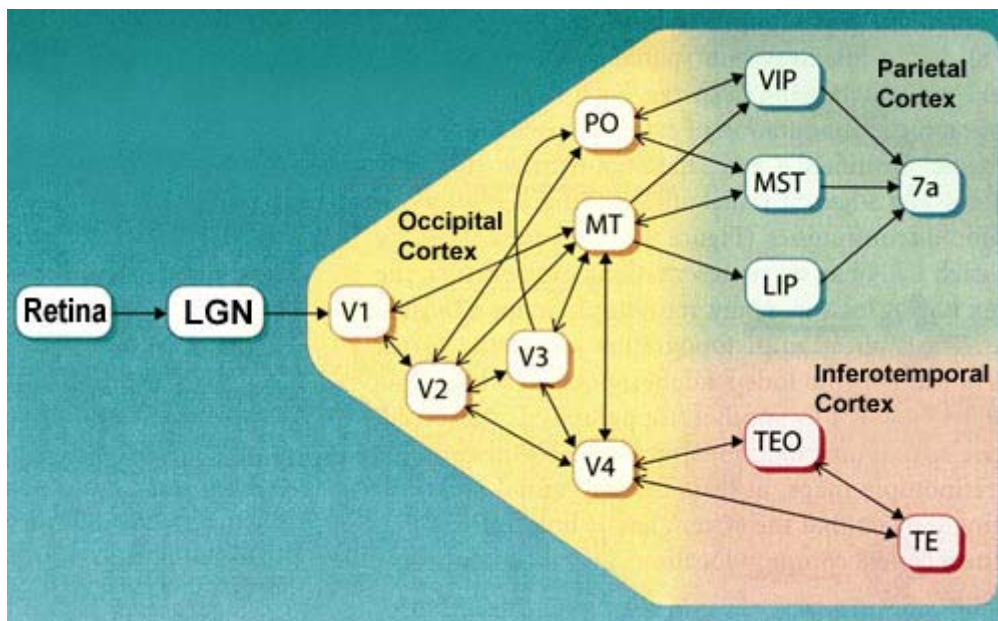
¹²² dorsal pathway

¹²³ dichotomy

کنش^{۱۲۴} نیز گفته می شود، چرا که با یکی کردن ارتباطات سه بعدی بین بدنمان و محیط اطرافمان، امکان تعامل موثرتری با این محیط ایجاد می شود (شکل ۲-۱۹). از سوی دیگر به نظر می رسد که مسیر بینایی بطنی برای شکل گیری معرفی آگاهانه کیفیت اشیاء باشد. بنابراین علاوه بر طبقه بندی مسیر بینایی بطنی و پشتی از حیث عملکرد، از حیث بینایی آگاهانه و ناآگاهانه نیز بین این دو مسیر می توان طبقه بندی ایجاد کرد.

ناحیه V3 انشعابات اصلی را از V2 دریافت کرده و به نواحی V4 و MT می فرستد. نورون های زیادی در ناحیه V3 هستند که ویژگی های مشابهی به ناحیه V2 دارند به عنوان مثال بیشتر آنها برای موقعیت انتخاب شده اند اما هنوز ناشناخته های زیادی در مورد ناحیه V3 وجود دارد و همچنین دارای سلول هایی با ویژگی های پیچیده ترمی باشد. به عنوان مثال بعضی از این سلول ها به رنگ و حرکت حساسند؛ ویژگی هایی که معمولاً در مراحل بعدی پردازش سیگنال های بینایی در مغز تحلیل می شوند.

سلول های ناحیه IT اتصالات زیادی از ناحیه V4 دریافت می کنند و قادر هستند به طیف وسیع رنگ ها و اشکال ساده پاسخ دهند. این سلول ها نقش مهمی در حافظه بینایی بازی می کنند و علاوه بر آن یک ناحیه کلیدی برای تشخیص شیء محسوب می شوند.



شکل ۲-۲۰: اتصالات نواحی بینایی قشر مخ به یکدیگر

مسیر خلفی از چند ناحیه قشری تشکیل می شود که شامل نواحی گیجگاهی میانی^{۱۲۵} (MT یا V5) و نواحی فوق گیجگاهی میانی^{۱۲۶} (MST) و نواحی آهیانه ای خارجی و بطنی^{۱۲۷} (VIP و LIP) می شود.

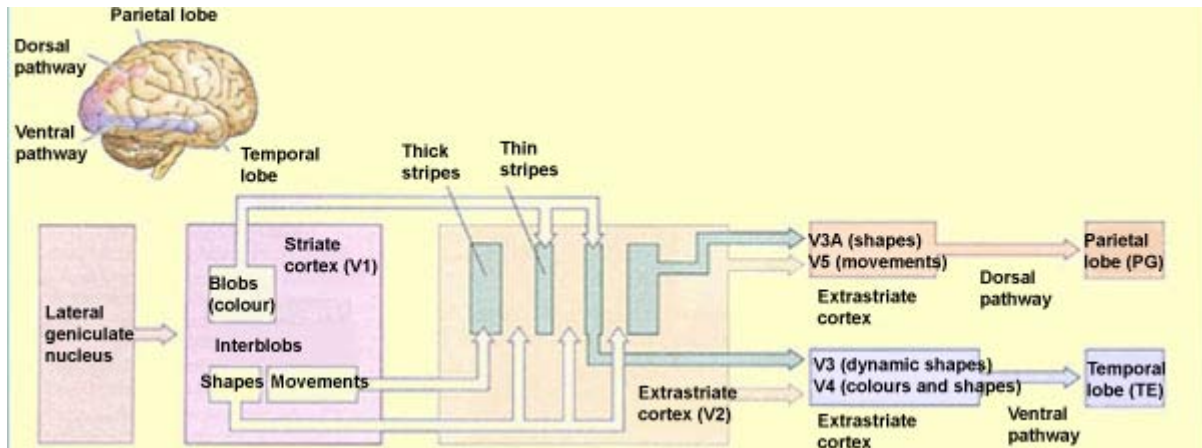
¹²⁴ action pathway

¹²⁵ medial temporal

¹²⁶ medial superior temporal

¹²⁷ ventral and lateral interparietal

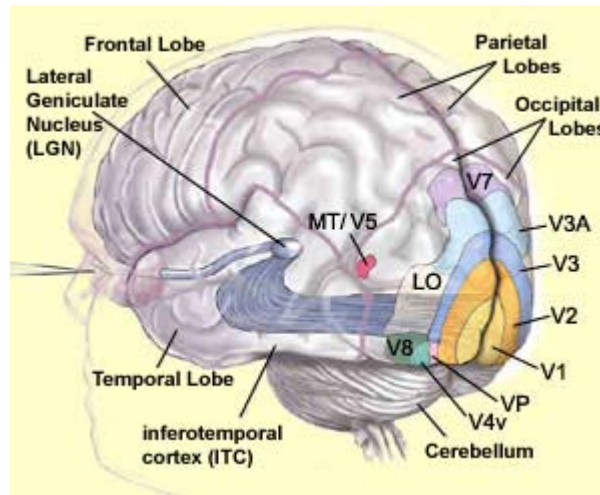
ناحیه V5 یا (MT) عمدتاً به ادراک حرکتی مربوط می شود. این ناحیه تصاویر V2 و V3 را دریافت می کند. همچنین این ناحیه تصاویر لایه IV-B در ناحیه V1 را دریافت می کند، همان لایه ای که به طور قابل توجهی بخشی از کانال درشت سلول است که شامل تحلیل حرکت اشیاء می شود. این کانال همچنین تا ناحیه V2 ادامه می دهد، جایی که نوارهای ضخیم متمرکز شده اند (شکل ۲-۲۱) که شامل میزان زیادی اکسیداز سیتوکروم هستند.



شکل ۲-۲۱: نواحی قشری موجود در مسیرهای بینایی و کانالهای موازی آنها

پیشنهاد شده است که این جداسازی بین سیگنال های درشت مولکول و ریزمولکول تا بالاترین سطح تحلیل بینایی حفظ شود. همچنین تفاوت عملکرد بین مسیر بطنی و خلفی ممکن است به سهم کانال P-IB در گذشته و کانال M در حال حاضر نسبت داده شود.

برای سلول های ناحیه MT به نظر می رسد که حرکت یک شئ خود آن مهمتر باشد، به قدری که ناحیه قشری به عنوان پایه هایی برای موقعیت حرکت شئ برنامه ریزی شده است، درست مثل ستون های راستای خط در ناحیه V1. بعضی از سلول ها در ناحیه MT حتی به نظر می رسد که برای پاسخ گویی به خود مسیر حرکت نیستند، بلکه برای پاسخ گویی به درک مسیر حرکت هستند. به عنوان مثال دو گروه از خطوط مشخص شده در شکل هر کدام با حرکتی ۴۵ درجه به اطراف در جهت عمود، می توانند درکی در جهت حرکت عمودی، در جایی که یکدیگر را قطع می کنند، داشته باشند. در ناحیه V1 سلول هایی که ترجیحی برای زاویه ۴۵ درجه دارند بهترین پاسخ را به این نوع محرک خواهند داشت. اما در ناحیه MT به طور عادی سلول های بسیاری کهبه مسیرهای عمودی پاسخ می دهند، به طور قابل قبولی به دو محرک با زاویه ۴۵ درجه نیز پاسخ دهند و به عبارت دیگر به مسیر ظاهری حرکت پاسخ می دهند.



شکل ۲-۲۲: مکان نواحی دخیل قشر مخ در ادراک بینایی انسان

پس از ناحیه MT (شکل ۲-۲۲)، نواحی دیگری مانند ناحیه MST وجود دارد که شامل تحلیل حرکت است. سلول‌های این ناحیه نه تنها مانند سلول‌های ناحیه MT به حرکت خطی حساسند، بلکه به حرکت شعاعی (رفت و برگشت به یک نقطه) و حرکت دایره‌ای (ساعتگرد و غیر ساعتگرد) نیز حساسند. همچنین این ناحیه دارای گزینندگی برای ساختارهای پیچیده حرکتی هستند، حرکاتی مانند اتفاقاتی که وقتی شما به سمت چیزی می‌روید در اطراف شما رخ می‌دهد.

حتی نورون‌های مشخصی در ناحیه فوق‌گیجگاهی چند حسی^{۱۲۸} (STP) وجود دارد که به حرکت‌های بیولوژیکی که برای بقای زندگی افراد لازم است، پاسخ می‌دهند. از جمله این موارد می‌توان به شناسایی عضوی از گونه جانوری مشابه از روی طرز راه رفتن او اشاره کرد.

مغز شما ممکن است این حجم زیاد از داده‌های مرتبط با حرکت را که به وسیله مسیر خلفی جمع‌آوری شده، برای مقاصد زیادی استفاده کند: برای استخراج اطلاعات مربوط به اشیایی که وقتی حرکت می‌کنید در میدان بینایی شما هستند، یا به منظور هدایت حرکت‌های شما، یا جهت هماهنگی حرکت‌های چشمی شما، یا جهت تشخیص حرکت اشیای اطراف شما که به بقای زندگی شما کمک می‌کند (مثلاً جهت دقیق حرکت اتومبیلی که به سمت شما می‌آید).

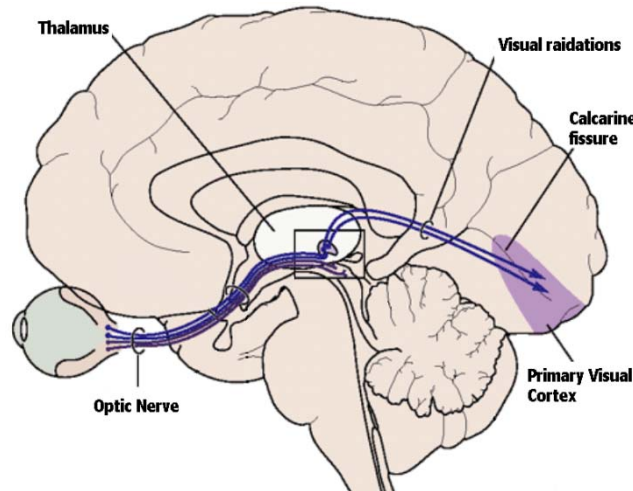
۲-۴-۲- مدل‌های رفتاری بخش بینایی قشر خاکستری مخ

قشر بینایی در قطعه پس‌سری در مغز انسان قرار گرفته است و حدود ۱۵ درصد جسم خاکستری را تشکیل می‌دهد (Kandel, Schwartz et al. 2001). اعصاب بینایی که از چشم خارج می‌شوند پس از عبور از LGN به قشر ابتدایی بینایی^{۱۲۹} وارد می‌شوند (شکل ۲-۲۳). قشر ابتدایی بینایی، بزرگترین قسمت از قشر بینایی است که نورون‌هایی با

¹²⁸superior temporal polysensory

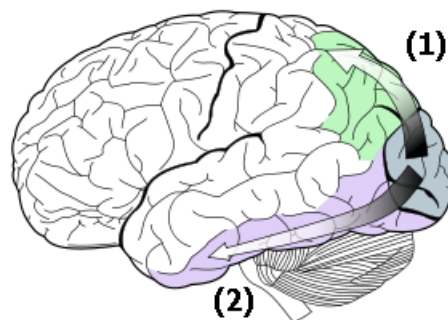
¹²⁹ primary visual cortex

خصوصیات مشابه دارد. این ناحیه معادل ناحیه ۱۷ در تقسیم بندی برادمن^{۱۳۰} می باشد. قشر ابتدایی بینایی که به نام V1 نیز شناخته می شود خروجی نورون های خود را به نواحی مختلفی می فرستد.



شکل ۲-۲۳: قشر ابتدایی بینایی. اعصاب بینایی پس از خروج از چشم و گذر از LGN به قشر ابتدایی بینایی وارد می شوند. (شکل از (Kandel, Schwartz et al. 2001))

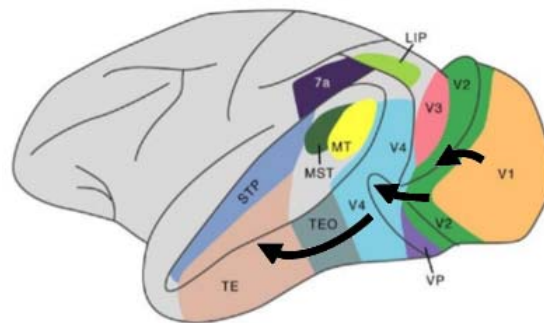
پردازش تصویر ورودی در قشر بینایی پس از ناحیه V1 در دو مسیر کلی به نام های مسیر خلفی و مسیر بطنی ادامه مییابد (شکل ۲-۲۴). یکی از نظریات رایج در مورد پردازش در قشر بینایی، جدایی تشخیص شیء از تشخیص مکان اشیا می باشد (Mishkin, Ungerleider et al. 1983). مطابق با این نظریه، داده های مربوط به تصویر در قشر بینایی مغز در دو مسیر جداگانه پردازش می شوند. مسیر بطنی از V1 شروع می شود و پس از عبور از V2 و V4 به IT ختم می شود. در این مسیر، اطلاعات مربوط به هویت اشیا در تصویر استخراج می شوند. در مسیر دیگر که از V1 به قشر آهیانه می رود اطلاعات مربوط به مکان اشیا پردازش می شود. در این پژوهش تمرکز روی مسیر بطنی است که اشیا موجود در تصویر را تشخیص می دهد.



شکل ۲-۲۴: دو مسیر پردازش داده ها در قشر بینایی. داده های تصویر در قشر بینایی در دو مسیر بطنی و خلفی پردازش می شوند. مسیر خلفی با پیکان شماره (۱) و مسیر بطنی با پیکان شماره (۲) مشخص شده اند.

یکی از مهمترین مطالعاتی که در مورد خصوصیات نورون های قشر بینایی صورت گرفته است مقاله ای است که هابل در سال ۱۹۶۲ ارائه کرد (Hubel and Wiesel 1962). این مقاله رویکرد مطالعه‌ی نورون های قشر بینایی را تغییر داد و جهت مطالعاتی که پس از آن انجام شد را دگرگون ساخت. در این مقاله، نورون ها در قشر ابتدایی بینایی به واحدهایی تعبیر شدند که خصوصیتی را از تصویر استخراج می کنند. هابل در این مقاله، نورون های V1 را به دو دسته تقسیم کرد: نورون های ساده که خصوصیتی مانند تغییر شدت روشنایی در تصویر را در مقیاس کوچک تشخیص می دهند و نورون های پیچیده که از تعدادی از نورون های ساده ورودی دریافت کرده و خروجی خود را برابر با ترکیبی از ورودی‌ها تنظیم می کنند. این نخستین مقاله ای بود که تشخیص خصوصیت را به نورون ها و همچنین ساختار سلسله مراتبی را به قشر بینایی نسبت داد.

در طی سال های گذشته مطالعات بسیاری بر روی نورون های قسمت های مختلف مسیر بطنی انجام شده است که نتیجه آنها منجر به فرضیات مختلفی در مورد عملکرد و ساختار اتصال این نورون ها شده است. یکی از این فرضیات، ساختار مسیر بطنی را متشکل از نواحی مختلفی می داند که به صورت سلسله مراتبی سازمان یافته اند. طبق این نظریه، مسیری که داده های تصویر در مغز طی می کنند از نواحی V1، V2، V4 و سپس IT می گذرد. نحوه ی اتصال این نواحی در شکل ۲-۲۵ نشان داده شده است. در این سلسله مراتب، هر ناحیه از ناحیه قبلی ورودی دریافت می کند و خروجی خود را به ناحیه بعدی فرستد. در این سلسله مراتب، خصوصیتی که نورون ها در تصویر تشخیص می دهند به تدریج پیچیده تر شده و مقاومت آنها در برابر تغییرات بیشتر می شود.

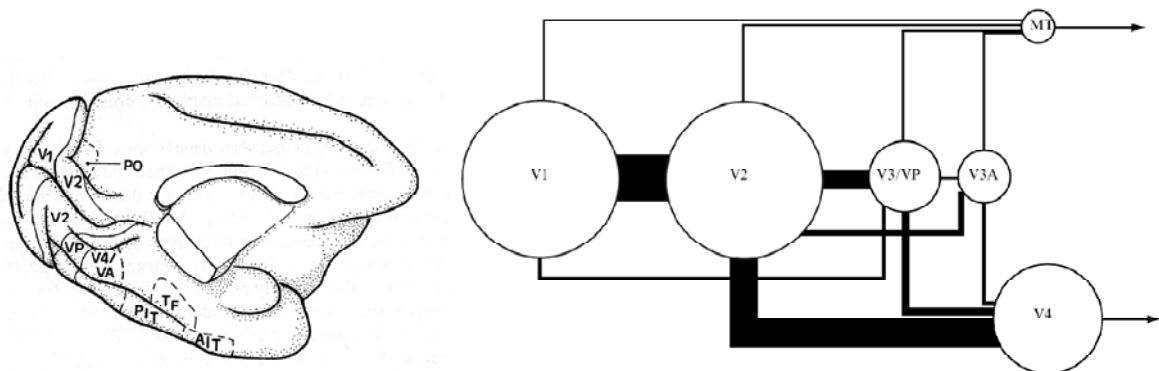


شکل ۲-۲۵: مسیر بطنی در قشر بینایی. مسیر بطنی شامل نواحی V1، V2، V4 و دو ناحیه TE و TE0 در IT می شود. در شکل بالا پیکان های سیاه رنگ جریان داده ها در طول این مسیر را مشخص می کنند.

در حالی که جریان شکمی که به V4 و IT می رود کد کننده شکل و رنگ در نظر گرفته شده است، جریان پشتی از ناحیه MT (V5) می گذرد و تحرک و نسبت های مکانی را کدگذاری می کند. البته این توصیف به نوعی کارتونی است زیرا تصویر کامل شامل بازخوردها و ارتباطات متقابل بین تقریباً تمام بخشهای قشر خاکستری مخ است (Felleman and Van Essen 1991).

مدلهای کلاسیک پردازش تصویر فرض می کنند که لایه های تصویری بالاتر مانند V2 و V4 به مرور شکلهای پیچیده تری را کدگذاری می کنند (Lennie 1998; Felleman and Van Essen 1991). نورونهای این نواحی بیشتر و

بیشتر نسبت به پاسخ خود گزینندگی دارند و محرکهای ساده ای که برای مشخص کردن خصوصیات سلولهای ساده و پیچیده استفاده می شوند، قابلیت متمایز کردن نورونهای سطح بالاتر را به اندازه کافی ندارند. بطور مشابه مدلهایی که بر اساس این ویژگیها بنا شده اند در پیش بینی رسته های جدید تصاویر و فیلمها بسیار ضعیف عمل می کنند (David and Gallant 2005). برای بررسی خصوصیات کدگذاری، پژوهشگران تجربی باید به محرکهای ساختگی پارامتری شده (و معمولاً مخصوص همان آزمایش) روی آورند که فرضیه خاصی در مورد کدگذاری نورونی می آزماید. سلولهای V2 تا حد زیادی خواص V1 را از خود نشان می دهند ولی با این وجود میدان گیرندگی آنها بزرگتر است و ممکن است شکلهای پیچیده تری نسبت به نورونهای V1 را کدگذاری کنند (Levitt, Kiper et al. 1994; Hegd'e and Van Essen 2007). از آنجایی که این نواحی بخشی از مسیر بطنی است که شامل نواحی سطح بالای حساس به شکل مانند ناحیه فوق آهیانه است، کدگذاری احتمالی آنها از شکل، کانتور، و بافت مورد بررسی قرار گرفته است. بسیاری از نورونها در V2 و V4 بخوبی به محرکهای غیرپنجره ای مانند الگوهای قطبی یا هایپربولیک (Gallant, Braun et al. 1993)، زوایا (Ito and Komatsu 2004)، شکلهای (Hegd'e and Van Essen 2000) و کانتورهای منحنی (Pasupathy and Connor 2001) پاسخ می دهند. با اینحال این مطالعات پراکنده به دشواری تطبیق داده می شوند زیرا جنبه های متفاوتی از بازنمایی نورون از تصویر را بررسی کرده اند و نتایج آنها در زمینه کدگذاری صحنه های طبیعی پیچیده ای مانند چهره انسان بسادگی قابل تفسیر نیست.



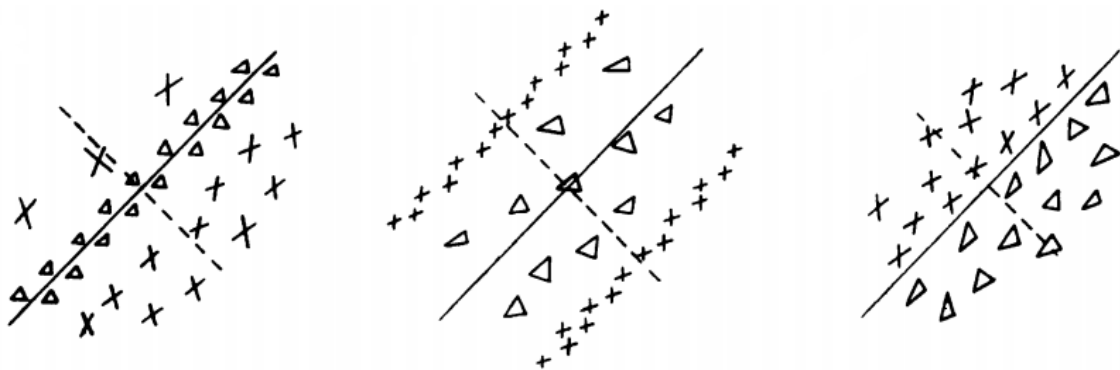
شکل ۲-۲۶: نواحی بینایی قشر مخ میمون ماکاگو و اندازه نسبی در بخش پس سری قشر مخ

برای نشان دادن هرچه بیشتر پیچیدگی این موضوع، باید به این نکته اذعان داشت که تحقیقات جدید حاکی از آن هستند که تفاوتی بین نواحی قشر خاکستری مخ در واقع بسیار ظریف هستند. مثلاً پاسخ تعداد زیادی از نورونهای V1 دارای خاصیت‌های پیچیده ای هستند و خاصیت پاسخهای نورونی در طول V1، V2 و V4 شدیداً هرگونه دسته بندی و جداسازی را زیر سوال می برند (Hegd'e and Van Essen, 2007). برای آشنایی بیشتر با هر یک از این نواحی و خصوصیات نورون های آنها، در ادامه به طور جداگانه به مرور هر کدام از این نواحی پرداخته می شود.

۲-۴-۳- لایه V1

این ناحیه بزرگترین قسمت از قشر بینایی مغز است که حدود ۲۰ درصد از این قشر را تشکیل می دهد. این قسمت در قطعه پس سری قرار دارد و مطابق با ناحیه ۱۷ در تقسیم بندی برادمن می باشد. در (Hubel and Wiesel 1962)، خصوصیت نورون های این قسمت، حساسیت به تغییر شدت روشنایی در مقیاس کوچک ذکر شد.

قشر بینایی اولیه (V1) است انواع گوناگونی از سلولها از جمله سلولهایی بنام «ساده» و «پیچیده» دارد (Hubel and Wiesel 1962; Hubel and Wiesel 1962). هر نورون در قشر بینایی دارای خصوصیتی به نام میدان گیرندگی^{۱۳۱} می باشد. میدان گیرندگی معادل قسمتی از تصویر ورودی است که تغییر در خصوصیات آن، فعالیت نورون را تحت تاثیر قرار می دهد. در ناحیه V1، نورون ها دارای ناحیه دریافت کوچکی هستند که اندازه آن در نقطه مرکزی دید حدود ۱° (معادل ۷×۷ پیکسل) می باشد. ناحیه دریافت در تصویر محلی می باشد و هر نورون قسمت خاصی از تصویر را از طریق ناحیه دریافت خود مشاهده می کند. شکل ۲-۲۷ نمونه ای از ساختار ناحیه دریافت برای نورونهای V1 را نشان می دهد. در این شکل، مثلث نقاطی در ناحیه دریافت را نشان می دهد که تاباندن نور به آنها نورون را از فعالیت باز می دارد. ضربدر نقاطی را نشان می دهد که تاباندن نور به آنها فعالیت نورون را تشدید می کند. همانطور که مشخص است مرز کاملی بین نقاط بازدارنده و نقاط وادارنده وجود دارد. در نورون های V1، این مرز در جهت های مختلفی مشاهده می شود.

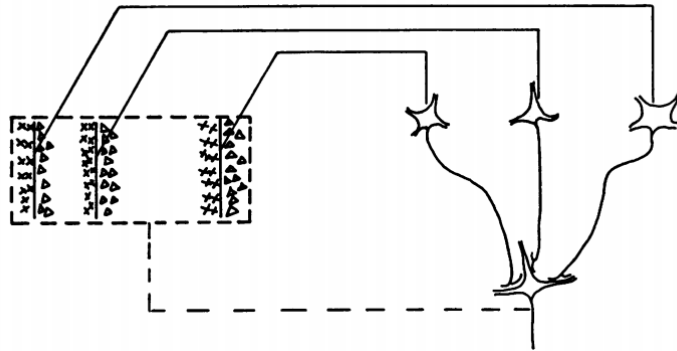


شکل ۲-۲۷: ساختار ناحیه دریافت برای چند نورون در ناحیه V1 (Hubel and Wiesel 1962). مثلث نشان دهنده نقاط بازدارنده و ضربدر نشاندهنده نقاط وادارنده می باشد. خط متصل برای هر ناحیه دریافت، جهت آن ناحیه دریافت را مشخص می کند. در نورون های V1، انواع مختلفی از جهت ها بین ۰ تا ۱۸۰ درجه مشاهده می شود.

تفسیری که هابل از این نورون ها داشت این بود که نورون های V1 تغییرات شدت روشنایی را در جهت های مختلف تشخیص می دهند. می توان گفت که این نورون ها مانند فیلترهای لبه عمل می کنند و لبه ها را در جهت های مختلف از تصویر استخراج می کنند. یکی از وجوه تمایز این نورون ها با نورون های پیچیده در میزان حساسیت آنها به موقعیت مکانی لبه می باشد. نورون های ساده به موقعیت مکانی لبه بسیار حساس هستند و با تغییر موقعیت آن، پاسخ این نورون ها به سرعت افت می کند. از طرف دیگر، نورون های پیچیده، لبه در یک جهت خاص را در یک محدوده مکانی که نسبت به نورون های ساده گسترده تر است تشخیص می دهند. این خصوصیات باعث شد که هابل، نظریه پردازش سلسله مراتبی برای این نورون ها را ارائه دهد (شکل ۲-۲۸).

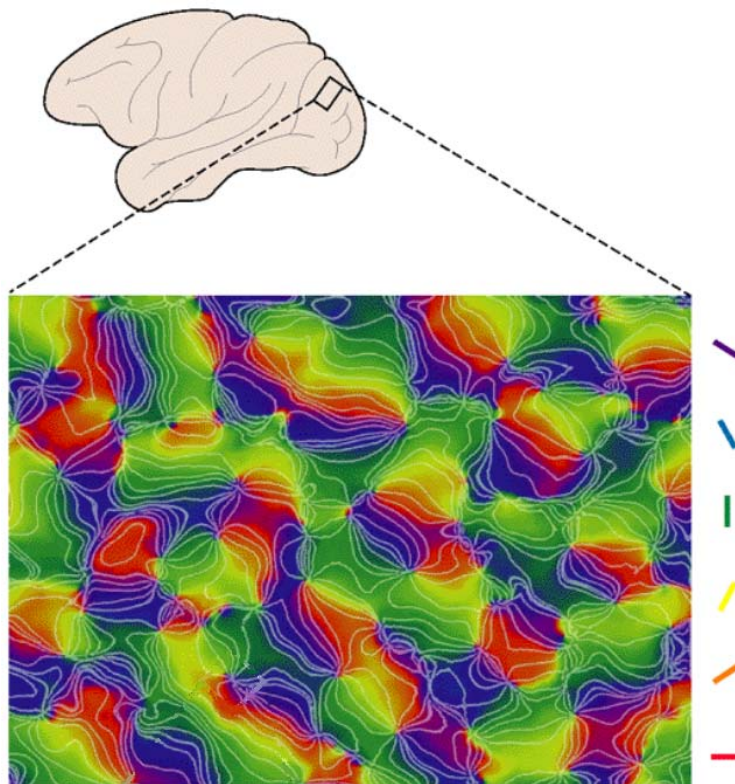
نکته دیگر در مورد نورون های V1، ترکیب قرار گرفتن آنها در قشر ابتدایی بینایی با توجه به جهت مورد نظر آنها می باشد. گفته شد که نورون های V1 انواعی از جهت ها را در تصویر ورودی تشخیص می دهند. نحوه قرار گرفتن این نورون ها در قشر ابتدایی بینایی بدین صورت است که نورون های با جهت های مشابه در کنار یکدیگر قرار گرفته اند. البته بریدگی هایی نیز در این فضاییکناخت مشاهده می شود که طیفی از نورون ها با جهت های مختلف را شامل می شود.

¹³¹ receptive field



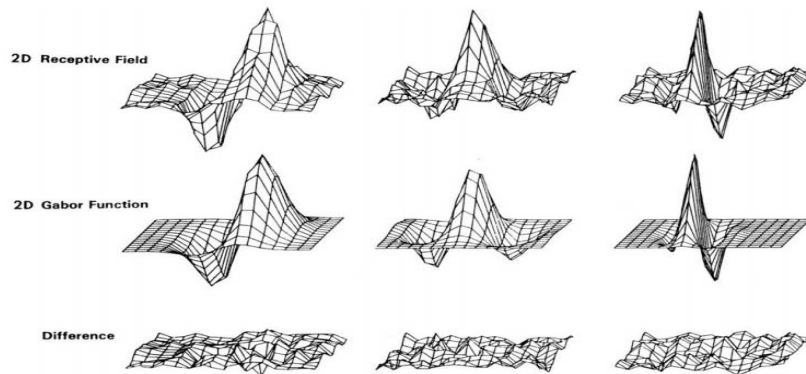
شکل ۲-۲۸: ساختار یک نورون پیچیده ارائه شده توسط هابل. در این شکل، نورون پیچیده در سمت راست از سه نورون ساده در سمت چپ ورودی دریافت می‌کند. این سه نورون در مکان‌های مختلف، لبه‌های یکسانی را تشخیص می‌دهند. بدین ترتیب نورون پیچیده قادر است که لبه با جهت خاص را در مکان‌های مختلف تشخیص دهد.

در شکل ۲-۲۹ نحوه قرار گرفتن نورون‌های V1 را بر اساس جهت آنها مشهود است. در این شکل، میله‌های کنار تصویر جهت‌های متناظر هر رنگ را نشان می‌دهند. دقت شود که جهت‌هایی که نورون‌های V1 تشخیص می‌دهند محدود به زوایای ذکر شده نمی‌باشند. این نورون‌ها طیف پیوسته‌ای از جهت‌ها بین زاویه ۰ تا ۱۸۰ درجه را تشخیص می‌دهند.



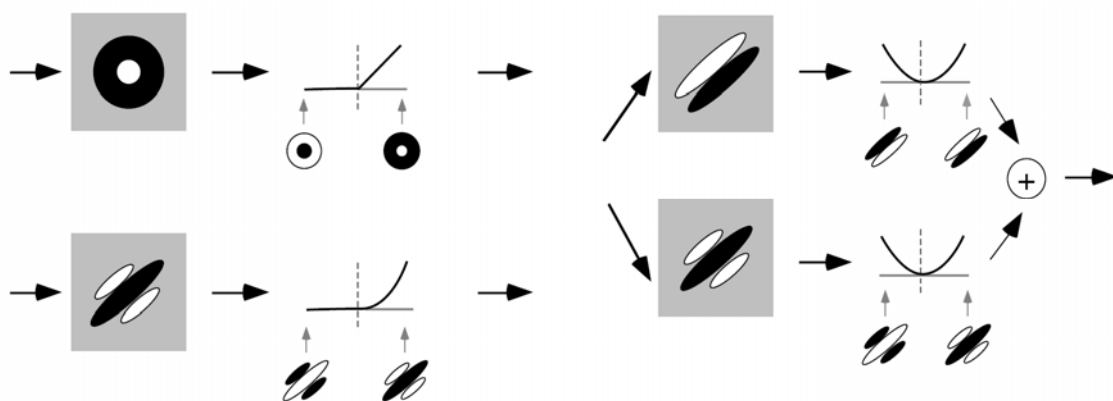
شکل ۲-۲۹: ترکیب قرار گرفتن نورون‌های V1 بر اساس جهت مورد نظر آنها (Kandel, Schwartz et al. 2001)

امروزه مشخص شده است که سلولهای ساده معمولاً الگوهای محلی راستادار را ترجیح می دهند و با یک لایه خطی و بدنبال آن یک لایه غیرخطی بخوبی مدل می شوند (شکل ۲-۳۱ پایین-چپ). محرکهایی که سلولهای ساده V1 را تحریک می کنند بخوبی توسط تابعهای دوبعدی گابور توصیف می شوند (Jones and Palmer 1987) (شکل ۲-۳۰). فرضیه ها در مورد سلولهای ساده، به دلیل پاسخ قوی آنها به میله ها و پنجره ها، آنها را کدگذاری کننده لبه ها در نظر می گیرند.



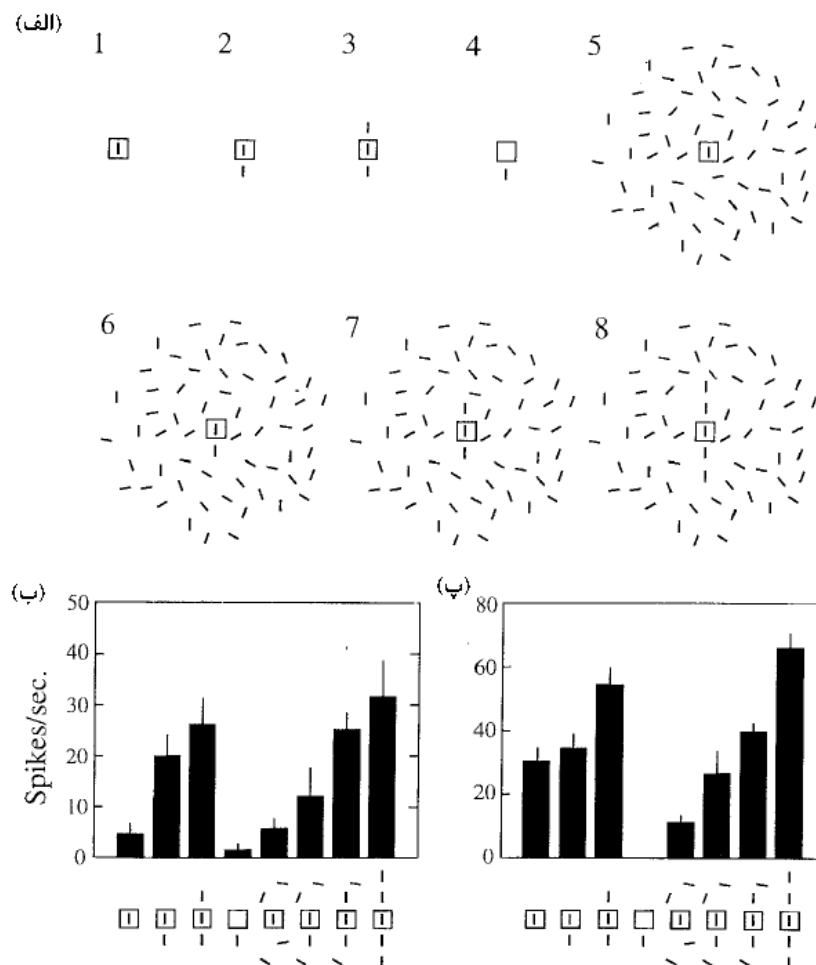
شکل ۲-۳۰: میدان گیرندگی سلولهای ساده V1 به خوبی توسط فیلترهای گابور توصیف می شود (Daugman 1989)

سلولهای پیچیده نیز دارای پاسخی قوی به میله ها و لبه ها هستند ولی به مکان دقیق لبه در میدان گیرندگی خود حساسیت کمتری دارند و معمولاً به لبه هایی با قطبیت مخالف نیز پاسخ یکسانی می دهند و هنگامی که با پنجره های سینوسی تحریک می شوند نسبت به فاز آن نامتغیرند. در مدل استاندارد سلولهای پیچیده که به مدل «انرژی» معروف است، دو ویژگی محلی و راستادار (معمولاً با 90° خروج از فاز تابع گابور) با تصویر کانوالو می شوند و خروجی آنها مربع شده و جمع می شود تا پاسخ نورون را بسازد که شکل ۲-۳۱ آنرا به خوبی نشان می دهد (Movshon, Thompson et al. 1978; Adelson and Bergen 1985; Heeger 1992; Heeger, Simoncelli et al. 1996).



شکل ۲-۳۱: مدل‌های استاندارد نورونهای بینایی اولیه (Carandini 2004): نورون LGN (بالا چپ)، نورون ساده V1 (پایین چپ)، و نورون پیچیده V1 (راست)

مطالعات فراوانی در مورد خصوصیات اتصالات افقی در ناحیه V1 قشر بینایی انجام شده است. دلیل این امر، اهمیت این اتصالات در میدان گیرندگی گسترش یافته^{۱۳۲} یک نورون می باشد. با گسترش یک محرک به خارج از میدان گیرندگی نورون، پاسخ آن نورون تحت تاثیر قرار می گیرد. در ناحیه V1 این امر به خاطر پدیده ای به نام جمع کردن کانتور^{۱۳۳} مورد توجه قرار گرفته است. نورونهای V1 به قطعات لبه حساس هستند. هنگامی که این لبه ها به خارج از میدان گیرندگی نورونهای V1 گسترش می یابند پاسخ نورون همچنان تغییر می کند و ممکن است تقویت و یا تضعیف شود. این پدیده به وجود اتصالات افقی بین نورونهای V1 نسبت داده شده است زیرا این اتصالات خصوصیتی از خود نشان داده اند که برای جمع کردن کانتور مناسب به نظر می رسد. شکل ۲-۳۲ اثر حضور محرک در میدان گیرندگی گسترش یافته بر پاسخ یک نورون V1 را نشان می دهد.

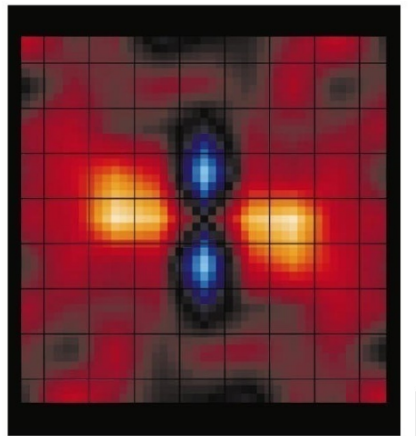


شکل ۲-۳۲: اثر حضور محرک ها در ناحیه دریافت گسترش یافته بر پاسخ نورون های V1. (الف) ترکیبات مختلف محرک ها در تصویر. مربع کوچک در هر شکل ناحیه دریافت اصلی نورون مورد نظر را نشان می دهد. (ب) پاسخ نورون به ورودیهای مختلف در (الف). هنگامی که محرکها تشکیل یک خط بدهند و در راستای یکدیگر باشند، پاسخ نورون تقویت می شود.

¹³² extended classical receptive field

¹³³ contour integration

شکل اتصالات افقی در نورونهای V1، منطبق با خصوصیات تصاویر موجود در طبیعت می باشد (Stettler, Das et al. 2002). در این ناحیه، اتصالات افقی به گونه ای گسترش یافته اند که نورونهای با جهت مشابه و نزدیک به هم را به هم متصل می کنند. هنگامی که محرک بهینه برای نورونهای مورد نظر در جهت یکدیگر باشند و تشکیل یک خط هموار بدهند، اتصالات افقی بین آنها وادارنده است. هنگامی که محرکها بر یکدیگر عمود باشند یا با هم زاویه تند و ناهموار داشته باشند این اتصالات بازدارنده هستند. شکل ۲-۳۳ اتصالات افقی مشاهده شده در لایه V1 را نشان می دهد.



شکل ۲-۳۳: ساختار اتصالات افقی در ناحیه V1 در قشر بینایی. در این تصویر نورون V1 در مرکز تصویر محرکی با جهت نشان داده شده در گوشه سمت راست تصویر را تشخیص می دهد. اتصالات وادارنده با رنگ آبی و اتصالات بازدارنده با رنگ زرد و قرمز نشان داده شده اند. در این تصویر، نورون های در راستای محرک نورون مرکزی دارای اتصال وادارنده با آن هستند.

۲-۴-۴- لایه V2

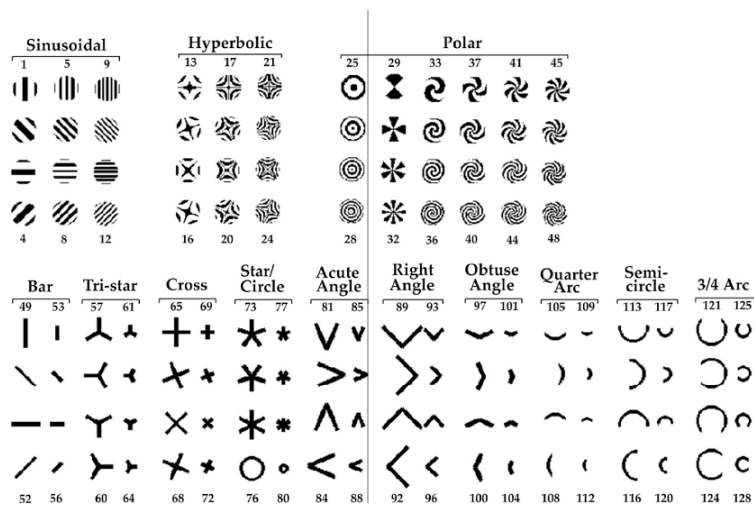
دومین ناحیه از سلسله مراتب پردازش در مسیر بطنی، ناحیه V2 می باشد. این ناحیه از نظر حجم تقریباً برابر با ناحیه V1 است و ورودی خود را از این ناحیه دریافت می کند. علیرغم اهمیت این ناحیه در پردازش بینایی، حجم مطالعاتی که در مورد این ناحیه انجام شده است بسیار کمتر از دیگر نواحی می باشد. با اینحال، در سال های اخیر مطالعات مناسبی در مورد این ناحیه انجام شده است که می تواند راهگشای بسیاری از پرسش ها باشد.

گزینندگی نورون های V2 پس از رشد اتصالات افقی در V1 توسعه می یابند. همچنین در هر لایه، اتصالات افقی پس از اتصالات ورودی از لایه پایین تر شکل می گیرند. بنابراین می توان گفت که نورون های V1 پس از یادگیری اتصالات افقی، افزونگی در تصاویر را در محدوده های کوچک حذف می کنند و حاصل را به لایه های بالاتر می فرستند. لایه V2 این ورودی را دریافت می کند و گزینندگی آن بر روی داده های خروجی V1 شکل می گیرد. سپس نورون های این لایه با برقراری اتصالات افقی بین یکدیگر، افزونگی در محدوده های بزرگتری را حذف می کنند. این مکانیزم در لایه های بالاتر تکرار می شود (Malmir and Shiry 2009).

با توجه به اینکه این ناحیه پس از V1 قرار گرفته است پیچیدگی خصوصیتی که نورون های این ناحیه تشخیص می دهند از نورون های V1 بیشتر است. این خصوصیت باعث شده است که نظریاتی در مورد اینکه نورون های V2 زاویه و

کانتور تشخیص می دهند شکل بگیرد (Hegd  and Van Essen 2000; Boynton and Hegd  2004; Hegd  and Van Essen 2007).

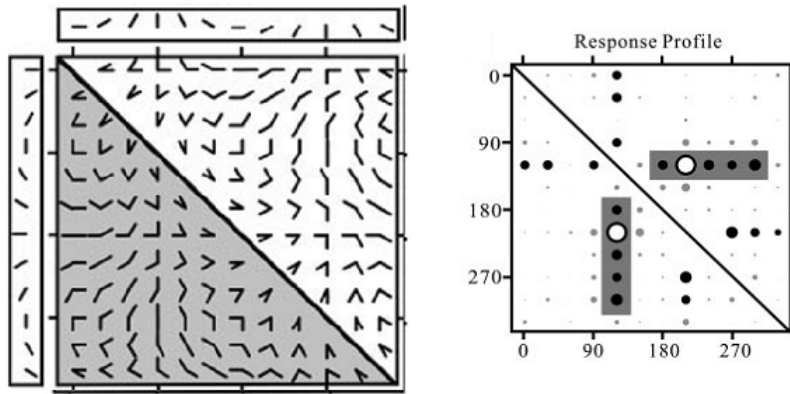
در تایید این ایده، آزمایشاتی انجام شده است که با انتخاب هوشمندانه مجموعه محرک ها، توانسته اند اطلاعات مناسبی در مورد خصوصیات نورون های V2 جمع آوری کنند. به عنوان مثال، در مرجع (Hegd  and Van Essen 2003) مجموعه ای از کانتورها و زوایا با جهت های مختلف برای تحریک نورون های V2 استفاده شده اند. این مطالعات با استفاده از تکنیک های ضبط تک نورون، توانسته اند پاسخ نورون های V2 را به مجموعه محرک ذکر شده ضبط کنند. مجموعه محرک در شکل ۲-۳۴ نشان داده شده است.



شکل ۲-۳۴: مجموعه محرک استفاده شده در (Hegd  and Van Essen 2003) برای آزمایش خصوصیات نورون های V2. این مجموعه شامل کانتور با زوایای مختلف و محرک های مشبک برای آزمایش تشخیص بافت می باشد.

در مجموعه ای از آزمایش ها که با این مجموعه محرک بر روی نورون های ناحیه V2 انجام شد این فرضیه تقویت شد که نورون های V2 به محرک های کانتور بیشتر از محرک های بافت پاسخ می دهند (Hegd  and Van Essen 2007; Hegd  and Van Essen 2003). در حقیقت بسیاری از نورون های V2 به این مجموعه محرک پاسخ می دادند و در میان این محرک ها، کانتور و زوایا در تحریک نورون های V2 نقش بیشتری داشتند. این مشاهدات مهر تأییدی بر این فرضیه بود که نورون های V2 در سلسله مراتب مسیر بطنی، خصوصیاتمانند زوایا و کانتورها را تشخیص می دهند.

در یکی از مطالعات جدید ناحیه V2 که بسیار مورد توجه قرار گرفته است، با استفاده از یک مجموعه زاویه با جهت های مختلف و زوایای مختلف توانست اطلاعات خوبی در مورد خصوصیات مورد نظر نورون های V2 به دست آورد (Ito and Komatsu 2004). در این مطالعه مشخص گردید که نورون های V2 به محرک های به شکل زاویه به خوبی پاسخ می دهند. بر اساس این مجموعه محرک، نورون های مورد نظر به سه دسته تقسیم شدند: نورون های که به یکی از خطوط تشکیل دهنده زاویه حساس هستند، نورون های که تنها به یک زاویه خاص در جهت های مختلف پاسخ می دهند و نورون هایی که به یک جهت خاص از زاویه حساس هستند. در شکل ۲-۳۵ این مجموعه محرک و پاسخ یک نورون به صورت نمونه نشان داده شده است.



شکل ۲-۳۵: مجموعه زوایای استفاده شده در (Ito and Komatsu 2004). در این مجموعه، زوایا با جهت های مختلف آورده شده اند. تصویر سمت راست پاسخ یک نورون به این مجموعه را نشان می دهد. این نورون به یکی از خطوط تشکیل دهنده زاویه حساس است.

۲-۴-۵- لایه V4

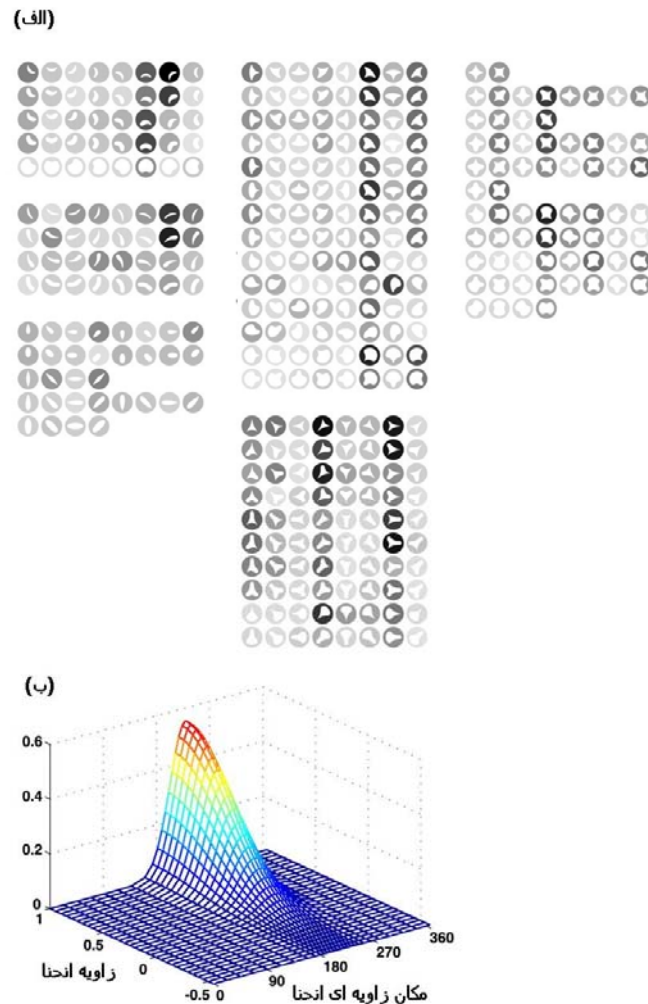
این ناحیه در مجاورت IT قرار دارد و نورون های آن خصوصیات نسبتاً پیچیده ای از تصاویر را تشخیص می دهند. یکی از مهمترین مطالعاتی که در مورد این ناحیه انجام شده است با استفاده از یک مجموعه محرک نشان داده شده در شکل ۲-۳۶ انجام پذیرفته است (Pasupathy and Connor 2001). در این آزمایش، هر محرک شامل یک منحنی بسته می باشد که کانتور آن در نقاط خاصی به صورت مقعر یا محدب طراحی شده است. هدف از این آزمایش تعیین این نکته بود که آیا نورون های V4 به قطعه خط های دور اشیا پاسخ مناسب می دهند. خطوط زاویه دار در این شکل شامل محدوده ای از زوایای محدب و مقعر می باشند که نسبت به مرکز شی دارای زاویه خاصی هستند. این مجموعه محرک به گونه ای انتخاب شده است که دو خصوصیت مهم را در محرک های خود نشان می دهد و می توان هر شکل را در فضای دو بعدی با یک نقطه نشان داد. یکی از این ابعاد انحنای کانتور^{۱۳۴} محرک می باشد و دیگری زاویه ای که انحنای کانتور در آن زاویه قرار دارد و نسبت به مرکز شی سنجیده می شود.

آنچه که در نتایج این آزمایش مشهود بود پاسخ هدفمند نورون های V4 به انحنا با زاویه های خاص بود. به عنوان مثال نورونی که در شکل ۲-۳۶ بالا نشان برای نمونه آورده شده به انحنای محدب در سمت راست مرکز شی پاسخ قوی می دهد. در این شکل، میزان تاریکی پس زمینه هر محرک نشان دهنده شدت پاسخ نورون به آن محرک می باشد. همانطور که در این شکل مشاهده می شود، انحنای محدب در سمت راست در تمام محرک هایی که پاسخ قوی داشته اند وجود دارد. نموداری که در شکل ۲-۳۶ پایین مشاهده می شود بیان دیگری از پاسخ این نورون می باشد. در این نمودار مشاهده می شود که بر اساس دو بعد زاویه انحنا و نقطه قرار گرفتن آن نسبت به مرکز شی، نورون در یک نقطه خاص حداکثر پاسخ را نشان داده است.

یکی از مدل های ارائه شده برای ناحیه V4، بر اساس مدل سلسله مراتبی HMax می باشد (Cadieu, Kouh et al. 2007). این مدل از لایه های مختلف نورون های ساده و پیچیده شکل گرفته که به طور متناوب در یک سلسله مراتب

¹³⁴curvature

سازمان یافته اند. این مدل در فصل بعد به تفصیل مورد بررسی قرار می گیرد. در این مدل، برای نورون های V4 از یک تابع گوسی استفاده شده است که همانند توابع پایه شعاعی (RBF) عمل می کند. نتایج آزمایشات، شباهت مدل ذکر شده و نورون های V4 را نشان می دهند.



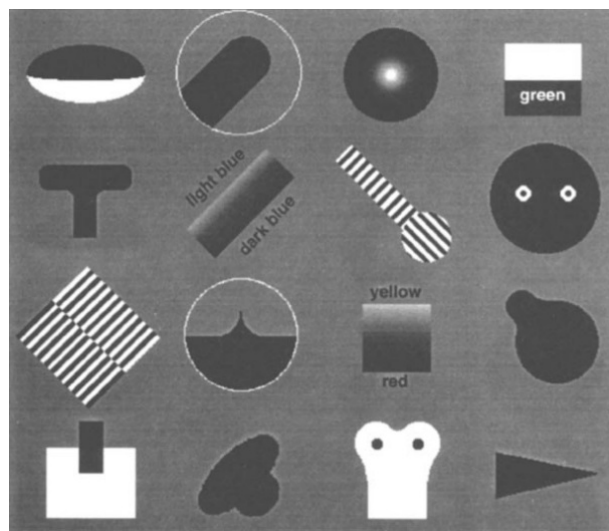
شکل ۲-۳۶: مجموعه محرک استفاده شده در (Pasupathy and Connor 2001) برای آزمایش نورون های V4 (الف) در این شکل میزان تاریکی پس زمینه نشان دهنده ی شدت پاسخ نورون به آن محرک می باشد. (ب) نموداری که میزان پاسخ نورون شکل الف را بر حسب انحنای زاویه و مکان انحناء نشان می دهد.

۲-۴-۶- لایه IT

ناحیه IT در قطعه مجانبی قشر مغز، آخرین ناحیه ای است که داده های بینایی را پردازش می کند بدون آنکه عوامل سطح بالا در این پردازش تاثیر داشته باشند. داده های بینایی پس از این ناحیه به قسمت های مختلفی از جمله قشر تداعی کننده در قطعه گیجگاهی و قشر پیش پیشانی^{۱۳۵} فرستاده می شوند (Miller and Cohen 2001). بنابراین ناحیه IT آخرین

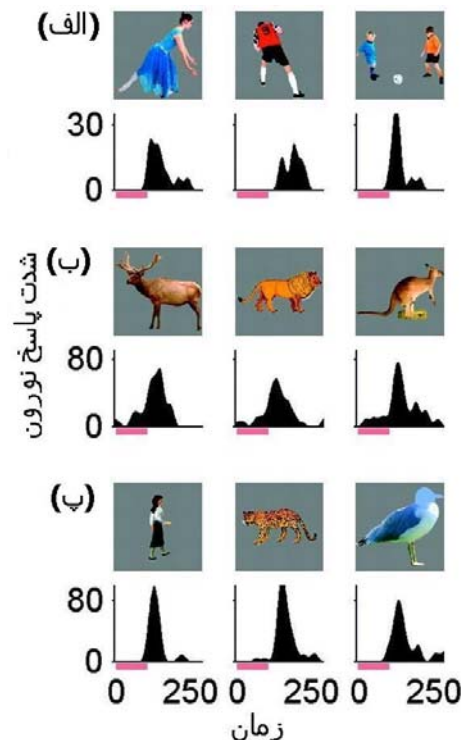
¹³⁵ prefrontal cortex

ناحیه ای است که داده های بینایی را پردازش می کند و تنها داده های حسی در عملکرد آن موثر هستند. نورون های این ناحیه همواره مورد توجه مطالعات عصب شناسی بوده و مطالعات فراوانی در مورد آنها انجام گرفته که در این بخش تنها به ذکر مهمترین موارد بسنده می شود. نکته مهم در مطالعه نورون های IT استفاده از مجموعه محرکی است که بتواند تمام فرضیه های احتمالی در مورد خصوصیات این نورون ها را شامل شود. در یکی از مطالعات مجموعه بسیار زیادی از اشیا گوناگون برای تحریک نورون های IT استفاده شد (Tanaka 1996). سپس هر شکلی که پاسخی قوی از نورون استخراج می کرد، در یک فرآیند ساده سازی به اجزای تشکیل دهنده خود تجزیه می شد. این ساده سازی تا هنگامی که پاسخ نورون افت قابل ملاحظه ای نداشت ادامه می یافت. بدین ترتیب، محرک هایی برای هر نورون پیدا می شد که پاسخ نسبتاً قوی در نورون ایجاد می کردند و تا حد امکان ساده بودند. مجموعه ای از این محرک ها در شکل ۲-۳۷ ملاحظه می شود. این شکل محرک هایی را نشان می دهد که بیشترین پاسخ را در نورون های IT ایجاد کرده اند.



شکل ۲-۳۷: مجموعه محرک هایی که در نورون های IT بیشترین پاسخ را ایجاد کردند (Tanaka 1996).

یکی از خصوصیات جالب در نورون های IT، حساسیت آنها به محرک های سطح بالا مانند چهره می باشد. بدین شکل که نورون های مجاور هم به یک شیء سطح بالا پاسخ یکسان می دهند. در مطالعه ایجاد مجموعه زیادی از تصاویر از دسته های مختلف اشیا به کار برده شدند و پاسخ نورون های IT به این محرک ها ضبط شد (Kiani, Esteky et al. 2007). نکته ای که در این آزمایش مشاهده گردید این بود که نورون های IT در دسته های مختلفی قرار می گیرند و هر دسته به نمونه های تصویر از یک دسته خاص از اشیا پاسخ بیشینه می دهند. می توان گفت که نورون های IT اشکال مختلف سطح بالا را به صورت کدهای دسته جمعی از پاسخ نورون ها نشان می دهند. نمونه ای از سه نورون حساس به دسته های مختلف در شکل ۲-۳۸ نشان داده شده اند.



شکل ۲-۳۸: سه نمونه از نورون های حساس به دسته اشیا (Kiani, Esteky et al. 2007) (الف) نورون حساس به شکل بدن انسان (ب) نورون حساس به شکل بدن چهارپایان (ج) نورون حساس به بدن انسان و چهارپایان و پرندگان. در هر شکل، پاسخ نورون بر حسب زمان و برای سه نمونه از دسته مورد نظر آورده شده است.

۲-۵- ویرگیهای آماری تصاویر طبیعی

۲-۵-۱- مقدمه

تصاویر طبیعی تصادفی نیستند، بلکه قاعده مندیهی آماری در آنها به وفور به چشم می خورد. اگر فرض کنیم که سیستم بینایی انسان برای کار روی تصاویر طبیعی سازگار شده باشد، این سیستم باید برای چنین قاعده مندیهی نیز بهینه شده باشد. بررسی های نظری در این راستا اخیراً منجر به درک عمیقی از خصوصیات پاسخ تصویری در قسمتهای اولیه از سیستم بینایی شده است. در ادامه به قاعده مندیهی آماری کشف شده در تصاویر طبیعی، میزانی که بینایی سطح پایین با آنها سازگار شده است، و پیشرفتهای جدیدی که مدلهای نظری در مدلسازی این رابطه داشته اند پرداخته می شود.

تصاویر طبیعی بسیار متغیرند. هنگامی که به بافت تصاویر طبیعی مانند پوست درخت نگاه می کنیم، علیرغم تنوع زیاد تصویری که بر روی شبکه می افتند بی درنگ آنرا می شناسیم. برای انجام چنین عملی، سیستم بینایی ما باید تمام انواع الگوهای طبیعی را بکاود تا بتواند تصویر را به عنوان نمونه ای از همان دسته تصاویر شناسایی کند. بنابراین بخش مهمی از بینایی شامل اخذ تنوع آماری تصاویر طبیعی می شود. در همین حال تصاویر طبیعی بسیار خاص نیز هستند. تفاوت آنها با تصاویر اتفاقی کاملاً محسوس است. توصیف قاعده مندیهی آماری آنها زمینه تحقیقاتی مهمی برای در علوم و مهندسی بوده

است. می توان هدف نهایی این تحقیقات را تولید مدلی احتمالاتی دانست که بتواند تصاویری تولید کند که از تصاویر طبیعی قابل تمییز نباشد. اگرچه همچنان راه درازی برای نیل به این مقصود باقی است، ولی دانش انسان درباره تصاویر طبیعی رو به فرونی است.

یک انگیزه مهم برای چنین تحقیقاتی فهم بهتر اینست که سیستم بینایی ما چگونه اطلاعات تصویری را کدگذاری می کند. بر طبق نظریه اطلاعات کدهایی که از نظر آماری کارتر هستند، همانهایی هستند که قاعده مندیهایی داده ها را به بهترین نحو اخذ می کنند. از آنجایی که سیستمهای بیولوژیکی زیر فشار تکاملی شدیدی هستند، می توان انگاشت که کدهایی که آنها استفاده می کنند بسیار کاراست. به علاوه، دانش درباره قاعده مندیهایی آماری داده ها، به سیستم اجازه انجام وظایف مهم بینایی مانند یافتن ویژگیهای جالب و یا پر کردن اطلاعات جا افتاده را می دهد. به طور کلی، سوالی که باید پاسخ داده شود اینست که: اهداف محاسباتی سیستم بینایی اولیه چیست؟ محدودیتهای محاسباتی بیولوژیکی آن کدامست؟ و تا چه اندازه خصوصیات پاسخ آن می تواند قاعده مندی آماری موجود در آنها را شرح دهد؟

در این بخش آنچه تاکنون درباره قاعده مندیهایی آماری تصاویر طبیعی کشف شده است و چگونگی توجیه رفتار بخش ابتدایی سیستم بینایی (بینایی سطح پایین) با توجه به آنها را کنار هم آورده می شود (مرورهای دیگری را نیز می توان در مقالات (Simoncelli and Olshausen 2001; Laughlin and Sejnowski 2003; Olshausen and Field 2005) یافت). ابتدا به قاعده مندیهایی آماری تک پیکسلها (مقادیر شدت روشنایی) در تصاویر طبیعی پرداخته می شود. سپس به قاعده مندیهایی آماری نواحی تصویر توجه می گردد. این نواحی در اکثر تحلیلها، به جای اینکه کل صحنه طبیعی را در بر بگیرند، به دلیل پیچیدگی محاسباتی به تکه های کوچک تصویر^{۱۳۶} محدود می شوند. با این وجود بدلیل اینکه اندازه میدان گیرندگی در سیستم بینایی نیز محدود است، این فرض آنچنان که در نگاه اول به نظر می آید محدود کننده نیست (Doi and Lewicki 2005).

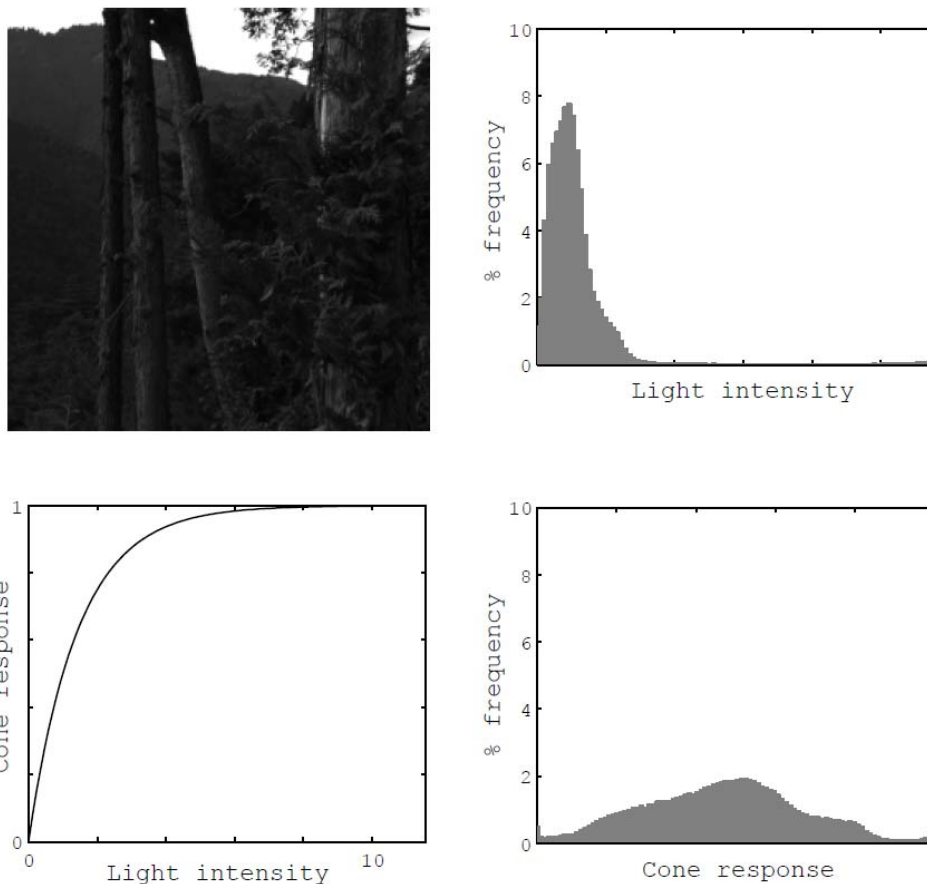
هنگامی که درباره مدل‌های نظری صحبت می شود باید دو نوع آنها را از یکدیگر تمیز داد. روش نخست معکوس مدل مولد داده هاست که در آن مدل احتمالاتی تولید داده ها دانسته فرض می شود و وظیفه سیستم بینایی استنتاج پارامترها و عوامل نهفته ایست که داده ها را تولید میکند (Dayan, Hinton et al. 1995; Olshausen 1996). مزیت این روش اینست که مفهومی خوش تعریف دارد و ارتباط واضحی با مدل احتمالاتی تصاویر طبیعی دارد. نوع دیگر مدل‌های بازنمایی هستند که هدف آنها استنتاج متغیرهای مخفی فرایند مولد نیست، بلکه برای ارضای قیودی برای بازنمایی مانند ناهمبسته سازی یا انتقال بیشینه اطلاعات طراحی شده اند (Doi and Lewicki 2005).

¹³⁶ image patches

۲-۵-۲- آمارگان تک پیکسلی

۲-۵-۲-۱- قاعده مندیه‌ای آماری

شدت روشنایی تصاویر طبیعی بسیار به مقادیر کوچک متمایل است (شکل ۲-۳۹ بالا-راست). این تحلیل بر مبنای تصاویری بدست آمده است که مقدار پیکسل آنها به طور خطی با شدت روشنایی مرتبط است (Ruderman, Cronin et al. 1998; Van Hateren and Van der Schaaf 1998; Doi, Inui et al. 2003) و شبیه شکل ۲-۳۹ بالا-چپ است. توجه شود که چنین داده‌هایی به عنوان ورودی برای چشم مناسب هستند در حالی که تصاویری که با دوربینهای معمولی فیلم/CCD گرفته می‌شوند شامل تبدیلات غیر خطی است و برای تحلیل مناسب نیستند.



شکل ۲-۳۹: شدت روشنایی خطی صحنه‌های طبیعی به سمت مقادیر کوچک متمایل است (بالا چپ) مثالی از شدت روشنایی خطی تصاویر طبیعی (بالا راست) هیستوگرام این تصویر (پایین چپ) مدل غیرخطی سلولهای مخروطی (پایین راست) هیستوگرام پاسخ غیرخطی سلولهای مخروطی به این تصویر. تصویر مربوطه در شکل ۲-۴۳ خواهد آمد.

۲-۵-۲-۲- ارتباط با خواص پاسخ

بسیاری از پژوهشگران فرض کرده اند که تابع لگاریتم می تواند غیرخطی بودن دریافت کننده های نوری مخروطی را مدل کند (Ruderman, Cronin et al. 1998; Van Hateren and Van der Schaaf 1998). تابع لگاریتم داده ها را طوری تبدیل می کند که متغیر تبدیل شده توزیع یکنواخت تری داشته باشد و از اینرو کنتراست تصویر افزایش می یابد. تحلیل دقیقتر نشان داده است که مدل غیرخطی بودن گیرنده های مخروطی با هیستوگرام تجمعی تصاویر طبیعی با شدت روشنایی خطی به خوبی سازگار است که به طور ضمنی اشاره به این دارد که غیرخطی بودن گیرنده های مخروطی به مثابه هموارسازی هیستوگرام عمل می کند (Doi, Inui et al. 2003). مدل تجربی غیرخطی بودن گیرنده های مخروطی از رابطه زیر بدست می آید:

$$x_{nl} = 1 - \exp(-kx_l) \quad (2-42)$$

که در آن x_l شدت روشنایی خطی، x_{nl} پاسخ غیرخطی گیرنده های مخروطی، و k پارامتری آزاد است که با سطح روشنایی^{۱۳۷} تطبیق می یابد (Baylor, Nunn et al. 1987). شکل رابطه ۲-۴۲ در شکل ۲-۳۹ پایین-چپ آمده است که پارامتر آن برای تصویر شکل ۲-۳۹ بالا-چپ تطبیق یافته است. هیستوگرام پاسخ غیرخطی گیرنده های مخروطی حاصل که در شکل ۲-۳۹ پایین-راست آمده است بیانگر اینست که مقادیر پیکسل نسبت به مقادیر خام شدت روشنایی توزیع بسیار یکنواخت تری دارند (تصویر شکل ۲-۳۹ بالا-چپ وقتی با پاسخ گیرنده های مخروطی تبدیل می شوند دارای کنتراست بهبود یافته ای می شوند، شکل ۲-۴۳). شیب مدل تجربی که از معادله ۲-۴۲ بدست می آید معمولاً اندکی کمتر از هیستوگرام تجمعی تصاویر طبیعی با شدت روشنایی خطی است که باعث می شود هیستوگرام حاصل اندکی با توزیع یکنواخت شکل ۲-۳۹ متفاوت باشد. یک مدل که خطای تخمین ورودی را با داشتن نویز ذاتی (که به آن نویز کانال اطلاق می شود) کمینه می کند در مقاله (Von der Twer and MacLeod 2001) آمده است و ادعا شده است که سازگاری با مدل هموارسازی هیستوگرام را بهبود داده است. مدل دیگری از روشهای تطبیق پذیری برای سریهای زمانی داده های شدت روشنایی بهره جسته است و مدل جامعتری را فراهم کرده است (Van Hateren and Snippe 2001).

۲-۵-۲-۳- ملاحظات

از آنجا که توزیع یکنواخت با داشتن بازه ثابتی برای بازنمایی بیشترین مقدار اطلاعات را منتقل می کند، خاصیت غیرخطی گیرنده های مخروطی به صورت کاراترین تبدیلی که از ظرفیت اطلاعاتی بازنمایی استفاده می کند تلقی می شود. به طور کلی این هدف - پیدا کردن تبدیلی از ورودی که بیشترین مقدار اطلاعات را منتقل کند - به نام کدگذاری بهینه (یا کاهش افزونگی) شناخته می شود (Barlow 1961; Atick 1992). این اصل زیربنایی روشهای نظری ایست که در این پژوهش بررسی می شود. توجه شود که این بهینگی به یک قید محدود است: اگر واریانس بجای بازه ثابت باشد، توزیع گوسی بهینه است و اگر متغیر مثبت باشد و میانگین آن ثابت باشد توزیع نمایی بهینه است (Simoncelli and Olshausen 2001). بنا به ادعای (Doi and Lewicki 2005) تا کنون توجه موثقی برای اینکه کدام قید برای سیستم بینایی اولیه مناسب

¹³⁷Luminance

است آورده نشده است، ولی بهرحال این نتایج نشان می دهد که غیرخطی بودن گیرنده های نوری را می توان به خوبی با داشتن بازه ثابتی از خروجی های نورونی نشان داد.

۲-۵-۳- آمارگان مرتبه دوم

برای تحلیل آمارگان نواحی تصویری، ابتدا به آمارگان مرتبه دوم پرداخته می شود. در قسمت بعدی به آمارگان مراتب بالاتر پرداخته می شود. در تحلیلهای آتی مقدار میانگین (همان آماره مرتبه اول) برابر صفر فرض می شود که این فرض مانند اینست که فرض شود که اطلاعات میانگین سطح روشنایی در بازنمایی حذف شود و سیستم بینایی تنها به تغییرات حول آن علاقه مند باشد.

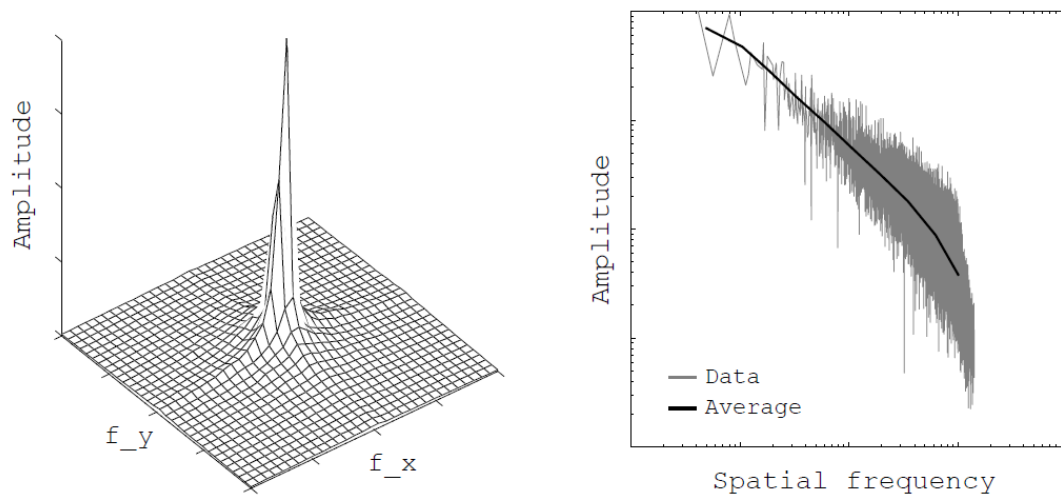
تعریف آمارگان مرتبه k ام عبارتست از $\langle \prod_k x_k \rangle$ که در آن $\langle \cdot \rangle$ عملگر میانگین گیری کلی روی نمونه ها را نشان می دهد و x_k مقدار پیکسل در مختصات k است. با فرض نامتغیری به ازای جابجایی در تصاویر طبیعی، میانگین کلی می تواند با میانگین گیری روی اندیس مختصات با حفظ موقعیت نسبی اندیسها جایگزین شود. آمارگان مرتبه اول برابر میانگین پیکسلهاست و در تمام مراتب دیگر آمارگان نیز دخیل است. آمارگان مرتبه دوم برابر کواریانس زوج پیکسلهاست که با ماتریس کواریانس بیان می شود. بقیه قاعده مندیهایی آماری را مجموعاً آمارگان مراتب بالا می نامند.

۲-۵-۳-۱- قاعده مندیهایی آماری

آمارگان مرتبه دوم کاملاً با طیف بزرگی تصاویر طبیعی توصیف می شوند. دلیل این امر اینست که کواریانس آنها همان خود-همبستگی تصاویر طبیعی است و تبدیل فوریه یک تابع خود-همبستگی با طیف توانی (= مربع طیف بزرگی) تصاویر طبیعی بدست می دهد. طیف بزرگی تصاویر طبیعی تقریباً متناسب با $1/f$ است که f فرکانس مکانی است (Field 1987). در شکل ۲-۴۰ مثالی از طیف بزرگی تصاویر طبیعی آورده شده است. می توان مشاهده کرد که بزرگی در طول راستاهای افقی و عمودی اندکی قویتر است که این به دلیل غالب بودن ساختارهای افقی و عمودی در صحنه های طبیعی است (مثلاً درختان و افقها). طیف بزرگی $1/f$ به معنی ساختار همبستگی مقادیر پیکسلی در تصاویر طبیعی است: همبستگی برای مؤلفه های فرکانس مکانی پایینتر قویتر است. در جایی که همبستگی در تصویر وجود نداشته باشد، طیف بزرگی مسطح خواهد بود.

علاوه بر تمایل کلی به طیف با بزرگی $1/f$ ، تصاویر رسته های گوناگون طبیعی دارای سبک خود در طیف بزرگی هستند (Torralla and Oliva 2003). به عنوان مثال تصاویر جنگلها دارای مؤلفه های مکانی-فرکانسی بالای زیادی هستند ولی تصاویر ساحل دریا بزرگی بیشتری را در فرکانسهای پایینتر و افقی تجربه می کنند که این مسئله برای رسته های زیادی آزمایش شده است. به طور مشابه عمق تصاویر طبیعی (یعنی نزدیکی و دوری آنها) نیز توسط طیف بزرگی تعیین می شود اگر تصاویر از نزدیک گرفته شده باشند، بزرگی در راستاها یکرخت تر و دارای مؤلفه های مکانی فرکانسی بالایی است، که با راستای مقید نشده دید نسبت به اشیا و جزئیات ریز در صحنه قابل توجه است. همینطور اگر تصویر از دور گرفته شده باشد، بزرگی در راستای محورهای عمودی و افقی قویتر است و در فرکانسهای مکانی پایینتر متمرکز است که با غالب بودن

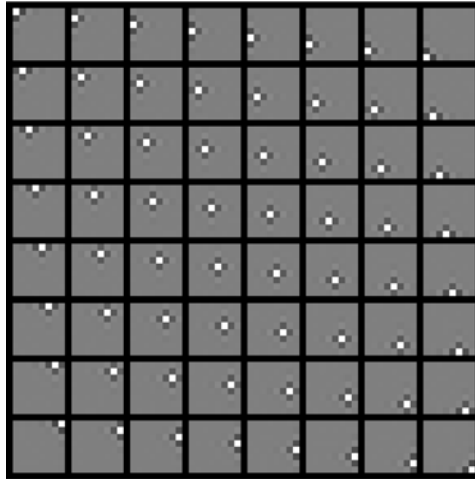
ساختارهای افقی و عمودی در یک نمای دور و افزایش ساختارهای بزرگ و هموار مانند زمینها، جنگلها و آسمان قابل توجیه است.



شکل ۲-۴۰: طیف بزرگی تصاویر طبیعی متناسب با $1/f$ است. داده های تصویری تحلیل شده در شکل ۲-۳۹ نشان داده شده است. (چپ) بزرگی فرکانس مکانی دوبعدی. (راست) بزرگی فرکانس مکانی که از تجمیع بزرگی دوبعدی روی تمام راستها بدست آمده است. منحنی خاکستری داده های تمام فرکانسهاست و نسخه هموار شده آن (میانگین آن) با منحنی سیاه مشخص شده است.

۲-۵-۳-۲- ارتباط با خواص پاسخ

هدف سیستم بینایی اولیه، حذف افزونگی از بازنمایی پیکسلهای (یا گیرنده های نوری) بسیار همبسته و تبدیل آن به یک بازنمایی جدید است که در آن فعالیت نورونی دارای استقلال آماری نسبی (بیشتری) باشد و بنابراین کارتر باشد (Barlow 1961; Atick 1992). کاهش افزونگی در قالب آمارگان مرتبه دوم به معنی تبدیلی است که متغیر را دارای واریانس واحد و ناهمبسته با بقیه کند. چنین تبدیلی وایتنینگ نامیده می شود و با یک مجموعه از فیلترهای خطی قابل پیاده سازی است. پس از اعمال وایتنینگ به تکه تصاویر طبیعی یک ساختار فیلتری حول مرکز بدست می دهد که توافق زیادی به میدان گیرندگی سلولهای عقده ای شبکه (RGCها) و LGN دارد (Atick and Redlich 1990; Atick 1992; Atick and Redlich 1993; Bell and Sejnowski 1997). تعداد نامحدودی فیلتر وایتنینگ وجود دارد. اگر W_1 یک ماتریس فیلتر وایتنینگ باشد، ماتریس کواریانس خروجی آن باید برابر ماتریس همانی باشد، $W_{1x} W_{1x}^T = I$ ، که به طور ضمنی اشاره دارد که $W_1 \cdot \Sigma_x \cdot W_1^T = I$. حال با هر ماتریس متعامد دلخواه U ، مجموعه جدیدی از فیلترها مانند $W_2 = U \cdot W_1$ فرض می شود. می توان دید که ماتریس حاصل نیز یک ماتریس وایتنینگ است زیرا رابطه $W_2 \cdot \Sigma_x \cdot W_2^T = I$ را ارضا می کند. ماتریس وایتنینگی که ساختار حول مرکز را تولید می کند همانی است که توسط $\Sigma_x^{-1/2}$ نشان داده می شود (Atick and Redlich 1993; Bell and Sejnowski 1997) (شکل ۲-۴۱).



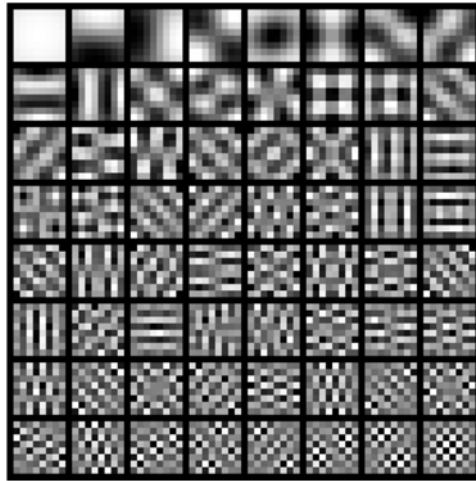
شکل ۲-۴۱: فیلترهای وایتنینگ برای تصاویر طبیعی سازماندهی میدان گیرندگی حول مرکز را از خود نشان می دهد. هر تصویر کوچک با یک فیلتر وایتنینگ برای تکه تصویرهای 8×8 متناظر است. در اینجا تمام مجموعه ۶۴ تایی آن به نمایش در آمده است.

۲-۵-۳-۳- ملاحظات

مدلی که در (Atick and Redlich 1990) و (Atick and Redlich 1992) پیشنهاد شده است مؤلفه نویز ورودی (معروف به نویز حسگری) و نویز فرایند فیلتر کردن پیش از وایتنینگ را نیز در خود محسوب می کند. با فرض اینکه واریانس نویز متناسب با میانگین سطح روشنایی باشد، مدل به تطبیق چشمگیری با میدانهای گیرندگی RCGها دست می یابد که شکل خود را بر اساس میانگین سطح روشنایی سازگار می کنند.

وایتنینگ عملیاتی است که قاعده مندیهایی آماری مرتبه دوم را از بین می برد. به عنوان یک ایده می توان از این خاصیت استفاده کرد و به جای حذف این قاعده مندیهایی از آنها استفاده کرد. اطلاعات رسته که در طیف بزرگی قرار دارد - چنانکه در بالا به آن اشاره شد - نیز می تواند استفاده شود ولی مقایسه مستقیمی از آنها در دست نیست (Torralba and Oliva 2003).

راه دیگر نیز کاهش ابعاد است. تحلیل مؤلفه اصلی (PCA) یک مجموعه بهینه از بردارهای پایه متعامد برای کاهش بُعد نشان می دهد و بازنمایی جدید کمترین خطای مربعی را در بازسازی دارد. شکل ۲-۴۲ یک مجموعه از فیلترهای PCA را نشان می دهد که الگوهای فرکانسی را بدون هیچ دخالت محلی نشان می دهد و بنابراین هیچ همبستگی آشکاری به سازماندهی میدانهای گیرندگی سیستم بینایی اولیه ندارد. با این حال PCA ممکن است قرابتی ساختاری با سیستم بینایی انسان داشته باشد. واریانس داده در زیر فضایی محدود متشکل از تعداد اندکی از مؤلفه های اصلی متمرکز شده است (مثلاً حدود ۲۰ تا ۳۰٪ مؤلفه های اصلی ۹۰٪ واریانس داده های تکه های تصاویر طبیعی را توصیف می کند). این ورودی تصویری دارای ابعاد بسیار زیادی است (پنج میلیون سلول گیرنده نوری مخروطی در هر شبکیه) و بنابراین ممکن است سیستم بینایی از آمارگان مرتبه دوم برای کاهش ابعاد استفاده کند.



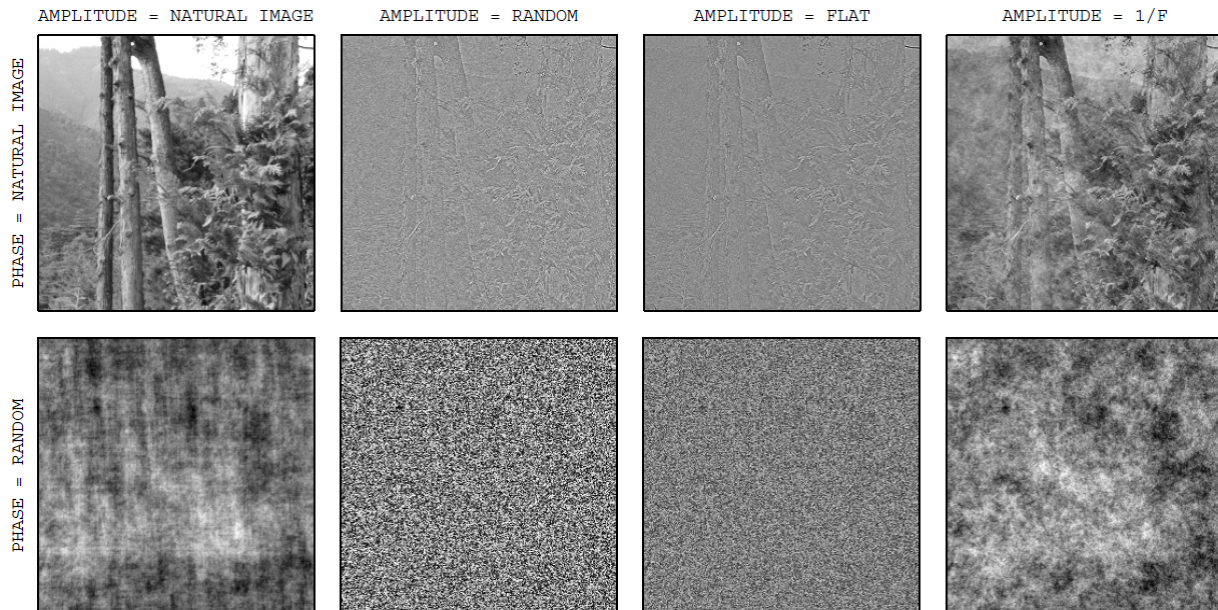
شکل ۲-۴۲: فیلترهای PCA تصاویر طبیعی محلی نیستند. این یک مجموعه از مؤلفه های اصلی برای تکه تصویرهای 8×8 پیکسل تصاویر طبیعی است که به ترتیب نزولی بزرگی از بالا به پایین و چپ به راست مرتب شده است.

۲-۵-۴- آمارگان مرتبه بالاتر

۲-۵-۴-۱- قاعده مندیهایی آماری

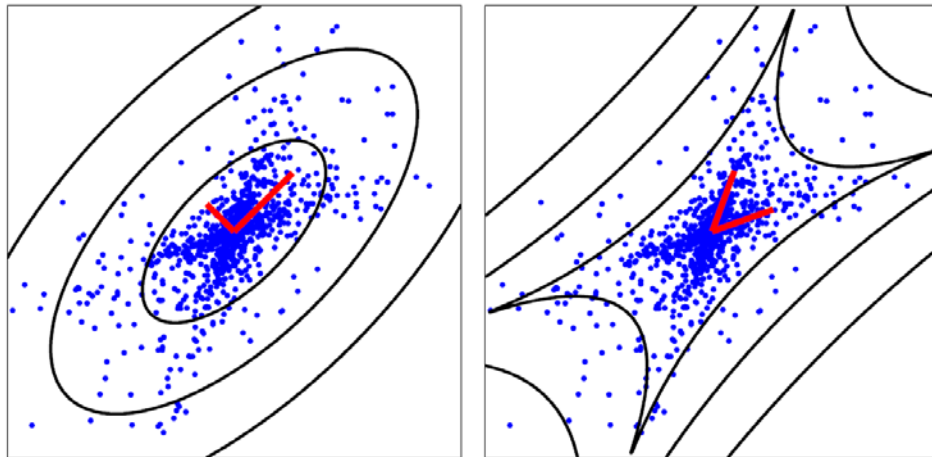
آمارگان مراتب بالاتر تصاویر طبیعی در طیف فاز تبدیل فوریه آنها مقیم هستند که مکمل طیف بزرگی این تصاویر ($=$ آمارگان مرتبه دوم) هستند. آنگونه که اشاره شد، طیف فاز دارای اطلاعاتی ایست که از لحاظ مفهومی مرتبط ترند زیرا اطلاعاتی در مورد لبه ها بدست می دهند: در مکان لبه ها فازهای فرکانسهای مختلف همراستا می شوند. این مسئله را می توان با بهم ریختن فاز یک تصویر طبیعی در عین حفظ طیف بزرگی آن مشاهده کرد (شکل ۲-۴۳). به راحتی می توان دید که کدام مؤلفه های فرکانسی با چه توانی ظاهر شده است (می توان مؤلفه فرکانس مکانی پایین و افقی متناظر با الگوی کنتراست برجسته بین آسمان و کوهها را مثال زد). بهرحال موقعیت آنها ($=$ فاز) هیچ ربطی به تصویر اصلی ندارد و به دلیل فقدان همراستایی فاز، الگوی لبه ماندی مشاهده نمی شود. به طریق مشابه تصاویری با بزرگی های ثابت یکنواخت، ثابت I/f و تصادفی با حفظ طیف فاز تصاویر طبیعی ساخته می شود: این تصاویر قابل شناسایی هستند و می توان تشخیص داد که لبه ها کجا قرار دارند. در عین شگفتی، تصویری که طیف فرکانسی اصلی را حفظ کرده است و طیف بزرگی آن برابر I/f است تصویری بسیار نزدیک به تصویر اصلی می دهد (شکل ۲-۴۳ بالا-راست) که نشان می دهد که اطلاعات طیف بزرگی نقش کمزنگتری در ادراک انسان دارد.

طیف بزرگی ماتریس کواریانس داده ها را شکل می دهد و بنابراین می تواند تمام قاعده مندیهایی داده ها را اگر از یک توزیع گوسی (با میانگین صفر) باشند اخذ کند. در مورد تصاویر طبیعی مشاهدات زیادی از رفتار بسیار غیرگوسی توزیعها صورت گرفته است (Field 1987; Field 1994). در شکل ۲-۴۴ یک نمونه از داده های دوبعدی غیرگوسی نشان داده شده است: فرض می شود که داده ها گوسی است و بررسی طیف بزرگی مدلی بدست می دهد که به وضوح در اخذ توزیع زیرین داده ها ناتوان است (شکل ۲-۴۴ چپ). برای مقایسه توزیع واقعی داده ها در شکل ۲-۴۴ راست ترسیم شده است که همان چیزی است که باید از داده ها تخمین زده شود.



شکل ۲-۴۳: طیف فاز نسبت به طیف بزرگی در ادراک ما سهم بیشتری دارد. هر تصویر از تلفیق طیفهای فاز و بزرگی ساخته می شود. از هر تصویر طبیعی یا تصادفی طیف فاز گرفته شده و در سطر پایین ترسیم می شود. از هر تصویر طبیعی یا تصادفی طیف بزرگی با بزرگی $1/f$ یا یکنواخت می گیریم. تناظر تصویر {فاز، بزرگی} = {تصویر طبیعی، تصویر طبیعی} مربوط به تصویر طبیعی اصلی است. تصویر {تصادفی، تصادفی} نیز تصویر اصلی تصادفی است و تصویر {تصادفی، $1/f$ } تصویر همبسته تصادفی است که طیف بزرگی آن $1/f$ است. توجه شود که تصویر {طبیعی، یکنواخت} نسخه وایتنینگ شده تصویر طبیعی است که معادل مدل RGC/LGN است.

یکی از شواهد غیرگوسی بودن تصاویر طبیعی پراکنده بودن توزیع آنهاست که به راحتی با کورتوسیس (نرمال شده) اندازه گیری می شود: $3 - \frac{\langle x^4 \rangle}{\sigma_x^2} = \tilde{\kappa}(x)$ که کورتوسیس آماره ای از مرتبه چهارم است. توزیعی پراکنده تر از گوسی است که کورتوسیس آن مثبت باشد ($\tilde{\kappa} > 0$). جهتایی در فضای تصویر وجود دارد که نگاشت خطی نقاط نمونه روی آنها توزیعی پراکنده دارد (Field 1987; Field 1994). توجه شود که چنین توزیع غیرگوسی اگر داده ها گوسی چند متغیره باشند امکانپذیر نیست زیرا هر نگاشت خطی از یک گوسی چندمتغیره گوسی است. جالب توجه است که نگاشت خطی روی تابع گابور - که به عنوان مدل تجربی سلولهای ساده V1 معرفی شده است (Daugman 1985; Jones and Palmer 1987) - یک توزیع پراکنده بدست می دهد که این ایده را القا می کند که سیستم بینایی از قاعده مندیهایی مراتب بالاتر استفاده می کند.



شکل ۲-۴۴: آمارگان مرتبه دوم برای مشخص کردن داده های غیرگوسی کافی نیستند. نقطه های آبی در هر دو قسمت از یک توزیع نامتعاد پراکنده نمونه برداری شده است. نمودارهای کانتور تطبیق گوسی به داده ها (چپ) و توزیع واقعی داده ها (راست) را نشان می دهد. میله های قرمز محورهای اصلی توزیع گوسی یا تابع پایه توزیع غیرگوسی آنها را نشان می دهد.

۲-۴-۵-۲- ارتباط با خواص پاسخ

مدلهای نظری کدگذاری پراکنده و تحلیل مؤلفه مستقل (ICA) نشان داده اند که با مدل کردن آمارگان مرتبه بالا و غیرگوسی تصاویر طبیعی امکان توجیه ساختار محلی و راستدار میدان گیرندگی سلولهای ساده وجود دارد (Olshausen, 1996; Bell and Sejnowski 1997; Hyvärinen, Hoyer et al. 2001). در هر دو مدل فرض می شود نقطه داده x توسط تجمیع تابعهای پایه \vec{a}_i که با متغیرهای مخفی s_i وزن دهی شده اند بوجود آمده اند:

$$\vec{x} = \sum_{i=1}^m s_i \vec{a}_i = A \vec{s} \quad (۴۳-۲)$$

$$A = [\vec{a}_1, \dots, \vec{a}_m] \text{ و } \vec{s} = (s_1, \dots, s_m)^T \text{ که}$$

مجموعه توابع پایه A چنان سازگار می شود که درستنمایی داده بیشینه شود:

$$L = \langle p(\vec{x} | A) \rangle \quad (۴۴-۲)$$

که احتمال داده A با بدست آوردن توزیع حاشیه ای متغیرها بدست می آید:

$$p(\vec{x} | A) = \int p(\vec{x} | A, \vec{s}) p(\vec{s}) d\vec{s} \quad (۴۵-۲)$$

این مدلها فرض می کند که اولاً متغیرهای مخفی مستقل آماری هستند و ثانیاً هر متغیر مخفی از یک توزیع پراکنده مانند لاپلاسیین گرفته می شود و تابع توزیع احتمال اولیه توأم را به شکل زیر بدست می دهد:

$$p(\vec{s}) = \prod_{i=1}^m p(s_i) \propto \prod_{i=1}^m \exp(-|s_i|) \quad (46-2)$$

مقدار بهینه A با استفاده از تپه نوردی صعودی برای یافتن بیشینه معادله ۲-۴۴ محاسبه می شود. فرض این مدل اینست که هدف محاسباتی سیستم بینایی استنتاج محتملترین بازنمایی محرک (که احتمالاً دارای نویز است) است که به صورت زیر فرموله می شود:

$$\hat{s} = \arg \max_s p(\vec{s} | \vec{x}, A) \quad (47-2)$$

متغیر \hat{s} متناظر با پاسخ نورونی محرک x است. فرض ICA اینست که A مربعی و معکوس پذیر است و بر اساس آن:

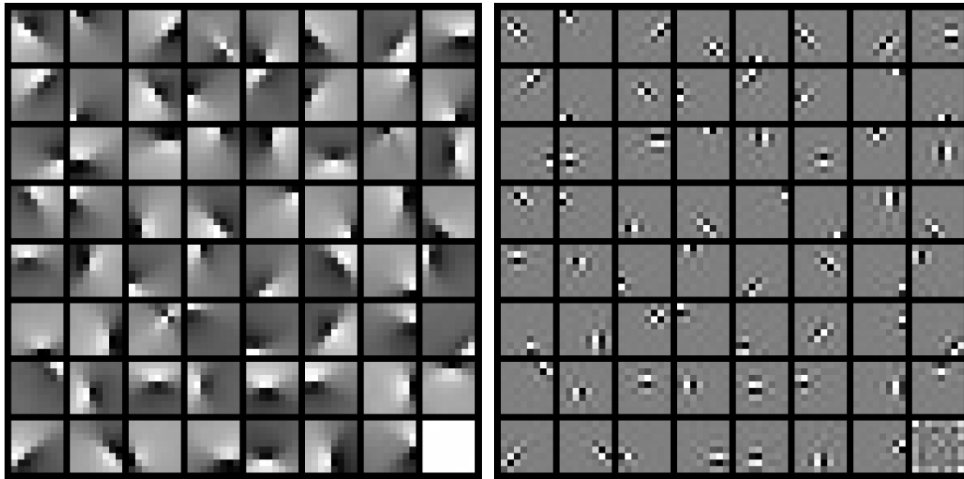
$$\hat{s} = A^{-1} \vec{x} = W \vec{x} \quad (48-2)$$

که \hat{s} تبدیل شده خطی x است (پس $P(s|x, A) = \delta(s - A^{-1}x)$). از سوی دیگر در کدگذاری پراکنده، نگاشت به این حالت خاص محدود نیستو باید معادله ۲-۴۷ که به شکل زیر نوشته می شود، حل شود:

$$\hat{s} = \arg \max_s \ln p(\vec{x} | \vec{s}, A) + \ln p(\vec{s}) = \arg \min_s \frac{1}{2\sigma^2} \|\vec{x} - A\vec{s}\|^2 + \sum_{i=1}^m |s_i| \quad (49-2)$$

که در آن σ^2 سطح خطای بازسازی را تعریف می کند. این بدین معنی است که \hat{s} با کمینه کردن توأم خطای مربعی بازسازی (جمله اول) و درستنمایی منفی تابع احتمال اولیه لاپلاسی (جمله دوم) بدست آید.

در شکل ۲-۴۵ پایه های ICA (A) و توابع فیلتر (W) برای تکه های تصویری طبیعی 8×8 نمایش داده شده است (و نتیجه کدگذاری پراکنده نیز مشابه است). توابع فیلتر با میدانهای گیرندگی ای که با روش همبستگی معکوس نگاشت شده اند متناظرند، و می توان دید که همه فیلترها (بجز فیلتری که مربوط به مؤلفه DC است) دارای سازماندهی راستادار و محلی است که شبیه میدانهای گیرندگی سلولهای ساده V1 است. توابع پایه با الگوهای که متغیرهای مخفی کدگذاری می کند متناظر است و به الگوهای لبه شبیه است (بجز مؤلفه DC). باید توجه کرد که توابع پایه ICA نیز نامتعامل است. اگر توابع پایه متعامد بودند با توجه به تعریف با فیلترها یکسان بودند (مانند حالتی که در PCA وجود دارد و می توان آنرا به چشم از روی شکل ۲-۴۵ تایید کرد). این نتایج نشان می دهد که میدانهای گیرندگی مشابه سلولهای ساده را می توان با جستجو به دنبال بازنمایی کارا بدست آورد، و اینکه فعالیت نورونی آنها احتمالاً صحنه های طبیعی را به صورت پراکنده و مستقل آماری کدگذاری می کند.



شکل ۲-۴۵: روش ICA فیلترهایی محلی و راستادار مشابه میدان گیرندگی سلولهای ساده V1 تولید می کند. این تصویر مجموعه توابع پایه (چپ) و فیلتر (راست) برای تکه های 8×8 پیکسلی است. توابع فیلتر و پایه به یکدیگر متناظرند و عنصری که مربوط به مؤلفه DC است در گوشه پایین سمت راست قرار دارد.

۲-۵-۴-۳- ملاحظات

قاعده مندیهایی آماری تکه های تصاویر طبیعی که تاکنون به آن اشاره شده است عبارتست از

- همبستگیهای مرتبه دوم آشکار؛
- وجود جهت‌های پراکندگی؛
- و اینکه متعامد نیستند.

توجه شود که توزیعی که در شکل ۲-۴۴ راست دیده می شود یک نمونه دوبعدی از چنین توزیعی است. مدلهایی مانند ICA و کدگذاری پراکنده، کدهای کارا را تشکیل می دهند ولی برخلاف وایتینگ (که تنها همبستگیها را از بین می برد) این روشها افزونگی مرتبه بالاتر را نیز از بین می برد. می توان به طور خلاصه به این نتیجه رسید که مراحل LGN/ها RCG افزونگی مرتبه دوم را از بین می برد و مرحله V1 مابقی این افزونگی ها را حذف می کند. همینطور باید توجه داشت که استقلال این متغیرهای مخفی تنها در محدوده مدل بدست می آید و وابستگیهای مراتب بالاتر حتماً باقی می ماند که در بخش بعد بررسی می شود.

۲-۵-۵-۵- توسعه های بیشتر

۲-۵-۵-۱- مدل‌های غیر خطی

با وجود آنکه ICA به دنبال بازنمایی مستقل آماری می گردد، اعمال آن روی تصاویر طبیعی از خود وابستگیهای باقیمانده را نشان می دهد (Hyvärinen and Hoyer 2001; Karklin and Lewicki 2003). به عنوان مثال واریانس ضرایب توابع پایه دارای وابستگی هستند: هنگامی که یک ضریب دارای بزرگی زیادی است، سایر ضرایب نیز تمایل به اخذ بزرگیهای

زیاد از خود نشان می دهند. این امر نشان می دهد که داده های نمی توانند با جمع خطی توابع پایه مستقل آماری کاملاً توصیف شوند. چنین وابستگی واریانسی ای در بین ضرایب مویجک نیز دیده می شود، که فرم دیگری از کدگذاری خطی تصاویر طبیعی هستند (Buccigrossi and Simoncelli 1999).

یک روش مستقیم که این نوع افزونگی را کاهش می دهد نرمال کردن هر ضریب با اندازه سایر ضرایب است که به بازنمایی آماری مستقلتری (!) می انجامد (Schwartz and Simoncelli 2001). از سوی دیگر می توان مدل مولد را توسعه داد تا متغیرهای مخفی مراتب بالاتر را پوشش دهد که این متغیرها به عنوان آبرپارامتر عمل می کند و وابستگی واریانس بین ضرایب ICA را در خود نگه می دارد (Karklin and Lewicki 2003; Karklin and Lewicki 2005). این مدل‌های غیرخطی برخی از خصوصیت‌های پیچیده پاسخ‌های نورون‌های V1 مانند کنترل بهره و نامتغیری به فاز را شرح می دهد که معتبر بودن فرضیه کدگذاری کارا را بیش از پیش تایید می کند. پیشنهاد شده است که وابستگی های باقیمانده بجای آنکه حذف شود، می تواند برای سازماندهی نقشه های توپوگرافیک نورون‌های ICA به خدمت گرفته شود (Hyvärinen, Hoyer et al. 2001).

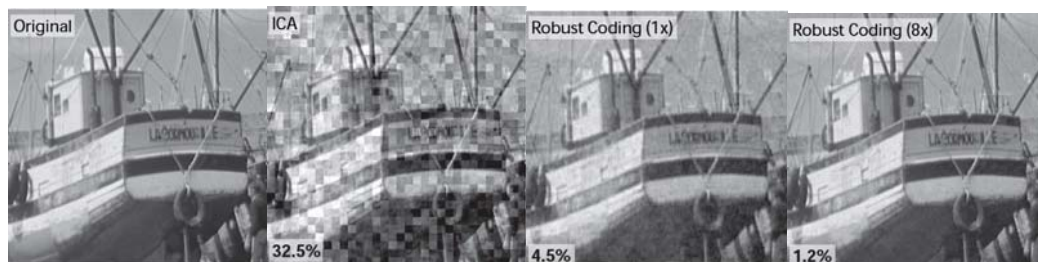
۲-۵-۵-۲- دخیل کردن محدودیت‌های بیولوژیکی

کدگذاری بهینه می تواند جنبه های زیادی از خصوصیات پاسخ سیستم بینایی اولیه را توجیه کند، ولی خصوصیت‌های مهمی باقی مانده است که تاکنون توضیح داده نشده است. یکی از این خصوصیت‌ها مهیت چند مقیاسه بازنمایی بینایی اولیه است، مخصوصاً تنظیم فرکانس مکانی در سلول‌های ساده V1 که در سه اکتاو در نقاط دور از مرکز شبکه رخ می دهد (De Valois, Albrecht et al. 1982; Anderson and DeAngelis 2004). در مقابل، کدگذاری پراکنده و ICA بازنمایی های تک مقیاسی را بدست می هند که در شکل ۲-۴۵ مشهود است. البته کدگذاری پراکنده در ابتدا با هدف بدست آوردن یک بازنمایی چندمقیاسه تهیه شده است (Olshausen 1996) ولی تحلیل‌های دقیقتر با استفاده از تصاویر طبیعی تنظیم شده در (Van Hateren and Van der Schaaf 1998; Doi, Inui et al. 2003) بازنمایی تک مقیاسه ای مشابه ICA را بدست داد. علاوه بر آن در فیزیولوژی سلول‌های تنظیم فرکانس مکانی زیادی وجود دارند که در کدگذاری پراکنده و ICA وجود داشته باشد هم تعداد آنها ناچیز است. بالأخص بازنمایی های چندمقیاسه همچنان در RCGها و LGN یافت می شود (Derrington and Lennie 1984; Croner and Kaplan 1995; Lee, Mumford et al. 2000; McMahon, Lankheet et al. 1998). در حالی که فیلترهای وایتینگ تنظیم تک مقیاسه بدست می دهد (شکل ۲-۴۱). سایر شکست‌های پیش بینی به کمک کدگذاری بهینه تنظیم رنگ در V1 در مقایسه با مدل ICA بوده (doi 2003) و همچنین تنوع شکل میدان گیرندگی سلول‌های ساده در مقایسه با سازماندهی میدان‌های گیرندگی کدگذاری بهینه و ICA است (Ringach 2002).

یک روش برای حل این مشکل پالایش مدل با تعبیه کردن قیود بیولوژیکی واقع بینانه تری در مدل است که فرم کد بهینه پیش بینی شده را تغییر می دهد. یک فرض که توسط مدل‌های قدیمی تر استفاده می شود اینست که نورونها دارای دقت بی نهایتی هستند زیرا فعالیت نورونها با اعداد نقطه شناور نمایش داده می شود. نورون‌های واقعی دارای ظرفیت محدودی هستند که تخمین زده می شود که به کمی یک بیت در هر اسپایک نرون باشد (Fairhall, Lewen et al. 2001;)

(Dhingra and Smith 2004). در چنین شرایط نویزی ای، استفاده از کدگذاری کارا به تنهایی می تواند مضر باشد زیرا افزونگی تنها دلیلی است که نویز را از سیگنال جدا می کند. در این حالت افزونگی سودمند است و می تواند برای غلبه بر این نویز ذاتی در ظرفیت محدود نورون استفاده شود (Barlow 2001). به طور کلی تر، باید بتوان اندازه جمعیت نورون را افزایش داد تا اطلاعات تصویری قویتر کدگذاری شود. از لحاظ فیزیولوژیکی اعتقاد رایج بر اینست که یک گلوگاه روی اعضای بینایی وجود دارد و این ورودی باید با ناهمبسته سازی توسط PCA نمایش داده شود. بهر حال این امر تنها برای نواحی جانبی شبکه صادق است. در ناحیه فوویا، تعداد سلولها همواره بدون هیچ گلوگاهی افزایش می یابد که احتمال وجود بازنمایی افزونه را به ذهن متبادر می کند.

اخیراً مدل‌هایی نظری تولید شده است که این قیود بیولوژیکی مربوط به محاسبات را به حساب می آورند. کدکننده و دیکد کننده خطی بهینه با استفاده از واحدهای نویزی مشابه بازنمایی نورونی در نظر گرفته می شود (اگر تعداد واحدها با ابعاد ورودی برابر باشد، آنها با توابع پایه و فیلترها متناظر خواهند بود) (Doi and Lewicki 2005; Doi and Lewicki 2006; Doi, Balcan et al. 2006). در این مدل که به آن کدگذاری مقاوم^{۱۳۸} گفته می شود، هدف محاسباتی بجای استنتاج متغیرهای نهفته، کمینه کردن خطای بازسازی است و بنابراین تعداد واحدهای کدگذاری دلخواه است و اجباری ندارد که با تعداد واحدهای مفروض مدل برابر باشد. این امر امکان مدل کردن هر ناحیه شبکه را فراهم می سازد که نسبت نورونها به گیرنده های نوری در آن دائماً در حال تغییر است و بین دو بازنمایی فراکامل (نسبت بزرگتر از یک) و فروکامل (نسبت کوچکتر از یک) سوئیچ می کند. تحلیل راه حل بهینه نشان داده است که کد بهینه با توجه سطح نویز کانال، ماتریس کواریانس ورودیها، و تعداد واحدهای کدکننده مدام تغییر شکل می دهد. بازسازی تحت تاثیر نویز کانال نسبت به PCA، ICA و موجک بهبود چشمگیری داشته است که این مسئله بخصوص برای بازنمایی فراکامل $8 \times$ دیده می شود (شکل ۲-۴۶). از لحاظ نظری می توان نشان داد که خطای بازسازی با افزایش تعداد واحدهای کدگذاری کننده می تواند تا حد دلخواه کوچک شود (Doi, Balcan et al. 2006).

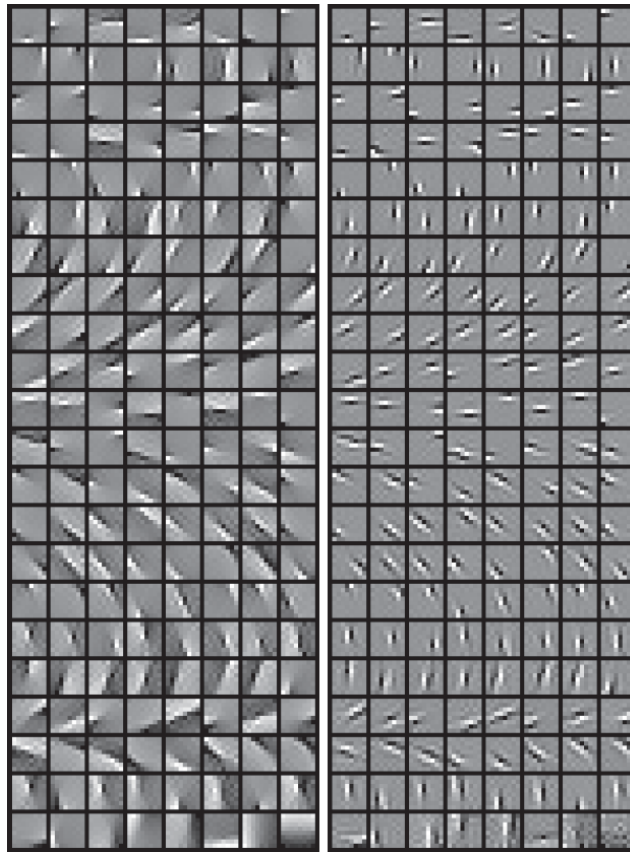


شکل ۲-۴۶: نتایج بازسازی با واحدهایی دارای ظرفیت محدود. در این آزمایش نویز کانال اضافه می شود به طوری که ظرفیت اطلاعاتی کانال به یک بیت به ازای هر ضریب برسد. در هر تصویر بازسازی شده، درصد خطای بازسازی نیز ذکر شده است.

در مدل فوق تابع هزینه توسط آمارگان مرتبه دوم داده ها و نویز بدست می آید. این مدل را چنان بسط داده اند که بازنمایی پراکنده باشد و برای بهبود کارایی کد، قاعده مندی مراتب بالاتر را نیز فاش کند. مثالی از کدکننده و دیکد کننده حاصل در شکل ۲-۴۷ آمده است، ظاهراً شبیه محصولات ICA (شکل ۲-۴۵) است ولی در عمل شامل تعداد بیشتری

¹³⁸robust coding

واحدهای کدگذاری کننده است و نیز تنظیم فرکانس مکانی بیشتری مشابه سلولهای ساده V1 از خود نشان می دهد. این مدل یک چارچوب ابتدایی را برای بکارگیری افزونگی در بازنمایی با داشتن تعداد دلخواهی واحد کدکننده مهیا کرده است.

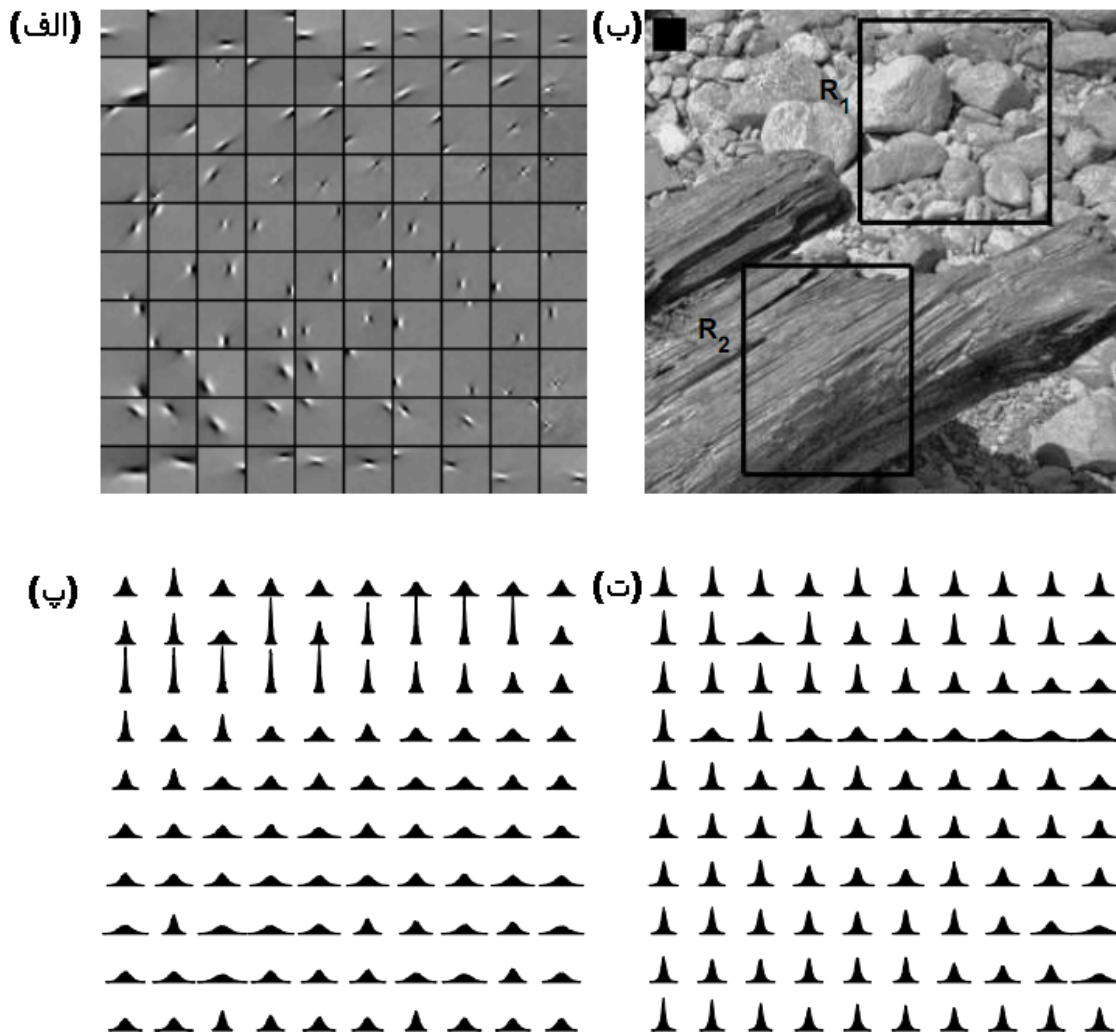


شکل ۲-۴۷: بازنمایی فراکامل پراکنده با واحدهای نوپز دار که به تنظیم فرکانس مکانی زیادی دست می آید. این تصاویر زیرمجموعه دیکدر(چپ) و کدکننده(راست) بردارهایی از تکه تصویرهای 8×8 پیکسلی طبیعی هستند که بر اساس فرکانس مکانی مرتب شده اند. مجموعه کامل مشتمل بر $8 \times 63 = 504$ واحد کدکننده است. ابعاد ذاتی ۶۳ است زیرا مؤلفه DC حذف شده است و این بازنمایی فراکامل $8 \times$ را نشان می دهد.

۲-۵-۵-۳- وابستگیها بر اساس زمینه^{۱۳۹}

یکی از مطالعات جدیدی که خصوصیات آماری تصاویر را مورد بررسی قرار داده، وابستگی بین واریانس ها را بر اساس زمینه تصویر تشریح می کند (Karklin and Lewicki 2005). در این تحقیق، پاسخ فیلترهای خطی گبر روی مجموعه ای از تصاویر طبیعی مورد آزمایش قرار گرفتند. برای تصاویر خاصی، پاسخ فیلترها از شرط مستقل بودن انحراف پیدا می کرد و توزیع آن مطابق با توزیع فرض شده نبود. شکل ۲-۴۸ پاسخ نمونه فیلترها را برای یک تصویر خاص نشان می دهد. همانطور که ملاحظه می شود، پاسخ فیلترها برای محدوده خاصی از تصویر که دارای بافت هم جهت با فیلتر می باشد، با توزیع پیش فرض مطابق نیست. در این مقاله، این خصوصیت از پاسخ فیلترها به زمینه تصویر نسبت داده شده است به طوری که زمینه های

متفاوت الگوهای متفاوتی از وابستگی را در تصاویر به وجود می‌آورند. این وابستگی‌ها بسته به اندازه زمینه بین پاسخ فیلترهای مختلف برقرار است.

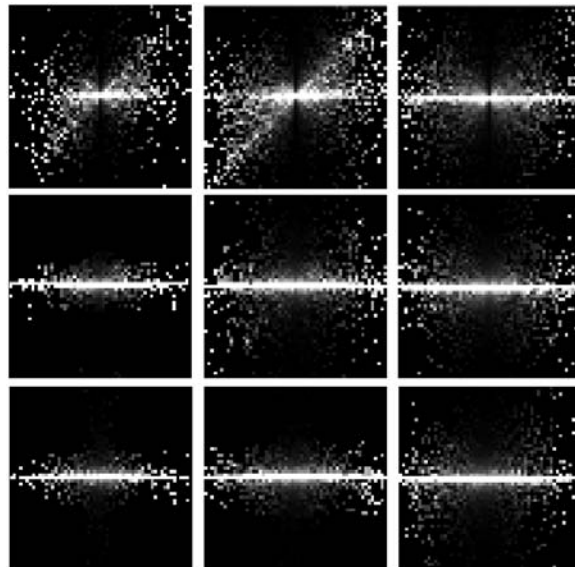


شکل ۲-۴۸: وابستگی پاسخ مؤلفه‌های مستقل به زمینه در تصویر. (بالا چپ) مجموعه‌ای از مؤلفه‌های مستقل استخراج شده از تصویر. (بالا راست) پاسخ مؤلفه‌های مستقل به تکه‌هایی از تصویر از نواحی R_1 و R_2 مورد تحلیل قرار گرفتند. (پایین چپ) پاسخ مؤلفه‌های مستقل به ناحیه R_1 (پایین راست) پاسخ مؤلفه‌های مستقل به ناحیه R_2 . پاسخ مؤلفه‌های مستقل با جهت مشابه در زمینه‌های ذکر شده با توزیع احتمال پیش فرض تفاوت دارد.

۲-۵-۵-۴- وابستگیها بر اساس کلاس اشیا

افزونگی موجود در تصاویر به کلاس تصاویر مورد بررسی بستگی دارد و بسیار پیچیده‌تر از وابستگی در همسایگی محدود می‌باشد (Malmir and Shiry 2009). شکل ۲-۴۹ توزیع احتمال توام پاسخ فیلترهای خطی گبر را نشان می‌دهد. این توزیع بر روی تصاویری از کلاس‌های مختلف اشیا جمع‌آوری شده است. سه جفت فیلتر با فواصل نسبی متفاوت از یکدیگر در سه

شکل ۲-۴۹ نشان داده شده اند. سه مجموعه از تصاویر برای محاسبه توزیع توام پاسخ ها استفاده شده اند: تصاویر صورت، تصاویر هواپیما و تصاویری از اشیا متفاوت.



شکل ۲-۴۹: توزیع احتمال توام پاسخ فیلترهای گیر به مجموعه تصاویر از کلاس های مختلف. هر سطر پاسخ یک زوج فیلتر با فاصله نسبی متفاوت را ارائه می کند. نمودارهای هر ستون با استفاده از مجموعه تصاویر مختلفی محاسبه شده اند. سه مجموعه از تصاویر برای محاسبه توزیع توام پاسخ ها استفاده شده اند: تصاویر صورت، تصاویر هواپیما و تصاویری از اشیا متفاوت. نمودارهای ستون سمت چپ با استفاده از تصاویر اشیا متفاوت محاسبه شده اند. نمودار ستون وسط از روی تصاویر چهره و نمودارهای ستون راست نیز با استفاده از تصاویر هواپیما محاسبه شده اند. زوج فیلتر سطر اول دارای فاصله نسبی ۷ واحد، سطر دوم فاصله ۱۵ و سطر سوم فاصله ۳۰ واحد می باشند. تصاویر از مجموعه ۱۰۱ شی استخراج شده اند

اولین نکته ای که در شکل ۲-۴۹ بارز است تفاوت توزیع توام برای دو مجموعه فیلتر در ستون های مختلف است. برای زوج فیلتر با کمترین فاصله نسبی، وابستگی واریانس روی هر سه مجموعه تصاویر مشاهده می شود. در سطر دوم، زوج فیلتر برای مجموعه تصاویر صورت و هواپیما وابستگی را نشان می دهند. ولی برای تصاویر از اشیا مختلف، وابستگی در پاسخ این فیلترها مشخص نیست. برای زوج فیلتر سطر سوم که دارای بیشترین فاصله از یکدیگر هستند، وابستگی واریانس تنها روی مجموعه تصاویر هواپیما مشخص می باشد. برای دو مجموعه دیگر تصاویر، این فیلترها به نظر مستقل می آیند (Malmir and Shiry 2009).

مدل های قبلی که بر اساس وابستگی واریانس بین پاسخ فیلترهای خطی گسترش یافته بودند، نمی توانند این پدیده را توجیه کنند. در این مدل ها، دو فیلتر یا به هم وابسته هستند یا از هم مستقل اند و این موضوع به مجموعه تصاویر مورد استفاده ارتباطی ندارد. در شکل ۲-۴۹ مشاهده می شود که این فرض نادرست است و وابستگی واریانس به مجموعه تصاویر استفاده شده بستگی دارد. توزیع احتمال توام برای پاسخ دو فیلتر خطی بسته به موقعیت قرار گرفتن آنها و مجموعه تصاویر مورد استفاده، ممکن است که وابسته باشند یا مستقل به نظر برسند. می توان گفت که اندازه ی همسایگی که در آن وابستگی واریانس وجود دارد با مجموعه تصاویر استفاده شده ارتباط مستقیم دارد. مدل های قبلی از مجموعه محدودی از تصاویر استفاده می کردند و این پدیده از نظر آنها پنهان مانده است.

این مشاهده یک توجیه ساده دارد. وابستگی واریانس نشان دهنده پاسخ بزرگ همزمان در دو فیلتر خطی می باشد. این حالت هنگامی اتفاق می افتد که فیلترهای خطی از ناحیه همسانی از تصویر ورودی دریافت کنند و این نواحی خواصی مطابق با گزینندگی آنها داشته باشند. هنگامی که ورودی فیلترها از تصاویر مختلفی باشند، این شرط تنها برای فیلترهای با فاصله نسبی کم برقرار می شود. دلیل آن نیز یکنواخت بودن بافت ها در تصاویر است که نواحیکسانی را در تصاویر تشکیل می دهند. به تدریج که فاصله فیلترها از یکدیگر زیاد می شود، احتمال دریافت آنها از نواحیکسان نیز کاهش مییابد و وابستگی واریانس از میان می رود. این حالتی است که در ستون دوم و سوم سطر اول شکل ۲-۴۹ اتفاق افتاده است (المیر ۱۳۸۸).

هنگامی که تصاویر از کلاس یکسانی مثل تصاویر چهره انسان باشند، دامنه مکانی وابستگی واریانس گسترش مییابد. در مجموعه تصاویری که از یک کلاس خاص هستند، بافت هایی که یکسان هستند نواحی بزرگتری در تصاویر مختلف تشکیل می دهند. بدین ترتیب وابستگی واریانس بین فیلترهای با فاصله نسبی بیشتر مشاهده می شود. این حالتی است که در سطرهای دوم و سوم شکل ۲-۴۹ مشاهده می شود. بنابراین می توان گفت که افزونگی های موجود در تصاویر طبیعی، به کلاس تصاویر وابسته است و فرض همسایگی محدود در وابستگی واریانس نادرست می باشد. در حقیقت افزونگی های موجود در تصاویر طبیعی بسیار پیچیده تر از این هستند که با یک لایه نگاشت خطی و به دنبال آن نگاشت غیر خطی در همسایگی های ثابت، به مؤلفه های مستقل تصاویر دست یافت (Malmir and Shiry 2009). این الگوهای پیچیده ی وابستگی راه را برای توجیه عملکرد لایه های مختلف از قشر بینایی مغز هموار می سازد.

نورون های قشر ابتدایی بینایی در مغز وضعیتی مشابه با فیلترهای ذکر شده در سطر اول شکل ۲-۴۹ دارند. این نورون ها ورودی خود را از تصاویری از کلاس های مختلف دریافت می کنند و پاسخ آنها اطلاعات بسیار کمی در مورد کلاس تصویر منتقل می کند. هنگامی که یک نورون V1 فعال می شود، تصویر ورودی متعلق به دامنه بسیار بزرگی از تصاویر می تواند باشد. این مطلب به خاطر سادگی خصوصیت تشخیص داده شده توسط نورون های V1 می باشد. نورون های قشر ابتدایی بینایی تغییر شدت روشنایی در یک محدوده مکانی خاص از تصویر را تشخیص می دهند که این خصوصیت تقریباً در تمام تصاویر یافت می شود. بنابراین نورون های V1 در همسایگی های محدود مکانی خصوصیت وابستگی واریانس را از خود نشان می دهند. افزونگی هایی که در محدوده های مکانی بزرگتر قرار دارند از دید این نورون ها پنهان هستند و در ناحیه V1 تنها وابستگی در محدوده های کوچکی از تصاویر ورودی حذف می شود. وابستگی هایی که در محدوده های بزرگتر و به کلاس خاص از اشیا مرتبط هستند توسط لایه های بالاتر قشر بینایی حذف می شوند.

۲-۵-۶- نتیجه گیری

مثالهایی که به طور اجمالی در این بخش مرور شد نشان می دهد که خصوصیات پاسخ سیستم بینایی اولیه می تواند با قاعده مندیهایی تصاویر طبیعی، اهداف محاسباتی مانند کدگذاری بهینه، و قیود بیولوژیکی مرتبط مانند نویز ذاتی توصیف شود. با اینحال چند مورد قابل ذکر باقی مانده است. این موارد شامل بازنماییهای موقتی مکانی (Olshausen 2002)، تلفیق قیود بیولوژیکی بیشتر (مانند کارایی انرژی یا طول سیم کشی محدود) با ایده های کدگذاری کارا (Laughlin and Sejnowski 2003)، تاثیر وظیفه و سیگنالهای حسی مرتبط بر کد نورونی (Lewicki 2002)، و شبیه سازی سیستمهای بینایی و محیطهای واقعی (شامل آپتیک و چینش حسگرهای چشم، حرکات چشم و ...) است. همچنین پیشنهاد شده است که

کدگذاری نورونی شنوایی طوری سازگار شده است که صداهاى طبیعى را کارا کدگذاری کند (Lewicki 2002). نتایج جدیدتر نشان داده است که خود اسپایکهای نورونی نیز نقشی اساسی در شکل گیری کدی کارا برای سیگنالهای شنوایی دارند (Smith and Lewicki 2005; Smith and Lewicki 2006). همه این نتایج نشان می دهند که اصل کدگذاری بهینه وقتی که با قیود بیولوژیکی واقع گرایانه ترکیب می شود شکلهای دیگری از کدگذاری حسگری و ادراک را به ما نشان می دهد.

[بررسی کارهای پیشین]

۱. مدل‌های نظری عملکرد قشر بینایی مغز
۲. مروری بر روشهای آشکارسازی چهره

۳-۱- مدل‌های نظری عملکرد قشر بینایی مغز

۳-۱-۱- مقدمه

دو روش کلی برای مطالعه آمارگان سیگنال‌های طبیعی وجود دارد. مدل‌های پایین به بالا به مطالعه قاعده مندیه‌های آماری تجربی بدست آمده از نگاشته‌های کم بُعد خطی و غیرخطی می پردازند. این روشها نورونهای قشر مخ را به عنوان عواملی برای انتخاب و دستکاری نگاشته‌ها می نگرند که هدف آنها بهینه سازی معیارهای احتمالاتی و نظریه محاسباتی است (Shannon 1949; Shannon and Weaver 1948) که از این قبیل معیارها می توان به پراکندگی (Field 1987) و کدگذاری کارا (که استقلال آماری یکی از مشتقات آنست) (Attneave 1954; Barlow 1961; Li and Atick 1994; Nadal and Parga 1997) اشاره کرد. بر خلاف این روشها، روشهای بالا به پایین (Neisser 1967; Hinton and Ghahramani 1997) بر مبنای خصوصیات احتمالاتی فرایندهایی است که سیگنالها را تولید می کنند و نورونهای قشر مخ را به صورت یک دستگاه مختصات می بینند که پارامترهای رویه آماری سیگنال را تعیین می کنند.

۳-۱-۲- کدگذاری بهینه

چه اصول محاسباتی زیربنای پردازش در قشر بینایی مغز را تشکیل می دهد؟ یکی از خصوصیات سیگنال های طبیعی که همواره مورد توجه بوده است افزونگی موجود در این سیگنال ها می باشد. در دهه های ۵۰ و ۶۰ میلادی و پس از معرفی نظریه اطلاعات توسط شانون، کارکرد این نظریه در سیگنال های طبیعی و قشرهای حسی مغز مورد توجه قرار گرفت. یک ایده که فرضهای اندکی درباره اهداف سیستم حسی و مکانسیمهای آن دارد اینست که چنین سیستمی باید اطلاعاتی را درباره ورودی خود حفظ کند و در همان حال افزونگی کد بکار رفته را کاهش دهد (Attneave 1954; Barlow 1961; Simoncelli and Olshausen 2001). این نظریه بر مبنای این مشاهده است که ورودی های حسی به مغز انسان دارای حجم بسیار زیادی می باشند. نواحی سطح بالا در مغز که وظیفه فکر کردن، تصمیم گیری و دیگر عملیات شناختی سطح بالا را بر عهده دارند دارای سرعت اجرایی کمی می باشند و حجم اطلاعاتی که می توانند در هر لحظه پردازش کنند از حجم داده های حسی که وارد مغز می شوند بسیار کمتر است. مطابق نظریه کد کردن بهینه، وظیفه قشرهای حسی، پردازش داده های ورودی است به نحوی که حاصل این پردازش، افزونگی نداشته باشد و حجم آن متناسب با اطلاعاتی باشد که منتقل می کند. به عنوان مثال، برای یک موجود زنده، داده هایی که از طریق بینایی به مغز آن وارد می شوند نسبت به اطلاعات مهم در تصاویر مانند شکارچی، طعمه یا جفت، دارای حجم بیشتری می باشند. قسمت های تصمیم گیری و حرکتی مغز نمی توانند متناسب با حجم انبوه داده‌های ورودی عمل کنند زیرا توان عملیاتی آنها پایین تر از این حجم داده است. برای چنین موردی، اطلاعات مهم نظیر وجود شکارچی یا شکار باید توسط قشر حسی از تصاویر استخراج شده و به قسمت های تصمیم گیری در مغز فرستاده شوند. این فرضیه که به نظریه کدگذاری بهینه معروفست، یک اصل کلی است که می تواند به هر نوع حسگری و سیگنال ورودی اعمال شود. با وجود آنکه این نظریه محدودیتهای خود را دارد (مثلاً به همه اطلاعات ورودی وزن یکسانی را اختصاص می دهد که بعضی از آنها ممکن است از بقیه اهمیت بیشتر یا کمتری برای عملکرد سیستم در محیط داشته باشند) (مالمیر ۱۳۸۸)، ولی باز هم به عنوان یک اصل منطقی برای تحلیل مراحل اولیه که با اطلاعات خام وارده از ارگانهای

حسی سروکار دارد مناسب است. نظریاتی مشابه با این نظریه در طی سال های مختلف و توسط افراد مختلف ارائه شده اند (Laughlin 1981; Atick 1992; Field 1994).

کارایی کد حسی به اینکه کد چقدر با آمارگان پیامهایی که مخابره می کند تطابق دارد، وابسته است (Shannon and Weaver 1949). در زمینه کدگذاری نورونی، این بدین معناست که برای اینکه کد کارا باشد، بازنمایی نورون باید مشخصاً با ساختار آمارگان ورودی خود سازگار باشد. به عنوان مثال، ورودی حسی که حاوی یک الگوی خاص از همبستگی است، اگر در لایه های اولیه پردازش ناهمبسته سازی شود، می تواند خیلی کارا تر منتقل گردد. این مسئله همچنین بدین معناست که بازنمایی که برای یک مجموعه از ورودیها یا نوع خاصی از محیط حسگری بهینه است، الزاماً برای انتقال نوع دیگری از اطلاعات بهینه نیست. بنابراین با مشخص کردن آمارگان ورودیهای سیستم بینایی (تصویر طبیعی) می توان به یک سیستم بهینه دست یافت و با استفاده از این توصیف می توان به درکی از بینایی بیولوژیکی دست یافت.

محیط حسی ما آکنده از ساختارهای آماری است. تصاویر صحنه های بیرون از منزل حاوی لبه ها، شکلها، بافتها و مرزهاست. ظاهر و بهم بندی این عناصر قاعده مندیهایی آماری را در ورودیهای حسی پدید می آورد که متفاوت از آمارگان الگوهای تصادفی است. این الگوها در طول فضای بینایی توزیع شده اند و به راههای غیرخطی پیچیده ای تعامل می کنند بنابراین یافتن یک توصیف فشرده چندان آسان نیست. با اینحال کدگذاری کارا یک چارچوب آماری را پدید می آورد که طبعاً با بسیاری از جنبه های حسگری در دنیای غیرقطعی مانند وجود نویز کنار می آید و کمیتهایی که به طور ضمنی در سیگنال تعبیه شده اند را استنتاج می کند.

داده هایی که از یک تصویر طبیعی استخراج شده است هنگامی که روی یک بردار تصادفی نگاشت می شود هیستوگرامی خواهد داشت که قله آن در صفر نوک تیزتر و دنباله های آن سنگین تر از توزیع گوسی خواهد بود (Field 1989; Daugman 1987). معیار تیزی قله، کورتوسیس است که مطابق تعریف آن کورتوسیس توزیع گوسی برابر صفر است و کورتوسیس مثبت نشانگر توزیعهای دارای قله های نوک تیزتر است. این ویژگی همانگونه که جلوتر با آن اشاره خواهد شد، نقطه عطفی در استفاده از نظریه کدگذاری بهینه در بازنمایی های تصویر خواهد بود. بحث شده است که چنین توزیعهای بهینه و پراکنده ای به کدهایی منجر می شوند که برای انتقال اطلاعات (Field 1987) و حافظه های انجمنی (Zetsche 1990; Field 1994) کارا تر از روشهای سنتی هستند و مطالعات اولیه با توجه به اینکه خروجی فیلترها نماینده میدان گیرندگی سلولهای ساده هستند، این بازنمایی را به قشر خاکستری مخ نسبت می دادند (Field 1994; Zetsche and Barth 1990; Daugman 1989; 1987).

۳-۱-۳- مدلهای پایین به بالا

این مدل ها اکثراً اطلاعات مورد نیاز خود را از مطالعات عصب شناسی که بر روی تک نورون ها انجام می شود به دست می آورند. سپس مدلی سیستمی یا شبکه عصبی ارائه می کنند که بتواند این رفتار را شبیه سازی کند. نکته مثبت این مدل ها، تعمیم پذیر بودن آنها به لایه های مختلف قشر بینایی می باشد (Malmir and Shiry 2009).

پیشرفتهای اخیر در آمارگان پایین به بالا بسیار چشمگیر بوده است. حجم زیادی از مطالعات به آزمودن خواص آماری محصول کانالو فیلترهای خطی با تصاویر پرداخته اند که این فیلترهای خطی معمولاً به گونه ای انتخاب می شوند که از لحاظ کیفی با میدانهای گیرندگی نورونهای شبکیه یا قشر مخ مطابقت داشته باشد. میدان گیرندگی بخش اولیه قشر بینایی مخ (یعنی سلولهای ساده) برای یک ناحیه مکانی، راستا، و فرکانس مکانی خاص تنظیم شده اند (Hubel and Wiesel 1962) و از اینرو فیلترهایی خطی که بتوانند این پارامترها را در خود بگنجانند و کنترل کنند مورد توجه قرار گرفته اند. از جمله این فیلترها می توان به فیلتر گابور اشاره کرد. فیلتر گابور از ضرب یک تابع سینوسی در تابع نمایی به دست می آید و نشان داده شده که مدل مناسبی از گیرندگی نورون های V1 می باشد (Jones and Palmer 1987):

$$F(x, y) = \exp\left(-\frac{(x_0^2 + \gamma^2 y_0^2)}{(2\sigma^2)}\right) \cos\left(\frac{(2\pi x_0)}{\lambda}\right) \quad (1-3)$$

$$x_0 = x \cos(\theta) + y \sin(\theta)$$

$$y_0 = -x \sin(\theta) + y \cos(\theta)$$

که در این معادله، θ راستا، γ نسبت کشیدگی، σ طول موثر و λ فرکانس مکانی فیلتر را مشخص می کنند.

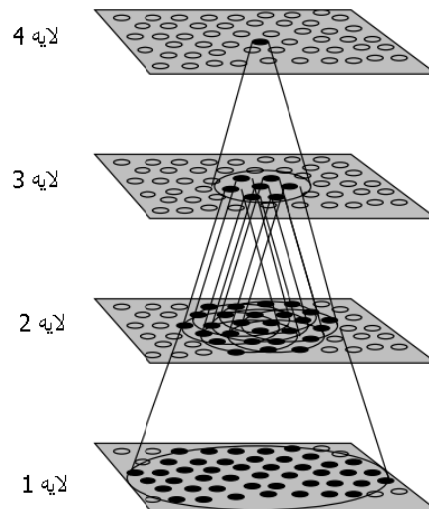
این میدانهای گیرندگی قرابت بسیاری به تجزیه چندمقیاسه موجک^{۱۴۶} دارند که در جامعه بینایی رایانه ای به توفیق زیادی دست یافته است. برای تصاویر طبیعی، مشاهدات تجربی فعالیتهای یک فیلتر خطی توزیعی با کورتوسیس بالا (بنابراین پراکنده) را از خود نشان می دهد (Field, 87). گروه هایی از فیلترهای خطی (که با توجه به پارامترهایی نظیر راستا، فرکانس، فاز، یا موقعیت مکانی مرتب می شوند) وابستگی خطی زیادی از خود نشان داده اند (Wegmann and Zetsche 1990; Simoncelli 1997; Zetsche and Barth 1990; Simoncelli 1997). که می توان آنها را به کمک واریانس نشان داد (Simoncelli 2001; Buccigrossi and Simoncelli 1999; Schwartz and Simoncelli 1997). اهمیت آمارگان واریانس بیشتر نیز در حوزه پیکسل نشان داده شده بود (Lee 1980) و در رشته هایی مانند گفتار (Brehm and Stammer 1987) و حتی اقتصاد (Bollerslev, Engle et al. 1994) نیز به آن پرداخته شده است. در ادامه به مدل‌های پایین به بالا و استفاده آنها از انواع وابستگیها برای تعیین قوانین یادگیری و ... پرداخته می شود.

۳-۱-۳-۱ مدل VisNet

این مدل بر اساس نظریه تشخیص نامتغیر اشیا^{۱۴۷} در قشر بینایی مغز توسعه یافته است (Rolls, Cowey et al. 1992). این مدل از مجموعه ای از لایه های مختلف تشکیل شده است که در هر لایه مکانیزم رقابت بین نورون ها با استفاده از اتصالات بازدارنده وجود دارد (Rolls and Stringer 2006). لایه های مختلف در یک معماری سلسله مراتبی به هم متصل هستند و اندازه ناحیه دریافت نورون ها در لایه های بالاتر افزایش می یابد. معماری VisNet در شکل ۳-۱ نشان داده شده است.

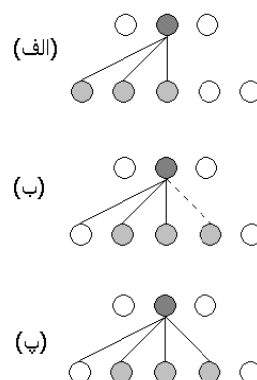
¹⁴⁶ multi-scale wavelet decomposition

¹⁴⁷ invariant object recognition



شکل ۳-۱: مدل VisNet. در این مدل، یک مجموعه از لایه ها به صورت سلسله مراتبی به هم متصل هستند. قوانین یادگیری در این مدل برای تشخیص نامتغیر اشیا توسعه یافته اند.

در این مدل برای یادگیری پاسخ نامتغیر به تبدیلات اشیا، قانون یادگیری تبدیلات پیوسته به کار برده شده است (Rolls and Stringer 2006). این قانون برگرفته از رفتار نورون های قشر بینایی مغز در مواجهه با محرک های مختلف می باشد. شکل ۳-۲ این قانون را توضیح می دهد. ایده اولیه بر اساس این مشاهده است که پاسخ سطح بالای یک نورون به محرک تا مدت کوتاهی پس از مشاهده محرک باقی می ماند (Földiák 1991). مطابق این قانون، یادگیری پاسخ نامتغیر در نورون ها از مشاهده تبدیلات پیوسته اشیا در محیط حاصل می شود. منظور از تبدیلات پیوسته اشیا این است که خصوصیات تصویر مانند اندازه یا مکان اشیا به صورت پیوسته تغییر می کنند. این پیوستگی در زمان سبب می شود که نورون های سطح بالاتر از نورون هایی که تبدیلات یک شی را نشان می دهند ورودی دریافت کند و اتصال بین آنها قوی تر شود.



شکل ۳-۲: یادگیری بر اساس تبدیلات پیوسته. (الف) نورون های خاکستری در لایه پایین نمایش یک شی هستند و نورون لایه بالاتر را فعال کرده اند. اتصال بین این نورون ها تقویت شده است. (ب) در اثر تبدیلات پیوسته، نمایش شی در لایه پایین تر تغییر می کند. در اینجا اتصال بین نورون جدید و نورون سطح بالاتر با خط چین نمایش داده شده است. (پ) نورون سطح بالاتر به خاطر حفظ سطح فعالیت خود، می تواند اتصال با نورون جدید را تقویت کند.

بر این اساس قانون یادگیری که به قانون ردیابی^{۱۴۸} موسوم است، برای به روزرسانی اتصالات بین نورون ها گسترش یافته است. این قانون گسترش یافته ی قانون هبی می باشد:

$$\nabla w_{ij}^t = \lambda \bar{y}_j^t (x_i^t - w_{ij}^{t-1}) \quad (2-3)$$

در معادله بالا، ∇w_{ij}^t تغییر وزن بین نورون های i و j در تکرار t را نشان می دهد، λ ضریب یادگیری است و x_i^t مقدار پاسخ نورون i را مشخص می کند. تنها تفاوت این قانون با قانون یادگیری هبی در جمله \bar{y}_j^t است که علاوه بر پاسخ کنونی نورون j ، ردی از پاسخ های قبلی این نورون را نیز در خود ذخیره می کند:

$$\bar{y}_j^t = \eta \bar{y}_j^{t-1} + (1-\eta)y_j^t \quad (3-3)$$

در معادله بالا، \bar{y}_j^t پاسخ نورون در تکرار t است و η ضریب اثر را نشان می دهد. ضریب اثر در بازه (۰،۱) انتخاب می شود و هرچه مقدار آن به صفر نزدیکتر باشد، اثر پاسخ های قبلی ضعیفتر می شود.

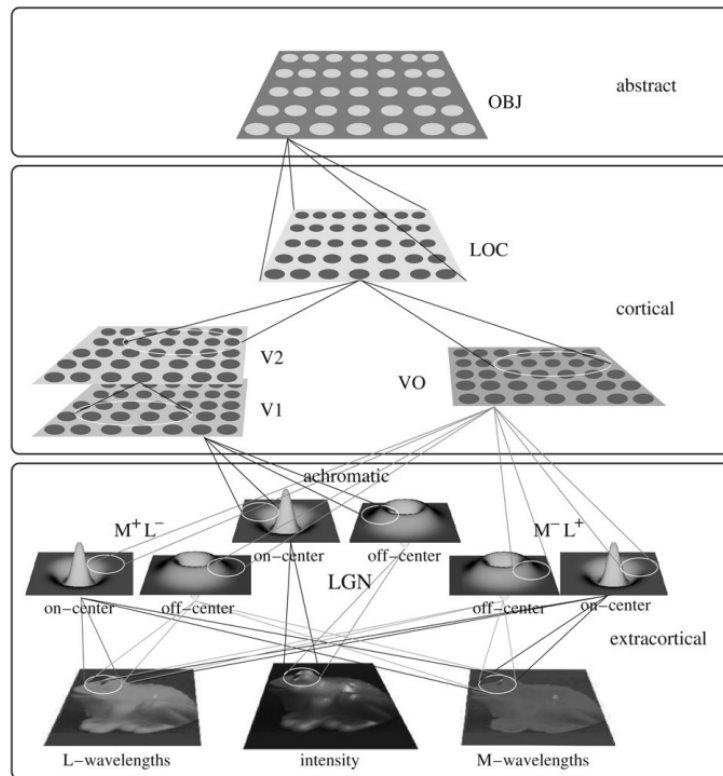
۳-۱-۳-۲- مدل نقشه های قشری مغز^{۱۴۹}

این مدل که به نام RF-Lissom نیز معروف است، بر اساس ایده شبکه های خودسازمانده گسترش یافته است (Plebe and Domenella 2007). مدل اصلی برای شبیه سازی ترکیب چینش نورون های V1 بر اساس راستا توسعه یافته بود (Sirosh and Miikkulainen 1997). این ایده زیربنایی سپس به دیگر نواحی قشر بینایی گسترش داده شده و توانسته است که بازه وسیعی از خصوصیات نورون های نواحی مختلف در قشر بینایی را شبیه سازی کند. خصوصیت اصلی این مدل در کلی بودن مکانیزم های ارتباطی ارائه شده بین نورونها می باشد. در هر لایه، نورونها از نورونهای لایه قبل و نورونهای مجاور خود ورودی دریافت می کنند. در تمام این اتصالات، قوانین خودسازمانده برای یادگیری مقادیر وزن ها به کار برده شده اند. ساختار کلی مدل در شکل ۳-۳ نشان داده شده است.

خصوصیت مهمی که در این مدل وجود دارد، پیدایش نقشه های قشری مشابه با الگوی موجود در قشر بینایی مغز می باشد. به عنوان مثال، لایه V1 در این مدل از مجموعه ای از نورون های LGN ورودی دریافت می کند. نشان داده شده که با قانون یادگیری خودسازمانده، در لایه V1 ترکیبی مشابه با ناحیه V1 قشر بینایی مغز پدیدار می شود. در لایه V2، نورون ها در یک روال یادگیری، محرک هایی به شکل زاویه را یاد گرفتند (Plebe 2007). همچنین در لایه ی آخر مدل، نورون ها همانند نورون های قشر IT به صورت خوشه بندی شده به اشیا مختلف حساسیت پیدا کرده اند.

¹⁴⁸ Trace rule

¹⁴⁹ Cortical maps



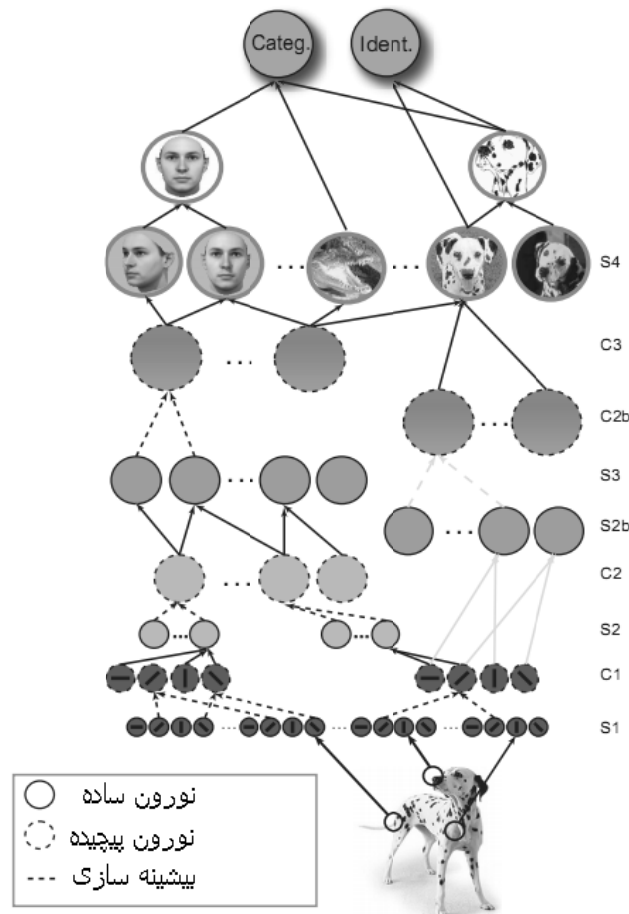
شکل ۳-۳: مدل نقشه های قشری مغز. اتصالات بین لایه ای و درون لایه ای در این مدل با استفاده از قوانین خودسازمانده یاد گرفته می شوند.

۳-۱-۳-۳- مدل HMax

در فصل قبل اشاره شد که نورون های ساده و پیچیده در قشر ابتدایی بینایی عمل تشخیص خصوصیت را به صورت سلسله مراتبی انجام می دهند. یکی از مدل هایی که بر اساس این مکانیزم ارائه شده مدلی است که به طور مستقیم از لایه های مختلف قشر بینایی الهام گرفته شده است (Serre, Wolf et al. 2007) (شکل ۳-۴). مبنای اصلی این معماری، مدل HMax است که در سال ۱۹۹۵ ارائه شده بود (Logothetis, Pauls et al. 1995). مدل HMax عملکرد نورون های پیچیده را با استفاده از عملکرد بیشینه^{۱۵۰} توجیه کرد. در این مدل، هر نورون پیچیده از یک مجموعه از نورون های ساده ورودی دریافت می کند و پاسخ خود را برابر با ماکزیمم پاسخ هر یک از ورودی های خود قرار می دهد. این مدل، در مقابل مدل جمع خطی قرار دارد. در مدل جمع خطی، هر نورون پیچیده، پاسخ خود را بر اساس جمع خطی از ورودی های خود تنظیم می کند (Salinas and Abbott 1997). هر دوی این مدل ها بر اساس مشاهده رفتار نورون های پیچیده در نواحی مختلف قشر بینایی الهام گرفته شده اند. در تایید هر دوی این مدل ها نیز مطالعات عصب شناسی مختلفی وجود دارد (Salinas and Abbott 1997; Gawne and Martin 2002; Lampl, Ferster et al. 2004). در برخی از این مطالعات، مدل بیشینه گیری با رفتار نورون ها بیشتر منطبق است و در برخی دیگر، مدل جمع خطی عملکرد نورون ها را بهتر توجیه می کند.

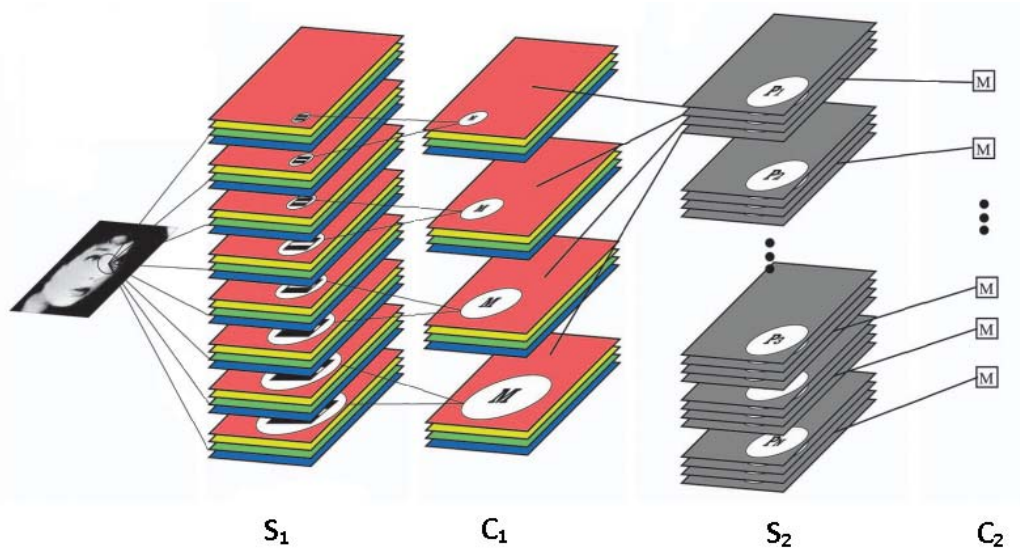
¹⁵⁰ Max operation

معماری سلسله مراتبی ارائه شده در این مدل، از یک مجموعه از نورون های ساده و پیچیده تشکیل شده است. هر لایه دارای تنها یک نوع نورون است و لایه های ساده و پیچیده به طور متناوب و پشت سر هم قرار گرفته اند. وظیفه نورون های ساده، تشخیص خصوصیات در تصویر ورودی است و نورون های پیچیده بر روی این سلول ها عمل بیشینه گیری انجام می دهند. بدین ترتیب، نورون های پیچیده به تبدیلات خصوصیت ورودی پاسخ نامتغیر نشان می دهند. لایه های ساده و پیچیده در این مدل به ترتیب با S_i و C_i مشخص شده اند که i شماره لایه را تعیین می کند.



شکل ۳-۴: مدل سلسله مراتبی بر اساس مدل HMax. لایه های این مدل به صورت تناوبی از نورون های ساده و پیچیده هستند. نورون های ساده وظیفه تشخیص خصوصیات را بر عهده دارند و نورون های پیچیده با عمل بیشینه گیری روی نورون های ساده، پاسخ نامتغیر نسبت به تبدیلات تصویر فراهم می کنند.

مدل دیگری بر مبنای این معماری معرفی شده است که دارای چهار لایه، متناظر با نواحی V1 و V4 در قشر بینایی است. در هر لایه، دو نوع نورون ساده و پیچیده وجود دارند. نورون های V1 با فیلترهای گابور در ۴ جهت و ۸ مقیاس اندازه مختلف مدل شده اند. نورون های لایه های بالاتر با عملگر بیشینه، به تدریج پاسخ نامتغیر به تبدیلات جهت و اندازه در اشیاء را فراهم می کنند. لایه C_1 از نورون های S_1 با دو مقیاس اندازه مختلف بیشینه می گیرند. نورون های S_2 از تمام مقیاس های مختلف ورودی دریافت می کنند و در نهایت نورون های C_2 از تمام جهت ها ورودی دریافت می کنند. جزئیات این مدل در شکل ۳-۵ مشاهده می شود.



شکل ۳-۵: جزئیات مدل ارائه شده بر اساس HMax. در این مدل، لایه اول با فیلترهای گابور در چهار جهت و با هشت مقیاس اندازه ارائه شده اند. لایه دوم از دو مقیاس بیشینه گیری می کند. لایه سوم از تمام مقیاس ها و لایه چهارم از چهار جهت ورودی دریافت می کند.

تفاوت این مدل با مدل HMax در نحوه انتخاب خصوصیات از تصاویر می باشد. در این مدل، یادگیری تنها در لایه چهارم صورت می پذیرد به این صورت که از یک مجموعه از تصاویر آموزشی، به صورت تصادفی یک مجموعه از پاسخ نورون های S_2 انتخاب شده و به عنوان خصوصیت نورون های C_2 در نظر گرفته می شدند. این نحوه یادگیری سبب شده است که نرخ تشخیص مدل ارائه شده تا حد چشمگیری افزایش یابد. این مدل در تصاویر حاوی بافت، محیط های شلوغ^{۱۵۱} و محیط های طبیعی، نرخ تشخیص شیء بالایی از خود نشان داده است.

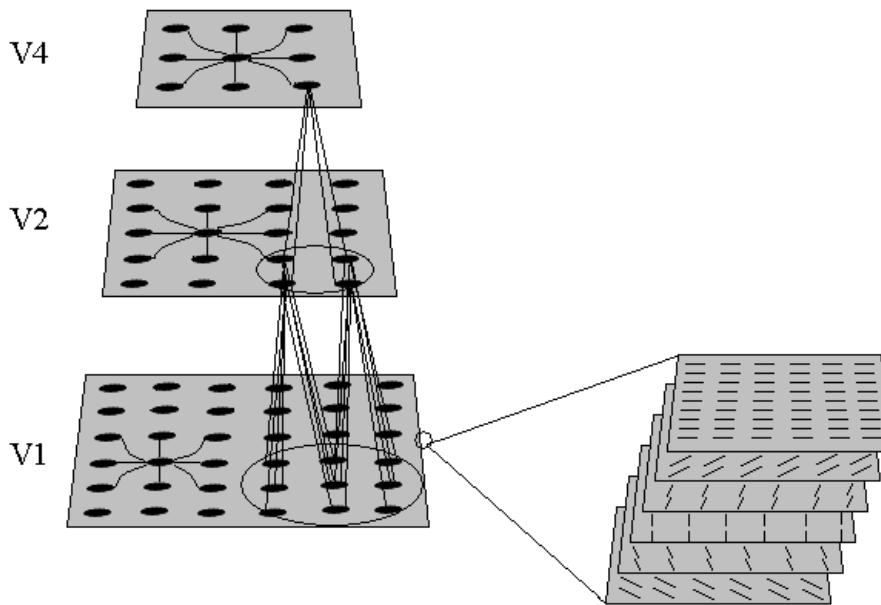
۳-۱-۳-۴- مدل نرمال کردن جانبی واریانس با تقسیم

این مدل (Malmir and Shiry 2009) یک شبکه عصبی چند لایه است که به صورت سلسله مراتبی سازماندهی شده است. شکل ۳-۶ این مدل را نمایش می دهد. مدل ارائه شده دارای سه لایه متناظر با لایه های $V1$ ، $V2$ و $V4$ می باشد. در هر لایه از این مدل، نورونها از لایه پایینتر و از نورونهای مجاور خود ورودی دریافت می کنند. وزن های اتصال به لایه پایینتر، گزینندگی نورون را مشخص می کنند و اتصالات افقی در رفع افزونگی بین پاسخ نورون های مجاور به کار گرفته می شوند. برای یادگیری وزن اتصالات بین لایه ها، از یک قانون هبی ساده استفاده شده است. بزرگی میدان گیرندگی نورونهای هر قشر حدوداً $2/5$ برابر لایه قبل در نظر گرفته شده است.

لایه $V1$ در این مدل با استفاده از فیلتر گابور در ۶ جهت مختلف مدل شده است. هدف از اتصالات افقی در هر لایه، حذف افزونگی بین پاسخ نورون های مجاور می باشد. در این مدل بر اساس مدل نرمال سازی با تقسیم (Schwartz and Simoncelli 2001)، برای حذف وابستگی بین پاسخ فیلترها، پاسخ هر فیلتر به واریانس آن تقسیم می شود. مدل نرمال

¹⁵¹ Cluttered

سازی با تقسیم بیان می‌دارد که هرگاه محرک نورون حسی از نظر مکانی گسترش یابد به نحوی که نورون‌های مجاور آن را نیز فعال کند، پاسخ نورون مرکزی کاهش می‌یابد. این کاهش در پاسخ با تقسیم کردن مدل شده است، یعنی پاسخ نورون بر مقدار پاسخ نورون‌های اطراف تقسیم می‌شود. مدل ارائه شده با رفتار نرمال سازی مشاهده شده در قشرهای سطح بالای مغز مطابقت دارد (Carandini, Heeger et al. 1997; Reynolds, Chelazzi et al. 1999; Zoccolan, Cox et al. 2005)



شکل ۳-۶: مدل سلسله مراتبی نرمالسازی با تقسیم. این مدل دارای سه لایه متناظر با نواحی V1، V2 و V4 در قشر بینایی مغز می‌باشد. در هر لایه اتصالات افقی و اتصالات ورودی از لایه پایین‌تر ورودی نورون را تعیین می‌کنند.

برای تخمین واریانس پاسخ هر نورون با فیلتر، به ازای تصویر ورودی جاری از پاسخ نورون‌های مجاور آن استفاده می‌شود. بر طبق اصل محلیت واریانس، نورونهای مجاور دارای واریانس مشترک هستند و امکان محاسبه واریانس پاسخ را ممکن می‌سازند. این مدل، مدل مالمیر، از ترکیب مربعات پاسخ نورون‌های اطراف برای حدس مقدار واریانس یک نورون کمک می‌گیرد:

$$\text{var}(L_x | L_y, y \in C_x) = \sum_{y \in C_x} w_{xy} L_y^2 + \sigma_x^2 \quad (4-3)$$

در اینجا، L_x پاسخ نورون x قبل از نرمال سازی می‌باشد و C_x همسایگی اطراف این نورون را مشخص می‌کند. σ_x^2 نیز بخشی از واریانس نورون x است که از پاسخ بقیه نورون‌ها مستقل است. هدف از اتصالات افقی، یادگیری واریانس پاسخ نورون می‌باشد و واریانس هر نورون به صورت تکراری و از روی پاسخ‌های گذشته آن به دست می‌آید و وزن‌های اتصالات افقی را به نحوی تغییر می‌کند که حاصل رابطه ۳-۴ به سمت این تخمین حرکت کند. مقدار تخمینی این دو مؤلفه واریانسی به صورت تکراری محاسبه می‌شود و به ازای داده‌های ورودی زیاد به مقدار واقعی نزدیک می‌شود. از آنجا که هدف از وزن‌های افقی، تخمین واریانس است، میانگین خطای مربعی (MSE) بین واریانس نورون و واریانس همسایگان آن باید کمینه

شود که برای اینکار از روش شیب نوردی نزولی استفاده می شود و مقادیر وزنهای جانبی، واریانس نوروونها و وزنهای بین لایه ای برورسانی می شود.

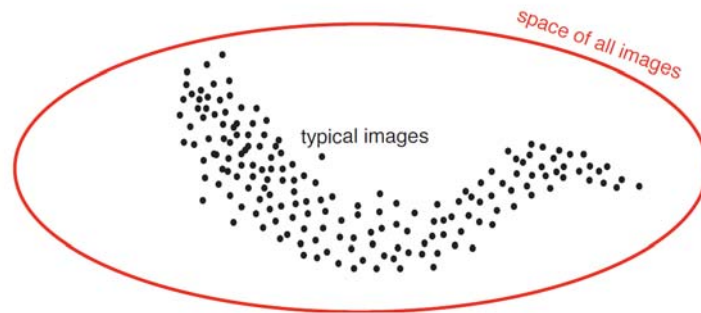
این مدل پدیده هایی مانند ساختار توپوگرافیک لایه V1، جمع کردن کانتور در اثر اتصالات جانبی V1، گزینندگی کانتور و بافت در لایه V2، گزینندگی زاویه در V2، کد کردن کانتور در لایه V4، و تشخیص اشیا در زوایای دید و نورپردازیهای گوناگون را توجیه می کند.

۳-۱-۴- مدلهای بالا به پایین

۳-۱-۴-۱- سیر تکاملی

مجموعه تمامی تصاویری که ممکن است وجود داشته باشند بسیار گسترده است ولی احتمال برخورد یک دستگاه تصویربرداری مانند چشم انسان به آنها یکسان نیست. از نخستین روزهای بررسیهای الکتروفیزیولوژیکی پردازش نورونی ورودیهای تصویری، فرضیه ای شکل گرفته است که طبق آن نوروهای سیستم بینایی اولیه، تصاویر طبیعی را طوری تجزیه می کند که گویی به جنبه های ساختار احتمالی آن حساس است (Attneave 1954; Barlow 1961; Simoncelli and Olshausen 2001). چنین آمارگانی در قلب روشهای کارا و موثر پردازش تصویر و کدگذاری آنها نهفته است (Schwartz, Pillow et al. 2006). می دانیم که دانش این توزیع غیریکسان در فضای تصویر توسط سیستم بینایی بیولوژیکی بدست آمده است و می تواند در بسیاری از کاربردهای پردازش تصویر و بینایی ماشین مفید واقع شود. به عنوان مثال، اگر ساده انگارانه به قضیه نگاه کنیم هنگامی که یک نفر به تصویری نگاه کند که آلوده به نوعی از نویز باشد، فرایند تخمین منبع اصلی تصویر را می توان به صورت بدست آوردن محتملترین تصویری دانست که به مشاهده نویزی شبیه باشد. مسئله فشرده سازی نیز غالباً به صورت استفاده از کسر بیشتری از بیتهای موجود برای کدگذاری ناحیه هایی از فضای تصویر که محتملترند دیده می شود، و مسائلی مانند بهبود تفکیک پذیری تصویر یا سنتز تصویر شامل مرحله ای هستند که در آن یک تصویر با احتمال بالا از یک توزیع انتخاب می شود (یا نمونه برداری می شود) که این انتخاب ممکن است مقید به شرایط ویژه ای نیز باشد.

برای تهیه یک مدل احتمال برای تصاویر مرئی، ابتدا باید تصمیم بگیریم که چه تصاویری را مدل کنیم. از دید کاربردی، این تصمیم گیری شامل انتخاب (الف) شرایط تصویربرداری مانند دیدگاه، تفکیک پذیری، ویژگیهای غیرخطی حسگرها و عملیات پس پردازشی و (ب) نوع صحنه و شرایط نورپردازی آن است که در تصویر ثبت می شود. اگر کسی به چنین مدلهایی برنخورده باشد ممکن است ایده اینکه تمام تصاویر از ظرف واحد احتمالاتی استخراج شده اند (شکل ۳-۷) کمی عجیب به نظر برسد مخصوصاً اینکه ویژگیها و خصوصیات هر تصویر دلخواهی مخصوص خودش است. به عنوان مثال، تصاویر طبیعی فضای طبیعی دارای ساختارهایی کاملاً متفاوت با خیابانهای یک شهر است و هر دوی آنها نیز به نوبه خود با چهره انسان متفاوتند.



شکل ۳-۷: مجموعه عام تمام تصاویر طبیعی. می توان این مجموعه را به صورت قطعی تعریف کرد. ولی این عضویت نرم^{۱۵۲} است و بهتر است برای توصیف آن از احتمال استفاده کرد.

دو وسیله برای فائق آمدن بر این معضل وجود دارد: نخست خصوصیات آماری که آنقدر پایه ای و ابتدایی باشند که لزوماً برای تمام صحنه های طبیعی به کار بیایند. دوم مدل‌های پارامتری که در آن مجموعه ای از اَبَرپارامترها^{۱۵۳} (که خود آنها احتمالاً متغیرهای تصادفی هستند) جزئیات رفتاری مدل را هدایت کنند و بنابراین درجه خاصی از تطبیق پذیری مدل را با منابع تولید تصویر فراهم کند. حال چگونه می توان یک مدل احتمالاتی برای تصاویر تولید کرد و آزمون؟ روشهای بسیاری برای اینکار بوجود آمده است ولی در اینجا برای سهولت فهم یک روش تجربی برای تصاویر گسسته سازی شده (پیکسلی) شرح داده می شود. دوربینهای دیجیتالی که امروزه یافت می شوند عکسهایی می گیرند که معمولاً حاوی میلیونها پیکسل هستند. استفاده از هیستوگرام تصاویر کاری بهبود یافته است زیرا اطلاعات بسیار کمی درباره محتوای تصویر دربر دارد. همانطور که در شکل ۳-۸ دیده می شود تصاویری با محتوای یکسان ممکن است دارای هیستوگرامهایی متفاوت باشند و بر عکس تصاویری دارای محتوای کاملاً متفاوت ممکن است هیستوگرامهای یکسانی داشته باشند.

اولین ایده خامی که به ذهن متبادر می شود بررسی مجموعه بزرگی از تصاویر به منظور بدست آوردن توزیع آنهاست. ولی در نگاه مجدد به این راه حل می توان به بی فایده بودن آن پی برد. میزان داده هایی که برای تخمین یک توزیع احتمال از نمونه های آن لازم است، با رشد ابعاد فضا (در اینجا تعداد پیکسلها) به طور نمایی افزایش می یابد، که این پدیده به طلسم تعدد ابعاد^{۱۵۴} معروف است.

بنابراین برای مدل کردن تصاویر، کاهش ابعاد فضای تصویر الزامی است. برای نیل به این مقصود دو نوع فرض برای ساده سازی استفاده می شود (Simoncelli 2005). فرض اول که به فرض مارکوف^{۱۵۵} معروف است، عبارتست از اینکه توزیع چگالی یک پیکسل هنگامی که روی پیکسلهای همسایگی مکانی کوچکی اعمال می شود مستقل از توزیع پیکسلهای خارج از همسایگی است. فرض دوم برای ساده سازی از تحمیل تقارنها با نامتغیریهای خاصی به ساختار احتمال بدست می آید. رایجترین مورد این نامتغیریها، نامتغیری به ازای جابجایی است (که همگنی^{۱۵۶} یا پایداری اکید حسّی^{۱۵۷} نیز نامیده می شود)

¹⁵² soft

¹⁵³ hyper parameters

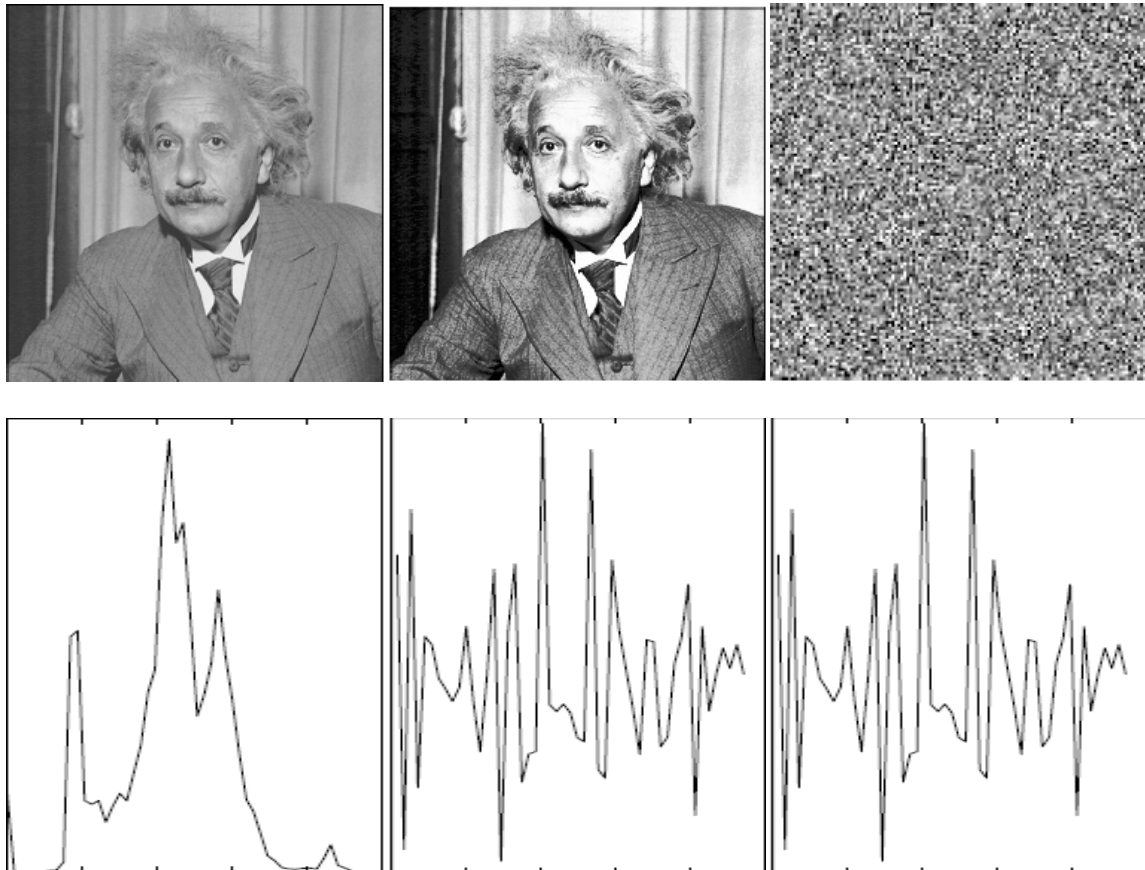
¹⁵⁴ curse of dimensionality

¹⁵⁵ Markov assumption

¹⁵⁶ homogeneity

¹⁵⁷ strict sense stationary

که در آن توزیع پیکسلها در یک همسایگی به موقعیت مطلق آن همسایگی درون یک تصویر وابسته نمی باشد. این مسئله به نظر منطقی می رسد زیرا با داشتن یک جابجایی افقی یا عمودی دوربین، تنها شدت روشنایی پیکسلها در آرایه پیکسلها شیفت پیدا می کند. باید توجه داشت که نامتغیری به ازای جابجایی در مرزها خوش تعریف نیست و معمولاً در پردازش تصویر این مناطق توجه ویژه ای را می طلبند.



شکل ۳-۸: هیستوگرام تصاویر، اطلاعات بسیار کمی درباره محتوای تصویر دربر دارد.

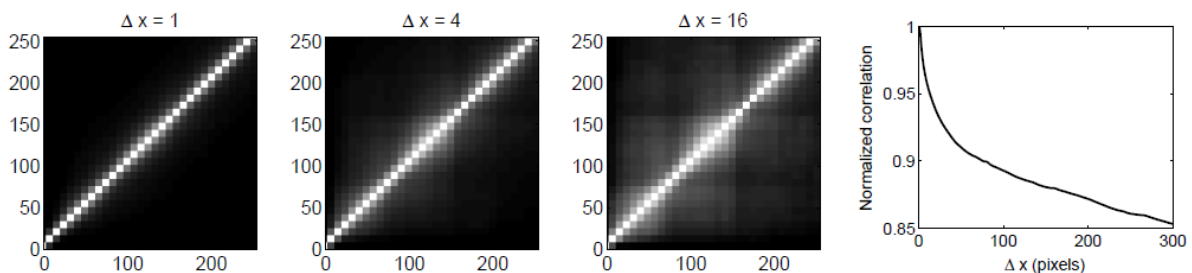
فرض رایج دیگری از این دست، نامتغیری با مقیاس است: تغییر اندازه تصویر ساختار احتمالات را تغییری نمی دهد. این فرض نیز بطور غیر رسمی اینگونه توجیه می شود که تنظیم فاصله کانونی دوربین (زوم) و یا لنز آن (جدا از اعوجاجات پرسپکتیوی) اندازه تصویر را تخمین می زند. همانند نامتغیری به ازای جابجایی، نامتغیری به ازای مقیاس نیز در مرزهای خاصی دچار مشکل می شود. نامتغیری به ازای مقیاس برای تصاویر گسسته شده با ابعاد ریزی نزدیک به اندازه پیکسل و یا ابعاد درشتی نزدیک به اندازه کل تصویر نیز دچار مشکل می شود.

با داشتن این فرضهای ساده سازی ساختاری، می توان مجدداً به مسئله تولید یک مدل احتمالاتی پرداخت. در سالهای اخیر محققان زمینه های مختلفی نظیر پردازش تصویر، بینایی رایانه ای، فیزیک، ریاضیات کاربردی و آمار طیف وسیعی از مدل‌های گوناگون را پیشنهاد داده اند. در ادامه به برخی از خصوصیات آماری ابتدایی تصاویر عکسبرداری شده و مدل‌هایی که از آنها بهره برده اند پرداخته می شود. اشاراتی به اینکه چطور درستی این مدلها با بررسی نحوه تطبیق آنها به

داده‌ها نشان داده می‌شود ارائه می‌شود ولی آزمون حقیقی زمانی صورت می‌گیرد که مدل برای یکی از کاربردهای چالش برانگیز پردازش تصویر مانند فشرده‌سازی یا رفع نویز استفاده شود.

برای حفظ تمرکز بحث تنها به تصاویر عکسبرداری شده، گسسته و خاکستری فام پرداخته می‌شود. بسیاری از اصول به سادگی قابل تعمیم به عکسهای رنگی (Buchsbbaum and Gottschalk 1983; Ruderman, Cronin et al. 1998) و یا دنباله‌ای از تصاویر موقتی (فیلم) (Dong and Atick 1995) و یا تصاویر تخصصی تر مانند پرتره‌ها، منظره‌ها، و بافتها نیز هستند. علاوه بر آنها روشهای کلی می‌توانند به تصاویر غیر مرئی مانند تصاویر پزشکی، تصاویر فرسوخ، رادار و سایر انواع تصاویر بُعد، و تصاویر نجومی اعمال شوند.

مدل گوسی: مدل کلاسیک آمارگان تصویر توسط مهندسان تلویزیون در دهه ۱۹۵۰ تشکیل شده است (برای مرور ر.ک. به (Ruderman and Bialek 1994)) که به بازنمایی و انتقال بهینه سیگنال علاقه مند بودند. ابتدایی ترین انگیزه برای تولید این مدل از این مشاهده بدست آمده بود که پیکسلهای مجاور معمولاً مقادیر شدت روشنایی مشابهی نیز دارند. این حقیقت براحتی با اندازه گیریهایی نظیر آنچه در شکل ۳-۹ (چپ) مشاهده می‌شود تایید می‌شود. هر نمودار پراکندگی مقدار یک جفت پیکسل با تفاوت مکانی داده شده را بدست می‌دهد. آنچه در این اندازه گیریها نهفته است همان فرض همگنی است که به آن اشاره شد: توزیعها از مکان مطلق درون تصویر مستقلند. با وجود اینکه پیکسلها از عکس یکسانی برداشته شده اند ولی بازهم نمایندگان خوبی برای آنچه در اغلب تصاویر عکسبرداری شده دیده می‌شود هستند.



شکل ۳-۹: (چپ) نمودار پراکندگی جفتهای پیکسل در سه جابجایی مکانی مختلف از پنج تصویر نمونه نشان می‌دهد. (راست) تابع خود همبستگی. تصاویر صحنه های خیابانهای شهر نیویورک که با دوربین دیجیتال Canon 10D گرفته شده است و در حالت سنسور خطی RAW پردازش شده است (که شدت روشنایی تصویر را تقریباً متناسب با شدت روشنایی نوری بدست می‌دهد). همبستگیها روی لگاریتم مهادیر شدت این سنسورها گرفته شده است (Ruderman and Bialek 1994)

آنچه در این نمودارها بخوبی خودنمایی می‌کند اینست که مقادیر پیکسلها بسیار همبسته اند: هنگامی که شدت روشنایی یک پیکسل بزرگ باشد دیگری نیز متمایل است که بزرگ باشد. ولی این همبستگی با زیاد شدن فاصله بین پیکسلها افت می‌کند. این رفتار در شکل ۳-۹ (راست) خلاصه شده است که خودهمبستگی (همبستگی پیکسلها به عنوان تابعی از فاصله آنها) را نشان می‌دهد. آمارگان همبستگی نشان داده شده در شکل ۳-۹ قید محکمی بر روی ساختار تصاویر قرار می‌دهد ولی یک مدل احتمالاتی کامل را فراهم نمی‌آورد؛ بخصوص با وجود توزیعهای احتمالاتی بسیاری که ساختار همبستگی (یا هم ارز آن کواریانس) مشابهی دارند. اکنون سوال اینجاست چگونه یک مدل را از این مجموعه انتخاب کرد؟ یک معیار

طبیعی، انتخاب توزیعی با بیشترین آنتروپی با حفظ محدودیت کواریانس است (Jaynes 78). حل کردن مسئله برای این توزیع نسبتاً راحت است و به یک توزیع گوسی چند متغیره می انجامد:

$$p(\vec{x}) \propto \exp\left(-\frac{1}{2}\vec{x}^T C_x^{-1} \vec{x}\right) \quad (5-3)$$

که در آن \vec{x} برداری حاوی تمام پیکسلهای تصویر (که برای سهولت با میانگن صفر در نظر گرفته شده اند) و $C_x = E\vec{x}\vec{x}^T$ ماتریس کواریانس است. توزیعهای گوسی در مختصرترین حالت خود با یک دوران به دستگاه مختصاتی بیان می شوند که ماتریس کواریانس آنها قطری باشد. با استفاده از جبر خطی استاندارد این هدف تامین می شود:

$$\vec{y} = E^T \vec{x} \quad (6-3)$$

که در آن E ماتریس متعامد حاوی بردارهای یکه C_x است به طوری که:

$$C_x = EDE^T \Rightarrow E^T C_x E = D \quad (7-3)$$

که در آن D یک ماتریس قطری است که حاوی مقادیر یکه متناظر است. هنگامی که توزیع احتمالاتی \vec{x} پایدار باشد (با فرض حل کردن تناوبی مشکل مرزها) ماتریس کواریانس C_x ، حلقوی خواهد بود. در این حالت خاص می دانیم که تبدیل فوریه یک تبدیل قطری ساز برای ماتریس E خواهد بود و حصول رابطه معرفی شده در معادله 3-7 را تضمین می کند. برای اینکه مدل تصویر گوسی کامل شود، تنها باید ورودیهای ماتریس قطری D را بدانیم که با مؤلفه های فرکانسی تبدیل فوریه متناظر است.

دو راه برای حصول یک جواب وجود دارد. نخست با نادیده گرفتن خطاری که زودتر داده شد، می توان فرض کرد که آمارگان تصویر به ازای مقیاس نامتغیرند. مخصوصاً باید فرض کرد که خواص آمارگان درجه دوم (کواریانس) تصویر نسبت به تغییر اندازه تصویر نامتغیر است. می توان نامتغیر بودن به ازای مقیاس را در دامنه فوریه چنین نشان داد:

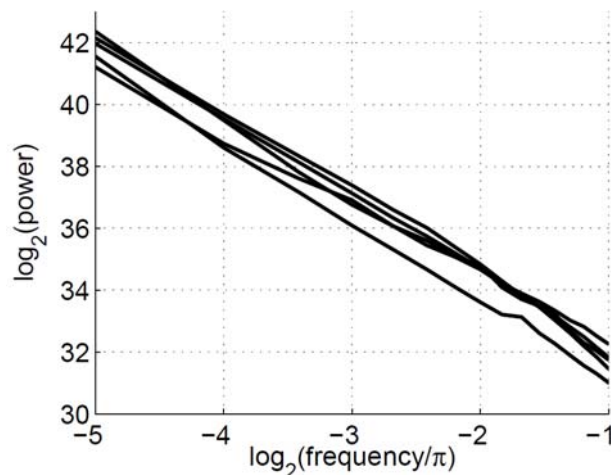
$$\mathbb{E}\left(|F(s\omega)|^2\right) = h(s)\mathbb{E}\left(|F(\omega)|^2\right), \forall \omega, s \quad (8-3)$$

بدین معنی که تغییر اندازه محور فرکانس شکل تابع را تغییر نمی دهد و تنها طیف را در یک ثابت ضرب می کند. تنها تابعی که یکسانی را تامین می کند قوانین توانی^{۱۵۸} است:

$$\mathbb{E}\left(|F(\omega)|^2\right) = \frac{A}{\omega^\gamma} \quad (9-3)$$

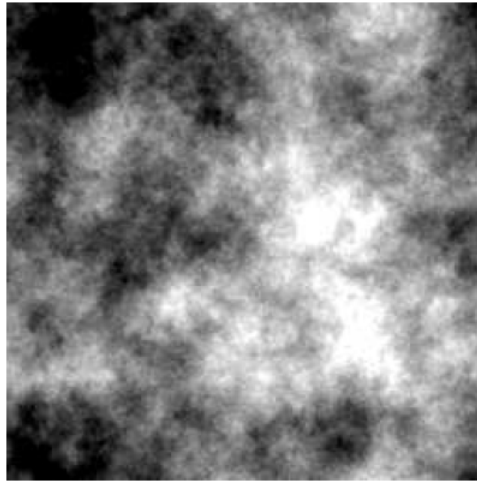
که توان آن نرخ نزول طیف را کنترل می کند. بنابراین فرضهای دوگانه نامتغیری به ازای جابجایی و مقیاس ساختار کواریانس تصاویر را به مدلی تنها با یک پارامتر مقید می کند. از سوی دیگر شکل طیف توان را می توان بطور تجربی بدست آورد

(Deriugin 1957; Field 1987; Tolhurst, Tadmor et al. 1992; Ruderman and Bialek 1994; Van der Schaaf and Van Hateren 1996). برای بسیاری از تصاویر متداول، طیف توان به کمک قانون توان به خوبی تخمین زده می شود و بر درستی خصوصیت نامتغیری به ازای مقیاس در آمارگان مرتبه دوم صحنه می گذارد. در این اندازه گیریهای تجربی، مقدار توان معمولاً نزدیک به دو است. مثالهایی از تخمین طیف توانی برای چند تصویر نمونه در شکل ۳-۱۰ آمده است. همچنین نشان داده شده است که نامتغیری به ازای مقیاس برای آمارگانی بجز طیف توانی نیز برقرار است (Ruderman and Bialek 1994; Turiel and Parga 2000). مدل طیفی همان مدل کلاسیک پردازش تصویر است. افزون بر در نظر گرفتن طیف تصاویر متداول، سادگی شکل گوسی به راه حلهایی برای فشرده سازی و رفع نویز تصویر منجر شده است که از ملزومات اولیه پردازش تصویر است.



شکل ۳-۱۰: تخمین طیف توانی برای پنج تصویر نمونه (که در شکل ۳-۹ توصیف شده بود) به عنوان تابعی از فرکانس مکانی که روی راستا میانگین گیری شده باشد. این نمودارها با یک قانون توانی که دارای نمایی اندکی بزرگتر از ۲ باشند بخوبی تخمین زده می شوند.

علیرغم سادگی و قابل محاسبه بودن مدل گوسی، می توان به راحتی مشاهده نمود که مدل توصیف نسبتاً ضعیفی ارائه می دهد. از دید دقیق می توان دریافت که با وجود اینکه مدل بزرگی ضرایب فوریه را قویاً محدود می کند ولی هیچ محدودیتی روی فاز آنها قرار نمی دهد. برای روشنی این موضوع، شکل ۳-۱۱ تصویری را نشان می دهد که فاز آن در سراسر تصویر به طور تصادفی انتخاب شده است و ظاهر تصویر به کلی از بین رفته است (Oppenheim and Lim 1981).



شکل ۳-۱۱: تصویری نمونه که بطور تصادفی از مدل طیفی گوسی انتخاب شده است.

به عنوان یک آزمون مستقیم می توان یک تصویر از توزیع را با تولید نویز سفید در حوزه فوریه ایجاد کرد، هر نمونه را با وزن مناسب $1/\omega^2$ مقدار دهی کرد و سپس با تبدیل معکوس به حوزه تصویر برگرداند. مثالی از چنین تصویری در شکل ۳-۱۱ دیده می شود. در واقع این آزمایشها که هر بار تصاویری ابر مانند تولید می کند نشانگر اینست که محدودیت کواریانس برای اخذ ساختار غنی تری از ویژگیهایی که در اکثر تصاویر طبیعی یافت می شوند ناکافی هستند.

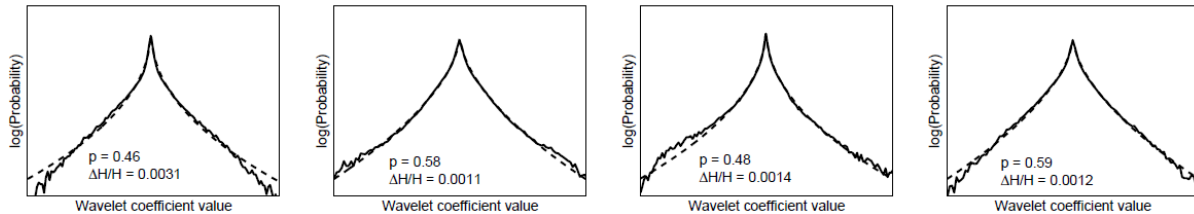
مدلهای توابع پایه خطی: برای دهه ها، ناقص بودن مدل گوسی مسئله ای آشکار و آزاردهنده بود ولی هرگونه پیشرفتی در زمینه مقید کردن فازهای فوریه به نظر بسیار دشوار می آمد. رابطه بین مؤلفه های فاز به آسانی اندازه گیری نمی شود که این امر از سویی بخاطر دشواری کار کردن با آمارگان توأم متغیرهای حلقوی است و از سوی دیگر به دلیل اینست که وابستگی بین فاز فرکانسهای متفاوت به نظر نمی آید که بخوبی توسط مدلی که خود از فرکانس محلی استفاده می کند مدل شود (Simoncelli 2005).

در دهه ۱۹۸۰ یک جهش عظیم با اقبال نویسندگان به شرح ویژگیهای غیرگوسی در تصاویر رخ داد. یک مدل آماری گوسی چند متغیره دارای این خاصیت است که تمام توزیعهای شرطی و حاشیه ای آن نیز گوسی است. ولی این نویسندگان متوجه شدند که هیستوگرام تصاویر طبیعی پس از عبور از فیلتر میان گذر شدیداً غیرگوسی هستند (Burt and Adelson 1983; Adelson and Bergen 1985; Daugman 1988; Mallat 1989; Zetsche 1990). توزیع فعالسازی تک تک فیلترهای خطی هنگامی که با تصویر کانالو می شود دارای کورتوسیس بالایی است؛ بدین معنی که پاسخ فیلتر اغلب صفر است ولی ندرتاً فیلتر به ساختارهای خاصی در تصویر پاسخی قوی می دهد (Field 1987). این توزیعهای حاشیه ای در حوالی صفر قله بسیار تیزتری دارند و دنباله^{۱۵۹} های آنها گسترده تر از توزیعی گوسی با واریانس مشابه است. مثلاً شکل ۳-۱۲ هیستوگرام سه تصویر را نشان می دهد که با تابع گابور^{۱۶۰} (یک شبکه سینوسی محدود شده با توزیع گوسی) فیلتر شده اند. توجه منطقی این رفتار اینست که تصاویر معمولاً دارای نواحی همواری هستند که توسط برخی از ویژگیهای محلی مانند خطها، لبه ها و یا گوشه ها از هم جدا شده اند. این نواحی هموار به پاسخ فیلتری کوچکی منجر می

¹⁵⁹ tails

¹⁶⁰ Gabor function

شوند که قله تیزی را در صفر ایجاد می کنند و ویژگیهای محلی پاسخهایی با بزرگی زیاد تولید می کنند که سازنده دنباله های کشیده شکل توزیع هستند. این رفتار ابتدایی الزاماً برای هر فیلتر میان گذری اعم از بدون جهت (حوالی یک مرکز) یا جهتدار صادق است ولی گاهی اوقات به پاسخهایی که از دیگران غیرگوسی تر هستند می انجامند.



شکل ۳-۱۲: هیستوگرام لگاریتمی یک زیرباند موجک برای چهار تصویر نمونه (که در شکل ۳-۹ توصیف شده بود) برای هر هیستوگرام، دنباله ها طوری قطع شده اند که ۹۹.۸٪ توزیع نسان داده شود. خط چینها توزیعهای گوسی تعمیم یافته هستند که به نمودار تطبیق داده شده اند و به طور کلی توسط رابطه ۳-۱۱ توصیف می شوند. نوشته های زیر هر هیستوگرام مقدار بیشترین درستنمایی p را نشان می دهد که برای تطبیق نمودار استفاده می شود و نیز آنتروپی مربوط به مدل و هیستوگرام (افتراق کولبک-لیبلر آنها) را به عنوان کسری از کل آنتروپی هیستوگرام نشان می دهد.

در دهه ۱۹۹۰ محققان بیش از پیش به تبدیلات خطی برای توصیف تصاویر طبیعی روی آوردند. یکی از تبدیلاتی که همبستگی بین داده ها را از بین می برد تحلیل مؤلفه های اصلی است. معیار اصلی در این تحلیل، ضریب همبستگی می باشد که یک معیار آماری درجه دوم است. در تحلیل مؤلفه های پایه، هدف یافتن نگاشتی از داده های موجود به داده هایی است که ضریب همبستگی بین آنها صفر باشد. در تحلیلی که بر روی تصاویر طبیعی انجام شده بود مؤلفه هایی به دست آمده بود که شباهت چندانی به گزینندگی نورون های $V1$ نداشتند.

یکی از مشکلات تحلیل مؤلفه های پایه در این است که مؤلفه های بدست آمده باید بر هم عمود باشند. مشکل دیگر از آنجا ناشی می شود که تحلیل مؤلفه اصلی تنها به ناهمبسته سازی داده های ورودی می پردازد و به وابستگی های مراتب بالاتر کاری ندارد. به بیان ساده، ناهمبسته سازی یعنی اینکه متغیرها را با استفاده از یک پیش بینی کننده خطی نتوان از روی هم پیش بینی کرد و این تعریف به طور ضمنی امکان وابستگیهای غیر خطی را تأیید می کند. مثلاً دو متغیر x و y را در نظر بگیرید که با یک رابطه سینوسی، $y = \sin(x)$ ، به هم مرتبطند. ضریب همبستگی برای این دو متغیر صفر خواهد بود در حالیکه این متغیرها وابستگی زیادی دارند. لبه ها، که همان همراستایی فاز در چندین مقیاس مکانی هستند، مثالی از وابستگی مرتبه بالا در یک تصویر هستند. از این قبیل می توان به شکل و انحناء نیز اشاره کرد. آماره های مرتبه دوم طیف درشتی^{۱۶۱} یک تصویر -و نه فاز آنرا- در بر دارد. درشتی یک آماره مرتبه دوم است و طیف درشتی یک تصویر لزوماً یک سری همبستگی با موجهای سینوسی است. همچنین تبدیل فوریه تابع خود-همبستگی یک سیگنال برابر با طیف توان (مربع طیف درشتی) آن است. اطلاعاتی که توسط تابع خود-همبستگی اخذ نشده اند، یعنی همان آماره های مراتب بالاتر، به طیف فاز مربوطند (Bartlett 1998).

مکانیسمهای کدگذاری که به فاز حساسند برای تولید یک سیستم ادراکی مهم هستند. فاز مکانی حاوی اطلاعات ساختاری است که نسبت به طیف درشتی، سیستم شناسایی انسان را بسیار قویتر هدایت می کند. مثلاً چهره ای که از طیف درشتی چهره اول و طیف فاز چهره دوم ساخته شده باشد، به عنوان چهره فرد دوم درک می شود (Bartlett 1998).

در اواسط دهه ۱۹۹۰ تعدادی از نویسندگان روشهایی برای بهینه سازی پایه ای از فیلترها جهت بهینه سازی غیرگوسی بودن پاسخها تولید کردند (Bell and Sejnowski 1997; Olshausen and Field 1997). تحلیل مؤلفه های مستقل و کدگذاری پراکنده از جمله روشهایی بود که برای جبران نقص های تحلیل مؤلفه های اصلی گسترش یافت. این روشها معمولاً با بهینه سازی آمارگان مراتب بالا مانند کورتوسیسی (گشتاور چهارم تقسیم بر مربع واریانس) کار می کرد. در این تحلیلها، که غالباً برای مسئله جداسازی منابع به صورت ناآگاهانه به کار می روند، معیارهای آماری مرتبه های بالاتر استفاده می شوند. گسترش این تحلیل به صورت همزمان و توسط افراد مختلف و با معیارهای مختلف صورت پذیرفت (Cardoso 1989; Inouye and Matsui 1989; Gaeta and Lacoume 1990; Comon 1992).

ارتباط بین نظریه کدگذاری بهینه و بازنمایی تصویر در قشر مخ با معرفی این دو روش مدلسازی تنگاتنگ تر شد. این دو مدل، یعنی تحلیل مؤلفه های مستقل (Comon 1992; Bell and Sejnowski 1997) و کدگذاری پراکنده (Olshausen 1996; Olshausen and Field 1997) نشان می دادند که از میان تمام کدهای خطی ممکن، فیلترهایی که نماینده میدان گیرندگی سلولهای ساده هستند، برای تصاویر طبیعی انتخاب بهینه هستند.

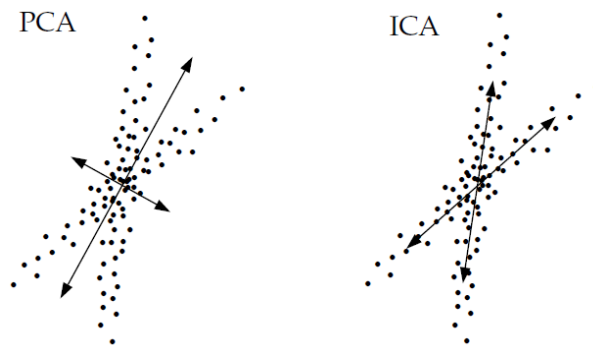
در مدل ICA هر تصویر \vec{x} (در اینجا برداری از شدت روشنایی پیکسلها) توسط یک ترکیب خطی از تابعهای پایه \vec{A}_j که با ضرایب s_j وزندهی می شوند بازنمایی می شود:

$$\vec{x} = \sum_j \vec{A}_j s_j \quad (10-3)$$

تلاش ICA بر این است که توابع پایه ای مانند \vec{A} پیدا کند که مستقلترین (و در نتیجه با کمترین افزونگی) ضرایب را داشته باشد. معمولاً تعداد توابع پایه برابر ابعاد داده است (به طوری که \vec{A} مربعی و معکوس پذیر باشد) و ضرایب از رابطه $\vec{s} = \vec{A}^{-1} \vec{x}$ محاسبه می شوند. از آنجایی که معیارهای تجربی استقلال آماری به سختی محاسبه می شوند آنها را تقریب می زنند. به عنوان مثال فرض می شود که ضرایب دارای توزیعی جداپذیر هستند $p(\vec{s}) = \prod_i p(s_i)$ و تحت چنین شرایطی توابع پایه بهینه سازی می شوند.

تحلیل مؤلفه مستقل (ICA) تعمیمی بر تحلیل مؤلفه اصلی است که علاوه بر وابستگیهای مرتبه دوم، وابستگیهای مراتب بالاتر را نیز در ورودی تفکیک می کند. همانگونه که اشاره شد، تحلیل مؤلفه اصلی یک راه کدگذاری وابستگیهای مرتبه دوم است و اینکار را با چرخاندن محورهای مختصات تا رسیدن به راستای بیشینه کواریانس انجام می دهد. مجموعه نقاطی که در شکل ۳-۱۳ دیده می شوند از دو توزیعی که در شکل مشهودند برداشته شده اند. تحلیل مؤلفه اصلی داده ها را با یک مدل گوسی چندمتغیره مدلسازی می کند و یک مجموعه محور متعامد را طوری قرار می دهد که دو توزیع کاملاً متداخل باشند. تحلیل مؤلفه مستقل الزام متعامد بودن محورها را ندارد و آنها را در راستای وابستگیهای آماری داده ها قرار می دهد. هر بردار وزن در ICA تلاش می کند که بخشی از وابستگیهای ورودی را کدگذاری کند، و لذا وابستگیهای بین عناصر

خروجی از بین می رود. تصویر کردن دو توزیع بر روی محورهای ICA دارای همپوشانی کمتری است و توزیع خروجی دو بردار وزن دارای کورتوسیس بالا خواهد بود.

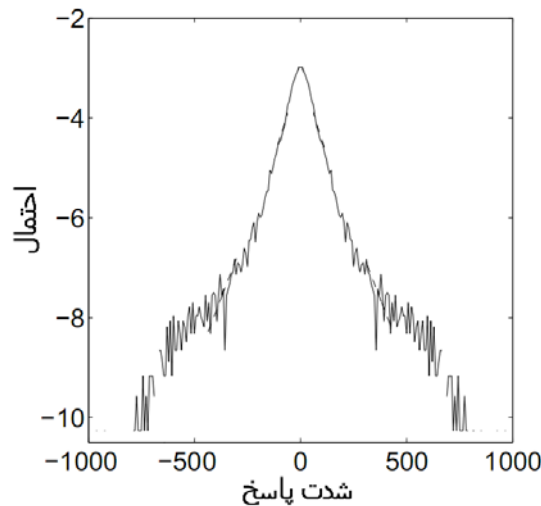


شکل ۳-۱۳: مثالی از توزیع داده های دوبعدی و محورهای تحلیل مؤلفه های اصلی (چپ) و تحلیل مؤلفه های مستقل (راست) متناظر

بل و سجنوسکی^{۱۶۲} الگوریتمی برای تفکیک مؤلفه های مستقل آماری یک مجموعه داده با استفاده از یادگیری بدون ناظر طراحی کرده اند. شالوده این الگوریتم بر اصل انتقال بیشینه اطلاعات بین نورونهای سیگموئیدی استوار است (Bartlett 1998). این الگوریتم اصل بیشینه سازی اطلاعاتی لینکسر (که در فصل ۲ به آن اشاره شد) را به حالت چند خروجی تعمیم می دهد و آنترپی توأم واحدهای خروجی را بیشینه می کند. راه دیگری برای بیان تفاوت بین PCA و ICA اینست که بگوییم PCA واریانس توأم خروجیها را بیشینه می کند درحالیکه ICA آنترپی توأم خروجیها را بیشینه می سازد. الگوریتم بل و سجنوسکی ماتریس وزن A^{-1} را با روش شیب نوردی صعودی روی آنترپی توأم \bar{y} یک تابع انتقال غیرخطی داده های ورودی، g ، $\bar{y} = g(A^{-1}\bar{x})$ ، پیدا می کند. بیشینه کردن آنترپی توأم خروجی هم ارز بیشینه کردن اطلاعات متقابل بین ورودی و خروجی (یعنی بیشینه کردن انتقال اطلاعات) است. دلیل این امر اینست که $H(\bar{y}|\bar{x}) = H(\bar{x}) + H(\bar{y}) - H(\bar{y}|\bar{x})$ ، که در آن تنها $H(\bar{y})$ به ماتریس وزنها A^{-1} مرتبط است چون $H(\bar{y}|\bar{x})$ نیز است. بیشینه کردن آنترپی توأم خروجی، اطلاعات مشترک بین تک تک عناصر را تشویق به کوچک بودن می کند (ر.ک. فصل ۲). هنگامی که تابع انتقال غیرخطی g با توزیع تجمعی سیگنالهایی که مربوط به داده های \bar{x} هستند برابر باشد، اطلاعات مشترک حتی با وجود مقیاس بندی و جابجایی به طور تضمین شده ای کمینه می شود (Bartlett 1998).

این امر در بادی امر متناقض می نماید، ولی در واقع بیشینه سازی اطلاعات در یک کد چندبُعدی با نظریه بارلو درباره کدگذاری کمینه آنترپی سازگار است. چنان که گفته شد بیشینه کردن آنترپی توأم خروجیها باعث کوچک شدن اطلاعات متقابل بین خروجیها می شود، ولی بعضی اوقات راه حلهای دیگری نیز موجود است که در آن اطلاعات متقابل غیرصفر است. اگر بدانیم که آنترپی توأم ثابت است (و در بیشینه مقدار خود قرار دارد)، راه حلی که اطلاعات متقابل را کمینه می کند، آنترپی های جانبی (تک تک) هر واحد خروجی را نیز کمینه می کند.

باید توجه شود که کاهش افزونگی کد نورونی تنها یکی از اصلهایی است که می تواند در پردازش نورونی مد نظر قرار گیرد. یکی دیگر از اصول شناخته شده کد کردن اطلاعات در قشرهای حسی کد کردن پراکنده^{۱۶۳} است. کد کردن پراکنده برای توضیح رفتار نورون های قشر حسی مغز ارائه شده است و بر مطالعات عصب شناسی استوار است. بر اساس نظریه کد کردن پراکنده، پاسخ یک نورون در اکثر مواقع صفر می باشد. در بقیه موارد پاسخ نورون از نظر مقداری بزرگ می باشد. شکل ۳-۱۴ نمودار توزیع پاسخ یک نورون نمونه را نشان می دهد.



شکل ۳-۱۴: نمودار احتمال پاسخ یک نورون بر اساس بزرگی پاسخ. در این نمودار پاسخ صفر بیشترین احتمال را دارا می باشد. همچنین احتمال پاسخ های بزرگ در نورون نیز بیشتر از توزیع گوسی است (Wainwright and Simoncelli 2000). شباهت پاسخ نورون به پاسخ فیلترهای شکل ۳-۱۲ الهام بخش است.

همانطور که از این شکل مشخص است، پاسخ نورون در اکثر مواقع صفر است. اما در مواردی که نورون پاسخ می دهد، مقدار پاسخ بزرگ می باشد. بر اساس این مشاهدات، فرضیه کد کردن پراکنده بیان می دارد که قشرهای حسی مغز اطلاعات را به صورت پراکنده و با تعداد کمی نورون فعال نشان می دهند و با نظریه کد کردن بهینه همراستاست. این فرضیه از نظر زیستی دارای مزایایی مانند حفظ کردن انرژی می باشد و با مشاهدات عصب شناسی تطابق دارد.

اتیک^{۱۶۴} کدهای فشرده و ناهمبسته ای مانند PCA را به دلیل کارایی کدگذاری می پسندید (Atick 1992). فیلد^{۱۶۵} کدهای پراکنده را برای چنین منظوری ترجیح می دهد (Field 1994). کدهای پراکنده دارای توزیع پاسخی با کورتوسیس بالا هستند که تجمع اکثر مقادیر در نزدیکی صفر است و مقادیر بزرگ مثبت و یا منفی نیز در دنباله های آن ندرتاً رخ می دهند. بنابراین بیشینه کردن پراکنده توزیع یک پاسخ معادل کمینه کردن آنتروپی است و کدهای پراکنده مزایایی مشابه کدهای آنتروپی دارند که از آن جمله می توان به تفکیک افزونگی های مراتب بالا و نیز افزونگی مرتبه دوم اشاره کرد. در چنین کدهایی، افزونگی بین عناصر ورودی به افزونگی درون الگوهای پاسخ تک تک ورودیها تبدیل می شوند که بدینسان تک تک خروجی ها، مگر در موارد معدودی، پاسخ یکسانی می دهند.

¹⁶³ Sparse coding

¹⁶⁴ Atick

¹⁶⁵ Field

با داشتن این رابطه بین کدهای آنتروپی و کدهای پراکنده، مزایای کدهای پراکنده که در (Field 1994) به آن اشاره شده است نیز در جهت تائید کدهای آنتروپی کمینه بارلو خواهد بود. کدهایی که تعداد نورونهای فعال را کمینه کنند برای تشخیص هم وقوعیهای مشکوک می توانند مفید واقع شوند. از آنجا که پاسخ غیرصفر هر واحد - با توجه به توزیع آن - نسبتاً متوقع نیست، روابط مراتب بالا بطور فزاینده ای کمیاب می شوند، و نتیجتاً هنگامی که در محرک ظاهر شوند اطلاعات زیادی را در بر خواهند داشت. فیلد تبیین این روش را نسبت به کدهای فشرده ای مانند تحلیل مؤلفه اصلی نشان داده است که در این روشها تعداد اندکی از سلولها دارای احتمال پاسخ زیادی هستند، و بنابراین ترکیبات مرتبه بالا در این گروه نسبتاً متداول است. در یک کد پراکنده توزیع شده، بازنمایی اشیای مختلف با صرف فعال شدن واحد، و نه میزان فعالیت آن نشان داده می شود. این بازنمایی از لحاظ نرخ سیگنال به نویز مزیت بیشتری دارد زیرا بدون توجه به میزان دقیق فعالیت، تنها لازم است مشخص شود که کدام واحدها فعال شده اند. مزیت دیگر کدگذاری پراکنده (از جمله برای بازنمایی چهره) قابلیت ذخیره آن در حافظه های انجمنی است. شبکه هایی با ورودی پراکنده حافظه بزرگتری دارند و بازیابی آنها با قسمتی از اطلاعات کارا تر است. فیلد شواهدی را ارائه کرده است که فیلترهای گابور جهتدار هنگامی که روی تصاویر طبیعی اعمال شوند کد پراکنده تولید می کنند و اگر روی تصاویر ساختگی که از نویز تشکیل شده اند اعمال شوند توزیع پاسخ آنها، توزیع گوسی خواهد بود. از آنجایی که این دو کلاس طیف درشتی مشابهی دارند و تنها در فاز متفاوتند، فیلد نتیجه گرفت که کدگذاری پراکنده با فیلترهای گابور اساساً به طیف فاز داده ها وابسته است (Bartlett 1998). اولشاوزن و فیلد^{۱۶۶} یک شبکه را برای بازسازی تصاویر طبیعی از روی ترکیب خطی تصاویر پایه مجهول آموزش دادند (Olshausen 1996). معیار مورد استفاده در این روش، کمینه میانگین خطای مربعی است که به تنهایی می تواند به یک ترکیب خطی از مؤلفه های اصلی تصویر همگرا شود. هنگامی که یک معیار پراکنده به تابع هدف اضافه شود، تصاویر پایه مشابه خصوصیات پاسخ سلولهای ساده V1 محلی هستند، راستای خاصی دارند، و از لحاظ مکانی متضادند^{۱۶۷}. بر اساس این نظریه، تحقیقی دیگر انجام شد (Olshausen 1996) که در آن، سعی شد با استفاده از نگاشتی خطی، نمایشی از تصاویر طبیعی به دست بیاید که تا حد امکان پراکنده باشد. برای این منظور، معیاری تعریف شد که میزان پراکندهی فعالیت ضرایب حاصل از نگاشت خطی را اندازه گیری می کرد. بهینه سازی این معیار در تصاویر طبیعی توابع پایه ای را نتیجه داد که شباهت بسیاری به گزینندگی نورون های V1 داشتند. مدل کدگذاری پراکنده نویز افزایشی گوسی را نیز در مدل خود گنجانده است، $\vec{x} = A\vec{s} + \epsilon$ ، و تعداد ضرایب کد کننده در آن (J در مدل ICA) محدود به ابعاد ورودی نیست. اگر نویز صفر نباشد یا J بزرگتر از ابعاد داده ورودی باشد (حالت فراکامل^{۱۶۸})، فعالیت نورونها دیگر تابعی قطعی از داده ها نمی باشد زیرا حالتها ی زیادی برای تولید یک مجموعه داده متصور است و به کمک استنتاج باید بهترین آنها را پیدا کرد. این امر مستلزم بکارگیری تکنیکهای تقریب برای یادگیری پایه های بهینه است و محاسبه احتمال ثانویه $p(\vec{s} | A, \vec{x})$ حالت ضرایب را دشوار می کند ولی از سوی دیگر به مدل امکان کار کردن با داده های حسی و تعداد متغیر نورون برای بازنمایی شکل را می دهد.

هر دو مدل وقتی با تکه های کوچکی از تصاویر طبیعی سازگار می شوند ویژگیهای راستاداری را می سازند که میدان گیرندگی سلولهای ساده را تداعی می کنند (شکل ۳-۱۵). این نتایج مستقیماً از اصل کدگذاری بهینه مشتق می شوند و فرضیه جدید را درباره عملکرد نورونهای V1 ارائه می دهند: بجای عمل کردن به عنوان لبه یاب و یا تحلیل تصویر از روی

¹⁶⁶ Olshausen and Field

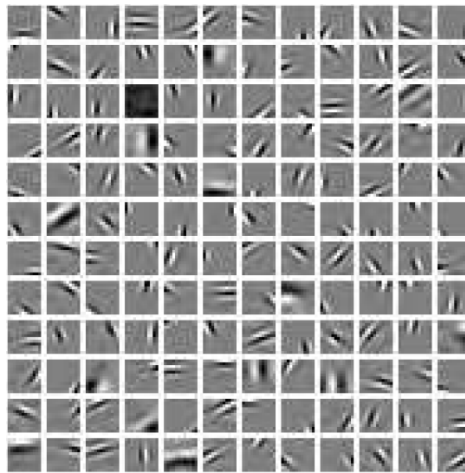
¹⁶⁷ spatially opponent

¹⁶⁸ overcomplete

طیف فرکانسی، نورونهای V1 ورودیهای شبکه را برای کاهش افزونگی پردازش می کنند و کد حاصل اختصاصاً با آمارگان صحنه های طبیعی تنظیم شده است. مهمتر از این مسئله اینست که این مدل نه تنها درباره ویژگیهای تک تک نورونها پیش بینی هایی را صورت می دهد، بلکه درباره نحوه ای که جمعیت نورونها اجماعاً محرک را بازنمایی می کنند نیز پیش بینی هایی دارد (Olshausen and Field 1997; Van Hateren and Van der Schaaf 1998; Lewicki 2002). از آنجایی که این روشها پایه های استواری در چارچوب آماری عمومی دارد و فرضهای اندکی درباره داده ها دارد، در کاربردهای پردازش تصویر متعددی نیز سودمند واقع شده است. این مدلهای داده های تصویری برای فشرده سازی، کاهش ابعاد، و بازیابی آماری تصاویر مفید است و این مسائل با استفاده از مدلهای بیان شده و مدلهای مرتبط قابل حل هستند (Hyvärinen and Hoyer 2000; Hyvärinen and Oja 2000; Lewicki 2002).



شکل ۳-۱۵: روش تحلیل مؤلفه های مستقل وقتی روی تکه تصویرهای طبیعی اعمال شود میدان گیرندگی شبیه سلولهای ساده بدست می دهد. (چپ) زیرمجموعه ای از ۱۰ بردار پایه ICA (وسط) فیلترهای مختص ICA که با الگوی محرک استفاده شده برای محاسبه شدن S متناظر است ($W=A^{-1}$). آنها قابل قیاس با میدانهای گیرندگی نورونها هستند. (راست) میدان گیرندگی بازسازی شده (محرکی بهینه) سلولهای ساده V1 که از آزمایشان فیزیولوژیکی بدست آمده است (رنگ سبز نماد روشنایی و رنگ قرمز نماد تاریکی است)



شکل ۳-۱۶: توابع پایه بدست آمده از بهینه سازی توابع حاشیه ای معیار کورتوسیس. شباهت این توابع با گیرندگی نورون های V1 انکار ناپذیر است.

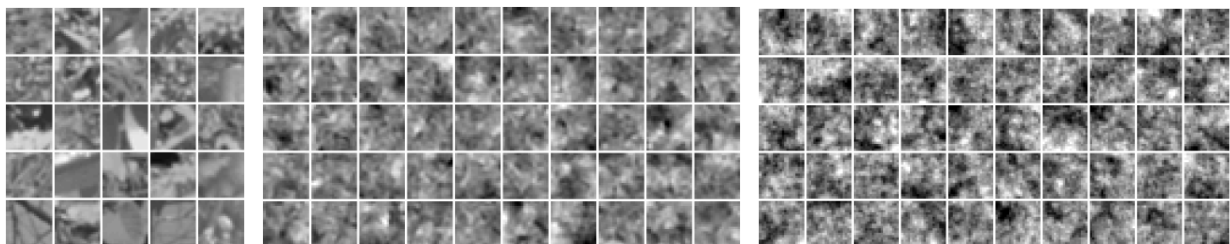
اهداف محاسباتی و قیود نورونی دیگری که برخی از آنها به کاهش افزونگی مرتبطند نیز به تصاویر طبیعی اعمال شده اند. مدلهای خطی که هزینه انرژی^{۱۶۹} را کمینه می کنند (Baddeley 1996)، فعالیت نورونها را ملزم می کنند که در

¹⁶⁹ energy expenditure

طول زمان به نرمی تغییر کنند (Hurri and Hyvärinen 2003)، یا شبه تعامد^{۱۷۰} یک پایه فراکامل را بیشینه می کنند (Inki and Hyvärinen 2001) نیز به مؤلفه های راستادار و محلی مشابهی منجر می شوند. مجموعه های پایه بدست آمده حاوی فیلترهایی در اندازه هایی گوناگون با پهنای باند فرکانسی تقریباً برابر یک اکتاو هستند (Simoncelli 2005) و یک دسته از آنها در شکل ۳-۱۶ آورده شده است. در نتیجه چندین توجیه عملکردی برای شکل میدان گیرندگی نوروها وجود دارد و پیش بینی های نظری زیادی صورت گرفته است که باید توسط آزمایشات فیزیولوژیکی صحت و سقم آنها مشخص شود.

مدل های ذکر شده مانند کد کردن پراکنده و تحلیل مؤلفه های مستقل با این هدف گسترش یافتند که نمایشی از تصاویر طبیعی به دست بیاورند که دارای کمترین افزونگی باشد. این مدل ها با در نظر گرفتن معیارهای آماری در یک نگاهت خطی از تصاویر، سعی در کم کردن میزان وابستگی بین ضرایب حاصل از این نگاهت خطی داشتند. با وجود اینکه این مدلها توانسته اند با موفقیت یک توجیه عملکردی برای پردازش تصویری اولیه در سیستم بینایی ارائه دهند، دچار محدودیتهایی نیز هستند.

نخست آنکه به دلیل خطی بودن مدلهای مولدزیربنایی آنها، تنها آن دسته از وابستگیهای آماری را در بر می گیرند که از ترکیب خطی کمیتهای مستقل حاصل شده اند. اگر این مسئله برای تمام ساختارها برقرار بود انتظار می رفت که اولاً تصاویری که از این مدلها نمونه گیری می شوند از نمونه های تصاویر طبیعی غیر قابل تمایز باشند و ثانیاً ضرایبی که برای کدگذاری تصاویر استفاده می شوند (و مستقل فرض شده اند) واقعاً مستقل باشند. ولی به سادگی می توان نشان داد که قسمت اول صادق نیست. هرچند که تصاویری که از مدلها نمونه گیری می شود از تصاویر تولید شده توسط مدل گوسی به ظاهر تصاویر واقعی طبیعی نزدیکتر است (شکل ۳-۱۷) ولی باز هم بسیاری از ساختارهای مشهود در تصاویر طبیعی را ندارند. مخصوصاً اینکه تصاویری که به طور تصادفی تولید می شوند فاقد ساختارهای تصویری مانند کانتورهای طولانی، بافتهای عادی، و نواحی همگن هستند (Karklin and Lewicki 2008).

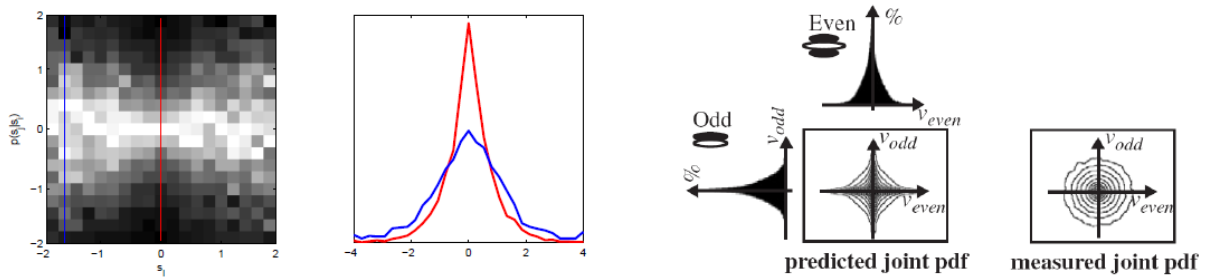


شکل ۳-۱۷: تصاویری که از یک مدل گوسی (راست) و یک مدل ICA با ضرایب پراکنده (وسط) تولید شده اند. هیچکدام از آنها مشابه تکه های تصاویر طبیعی (چپ) نیستند. (ر.ک. به متن)

دوم آنکه نشان داده شده است که ضرایب مدلهای خطی مستقل نیستند. یک مشاهده اینست که بزرگی ضرایب همبسته هستند (شکل ۳-۱۸ چپ) و ارتباطی غیر خطی دارند که این مدلها نمی توانند اخذ کنند. این الگوی وابستگی اولین بار بین ضرایب موجک دیده شد (Simoncelli 1997). قدرت این اثر به ارتباط بین دو فیلتر بستگی دارد؛ زوجهایی که در

¹⁷⁰ quasi-orthogonality

موقعیتهای مجاور درون تصاویر هستند یا راستاهای نزدیکی دارند قویترین همبستگی را از خود نشان می دهند. شاهد دیگری برای وابستگی مراتب بالا اینست که ضرایب به خوبی توسط توزیعهای مجزا توصیف نمی شوند یعنی توزیع توأم آنها را نمی توان از ضرب توزیعهای حاشیه ای بدست آورد (Zetzsche and Rhrbein 2001). این نکته مبین آنست که حتی با داشتن یک دیدگاه صرفاً آماری، کدگذاری خطی برای اخذ پیچیدگی توزیع تصاویر طبیعی کافی نیست. در پژوهش دیگری نیز به این نکته اشاره شده است که ضرایب زیرباند تصویر نیز مستقل نیستند که نوعی دیگر از وابستگی را معرفی می کند (Lyu and Simoncelli 2009).



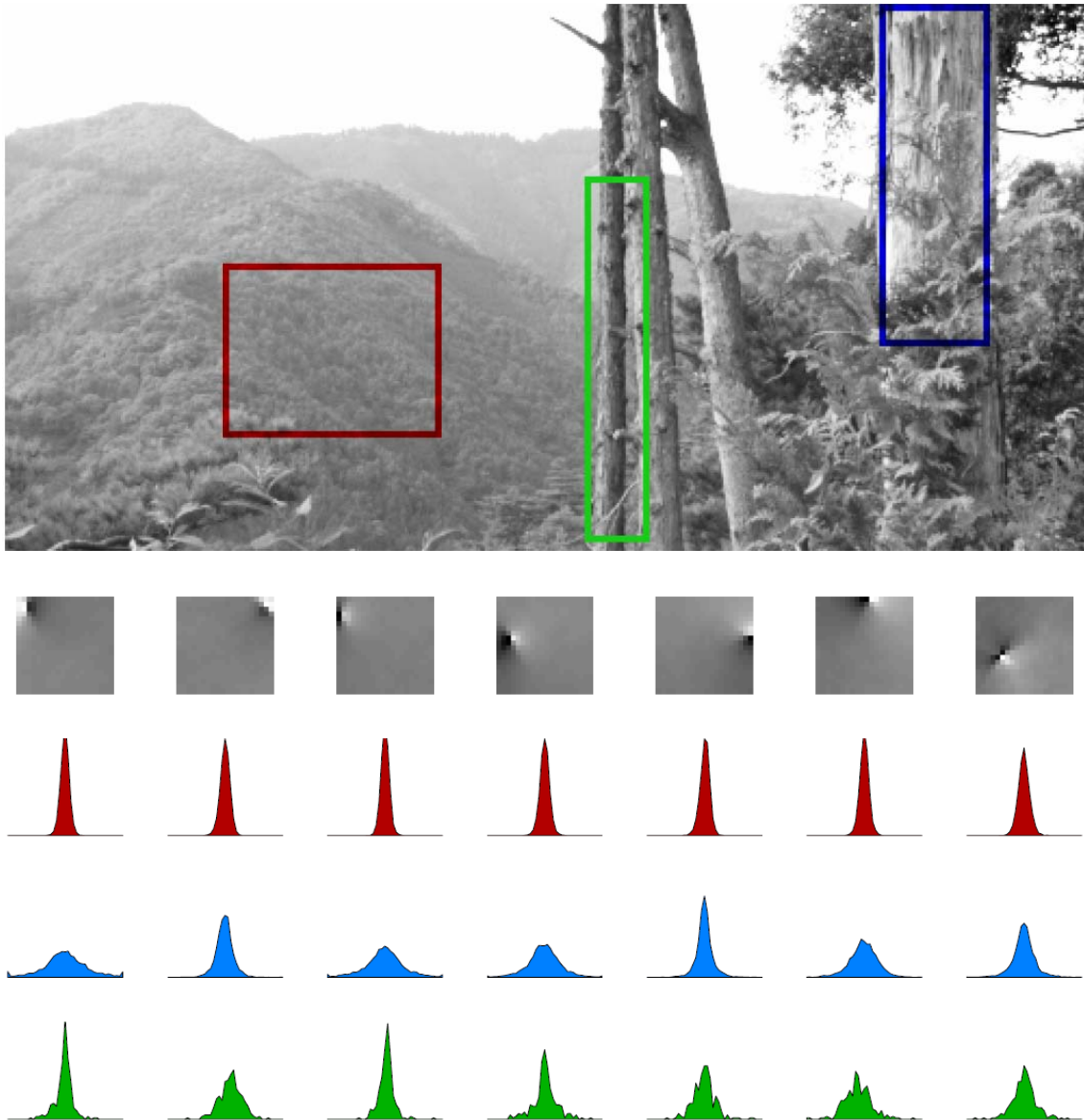
شکل ۳-۱۸: وابستگی بزرگی بین زوجهای ضرایب (چپ) هیستوگرام شرطی، هر برش عمودی توزیع s_j که روی مقادیر s_i شرطی است را نشان می دهد (که به طور قائم قرار دارد و شدت روشنایی آن برای هر ستون به طور مستقلی نرمال شده است) شکل پایونی این توزیع نشان می دهد که واریانس s_j با مقادیر بزرگتر $|s_i|$ افزایش می یابد. (وسط) توزیع احتمال s_j وقتی که $|s_i| < 0.2$ است (خط قرمز) و وقتی $|s_i| > 1.6$ است (خط آبی). (راست) توزیع توأم ضرایب با توزیع جداجدای توزیع احتمال برابر نیست (Zetzsche and Rhrbein 2001).

سومین کمبود این مدلها اینست که آنها فرض می کنند که آمارگان داده ها فرق نمی کنند و به عبارت دیگر توزیعهای احتمالاتی آنها ایستا^{۱۷۱} است. برای مثال به محض اینکه پارامترهای مدلر ICA تطبیق یافتند، توابع احتمال اولیه و پایه ثابت می شوند و به یک توزیع ایستا روی تمام مجموعه تصاویر منجر می شوند. این به شکل تابع احتمال اولیه بستگی ندارد و در مورد مدلهایی که تابع احتمال اولیه آنها سازگار پذیر و یا کاملاً غیرپارامتری است نیز ثابت است. با اینحال آمارگان تصاویر بر اساس محتوای آنها (شکل ۳-۱۹)، دگرگونی خصوصیات فیزیکی محیط یا شرایط اخذ تصویر تغییر می کند. همچنین گزارشاتی درباره تغییرات در آمارگان تصویر به عنوان تابعی از پهنای همسایگی مکانی مجموعه ضرایب موجد موجود است (Ruderman and Bialek 1994; Portilla, Strela et al. 2003; Schwartz, Pillow et al. 2006). در حالی که فرض تابع احتمال اولیه ایستا یک تخمین معتبر از توزیع واقعی داده ها روی بدنه عظیم تصاویر آموزشی بدست می دهد ولی تغییراتی که در زیرمجموعه های مجموعه داده ها دیده می شود را منعکس نمی کند.

همانطور که در شکل ۳-۱۷ نیز مشاهده شد، تبدیل خطی تکه های تصویر بازنمایی قابل قبولی را از ساختارهای بینایی که در تصاویر طبیعی دیده می شود را نشان نمی دهد. بافتها، لبه ها، گوشه ها، و سایر ویژگیهای ظاهراً مهم در بازنمایی متمایز نشده اند. قسمتهای ثابت در تصاویری از یک نوع خاص (مثلاً بافتهای همگن) باعث نوسان شدید خروجی فیلترهای خطی می شوند. به منظور شکل گیری یک بازنمایی پایدار (نامتغیر) روی تمام قسمتهای ثابت، یک مدل باید ویژگیهای مجردتری را از تصویر استخراج کند. متغیرهای نهفته درون مدل باید مشخصات مهم ولی مشاهده نشده ای را

171 stationary

بازنمایی کند که برای گروه بندی نقطه های داده یا یافتن داده هایی که از نظر آماری شبیهند کمک کند. مدل های فعلی چنین توصیف متراکمی از ساختارهای مرتبه های بالا بدست نمی دهد و نیاز مبرمی به مدل های قویتری که بتواند وابستگیهای مشاهده شده را در بر بگیرد، قاعده مندیهایی آماری مراتب بالاتر را توصیف کند، و بازنمایی سودمندی از ساختارهای پیچیده تصویری ارائه کند، احساس می شود (Karklin and Lewicki 2008).



شکل ۳-۱۹: واریانسهای ضرایب ICA از یک محتوا تا محتوای دیگر تصاویر تغییر می کنند. (بالا) تکه های تصویر نمونه برداری شده از سه ناحیه از تصویری طبیعی (برگرفته از (Doi and Lewicki 2005)). (پایین) هیستوگرام ضرایب هفت تابع پایه.

چهارمین کمبود این مدلها هم اینست که با وجود اینکه این مدل ها در شبیه سازی نورون های ساده قشر ابتدایی بینایی موفق عمل کردند ولی امکان توسعه به نورون های سطوح بالاتر قشر بینایی را نداشتند. نورون های پیچیده در قشر ابتدایی بینایی نسبت به نورون های ساده دارای ناحیه دریافت بزرگتری هستند و نسبت به تغییر مکان محرک مورد نظر در ناحیه دریافت خود حساس نیستند. به این ترتیب، نمی توان با استفاده از یک ترکیب خطی از تصویر ورودی به خصوصاتی

مانند گزینندگی نورون های پیچیده دست یافت (Malmir and Shiry 2009). از سوی دیگر پاسخ پس از لایه V1 پراکنده تر نمی گردد (Lyu and Simoncelli 2009).

در ضمن باید توجه داشت که صرف بدست آوردن مؤلفه های راستادار توسط ICA به این معنی نیست که این فیلترها بهترین پاسخها هستند. در بررسی فیلترهای میان گذر مشخص شده است که تمام آنها دارای تابع توزیع حاشیه ای پراکنده هستند (Lyu 2009).

تلاش برای گسترش تحلیل مؤلفه های مستقل شبیه سازی گزینندگی نورون های پیچیده آغاز شد. مدل های جدایی پذیر خطی شالوده خوبی برای مدل های آماری قویتر از تصاویر طبیعی فراهم کرده بودند و چندین توسعه از آنها برای اخذ وابستگی های مشاهده شده بین ضرایب خطی پیشنهاد شد. برای مثال در مدل ICA زیرفضا^{۱۷۲} (Hyvärinen and Hoyer, 2000)، ضرایب خطی دیگر مستقل فرض نمی شوند و بجای آن گروههایی از توابع پایه باهم همسایگی ها (زیرفضاها) بی را می سازند که در هر کدام انرژی ضرایب با هم همبسته است و به همین دلیل فعالیت در آن یا بسیار زیاد و یا نزدیک به صفر است. ایده اصلی در توسعه این مدل، وابستگی بین ضرایب مؤلفه های مستقل تصاویر طبیعی بود. پاسخ مؤلفه های مستقل به دسته هایی تقسیم می شد که پاسخ های درون هر دسته به یکدیگر وابسته و پاسخ های دسته های متفاوت از یکدیگر مستقل بودند. هر کدام از این دسته ها زیرفضای مستقل ویژگیها^{۱۷۳} نامیده شد. برای شبیه سازی خصوصیات نورون های پیچیده مربع ضرایب مؤلفه های مستقل با یکدیگر ترکیب شدند تا پاسخ نورون های پیچیده را شبیه سازی کنند. برای شبیه سازی رفتار غیرخطی نورون های پیچیده، این مدل از مربع پاسخ فیلترهای خطی استفاده می کند و یک ترکیب خطی از این پاسخ ها را به عنوان پاسخ نورون پیچیده مدل می کند. سپس توابع پایه خطی برای بیشینه کردن استقلال نرم برداری هر همسایگی (بجای استقلال تک تک ضرایب) تطبیق می یابند. با توجه به این فرض که متغیرهای درون یک زیرفضا دارای توان همبسته ای هستند، این ساختار وابستگی میان بزرگی ضرایب را می گیرد و به مدل اجازه می دهد که زیرفضاهای بهینه را برای توصیف وابستگی بیابد.

در شکل کلی تری از مدل که به ICA توپوگرافیک^{۱۷۴} معروف است، مجموعه های افزای توابع پایه وابسته با یک چینش توپوگرافیک که وابستگی های بزرگی ضرایب را بطور محلی توصیف می کند جایگزین می شود (Hyvärinen et al., 2001). این مدل به چند یافته جالب منجر می شود: توابع پایه خطی باهم ویژگیهای راستادار و محلی است که اینبار در مجموعه هایی مرتبط یا نقشه های توپولوژیکی مرتب شده اند. واحدهای همسایگی سرکشی شده ویژگیهای پیچیده تری از تکه های تصویری بازنمایی می کند و برخی ویژگیهای نامتغیر سلولهای پیچیده را نشان می دهد (Hyvärinen and Hoyer, 2000, 2001; Hyvärinen et al., 2001).

محدودیت اصلی این مدلها اینست که الگوی وابستگیها از قبل مشخص شده است و از روی داده ها یاد گرفته نمی شود. در ضمن مدل به روابط گسسته (و فقط مثبت) محدود است تا گروههای زیرمجموعه ای یا رابطه های توپوگرافیک بتوانند آنها پشتیبانی کنند. اگر فعالیت درون یک زیرفضا به عنوان مدل یک سلول پیچیده در نظر گرفته شود، آنگاه هیچ کدام

¹⁷² subspace ICA model

¹⁷³ independent feature subspace

¹⁷⁴ topographic ICA

از اثرات پیچیده فرونشانی توسط این مدل گرفته نمی شوند. همچنین بازنمایی ساختارهای تصویری مرتبه بالا به سرکشی به انرژی محلی محدود می شود بنابراین این مدلها از فقدان بازنمایی قاعده مندیهی آماری عمومی تصویرهای پیچیده مانند بافتها رنج می برند (Karklin 2007).

ایرادات وارد شده به مدلهای مبتنی بر توابع پایه خطی دامنگیر کدگذاری پراکنده نیز می شود. البته درباره کدگذاری پراکنده باید گفت که هنگامی که بازنمایی فراکامل باشد یا سطح نویز چشمگیر باشد، مرحله کدگذاری (بدست آوردن بازنمایی بهینه برای تصویر داده شده) غیرخطی است و ضرایب تابعهای پایه برای بازنمایی هر چه بهتر تصویر و بیشینه کردن پراکندگی کد رقابت می کنند. این رفتار رقابتی به رفتارهایی غیرخطی در پاسخ مدل می شود که با برخی از ویژگیهای سلولهای ساده مانند توقف انتها و فرونشانی اطرافیان در (V1 (Raina et al., 2007) مطابقت دارد. این امر نشانگر آنست که رقابت عمومی و بازنمایی بسیار پراکنده ممکن است مسبب برخی از خصوصیات غیرخطی شوند. با اینحال این مدل تمامی رفتارها یا پاسخهای غیرخطی را در نواحی تصویری بالاتر را نمی تواند شرح دهد (Karklin 2007).

مدل حاشیه ای موجک: به موازات این پیشرفتهای آماری، پژوهشگران زمینه های مختلفی در حال تهیه پایه های متعامد یکه چند مقیاسه برای سیگنال و تصویر بودند که امروزه به نام موجک¹⁷⁵ مشهورند. این پایه ها تخمینهای خوبی از پایه های بهینه ای مانند آنچه در شکل ۳-۱۶ نشان داده شده است بدست می دهند.

حال اگر تصویر را به یک بازنمایی موجک چند مقیاسه تبدیل کنیم، کدام مدل آماری برای مشخص کردن ضرایب مناسب خواهد بود؟ انگیزه آماری این انتخاب برای توابع پایه از شکل توزیعهای حاشیه نشأت می گیرد بنابراین این فرض که ضرایب درون یک زیرباند را مستقل و با توزیع یکنواخت¹⁷⁶ در نظر بگیریم توجیه پذیر است. با این مفروضات، مدل کاملاً با آمارگان حاشیه ای ضرایب - که مانند آنچه در مثال شکل ۳-۱۲ آمده به طور تجربی قابل تعیین است - معین می شود. برای تصاویر طبیعی، این هیستوگرامها به طور اعجاب آوری با توزیع دو پارامتری گوسی تعمیم یافته¹⁷⁷ (که به نامهای تابع نمایی کشیده شده یا تعمیم یافته¹⁷⁸ نیز معروف است) توصیف می شوند (Mallat 89, 47, Moulin 99, simoncelli 99).

$$p(x) = \frac{1}{Z(s, p)} \exp\left(-\left|\frac{x}{s}\right|^p\right) \quad (11-3)$$

که در آن ضریب نرمال سازی برابر $Z(s, p) = 2 \frac{s}{p} \Gamma\left(\frac{1}{p}\right)$ است. توان $p = 2$ با توزیع گوسی متناظر است، و توان $p = 1$ به توزیع لاپلاسین مربوط است. به صورت کلی مقادیر کوچکتر p به توزیعی منتهی می شوند که هم در صف متراکمتر است و هم دنباله های گسترده تری دارد (شکل ۲-۱۰). هر کدام از هیستوگرامهای شکل ۳-۱۲ که با نقطه چین رسم شده اند با سازگارترین نمونه این توزیعها متناظر است که پارامترهای این توزیع $\{s, p\}$ ، از بیشینه کردن درستنمایی داده ها تحت مدل بدست می آید. این مدل توزیع به طور قابل توجهی خوب روی هیستوگرامهای منطبق می شود که این مدعا توسط معیارهای

¹⁷⁵ wavelet

¹⁷⁶ independent and identically distributed (iid)

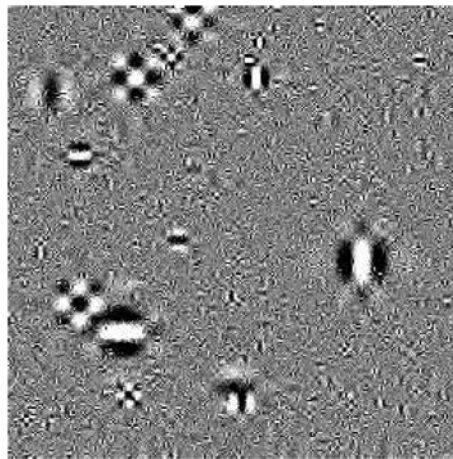
¹⁷⁷ generalized Gaussian

¹⁷⁸ stretched or generalized exponential

آنتروپی مربوطه که پایین هر نمودار آورده شده است اثبات شده است. مشاهده شده است که مقدار توان p معمولاً در بازه $(0/8, 0/4)$ است و عامل S نیز به طور یکنوا با مقیاس توابع پایه تغییر می کند بدین صورت که واریانس نسبتاً بالاتر مربوط به مؤلفه هایی با مقیاس درشت تر است.

این مدل حاشیه ای موجک به طور ملموسی از مدل کلاسیک گوسی (طیفی) بهتر است. به عنوان مثال در کاربرد فشرده سازی، آنتروپی توزیعی که بالاتر به آن اشاره شد از یک توزیع گوسی با واریانس مشابه بسیار کمتر است و مستقیماً به بالاتر رفتن کارایی کدگذاری می انجامد. در رفع نویز نیز استفاده از این مدل به عنوان توزیع اولیه تصویر منجر به بهبود فاحشی نسبت به مدل گوسی می شود (Simoncelli 96, Chipman 97, Abramovich 98, Moulin 99). تخمین زنهایی که از این روش بدست می آیند تابعهای هسته ای^{۱۷۹} غیرخطی ای هستند که در آن ضرایبی با بزرگی زیاد را حفظ می کنند. این تخمین زنها بهبود قابل ملاحظه ای را نسبت به تخمینهای خطی مدل گوسی نشان می دهند.

علیرغم تمام این موفقیتها، باز هم به آسانی می توان نشان داد که خاصیتهای مهمی از تصویر توسط مدل حاشیه ای موجک ثبت نمی شوند. هنگامی که تبدیل موجک متعامد یگه باشد، به آسانی می توان نمونه های آماری از مدل گرفت. شکل ۳-۲۰ نتیجه نمونه گیری از ضرایب بازنمایی موجک را مستقل از توزیهای گوسی نرمال شده نشان می دهد. پارامترهای توزیع برای هر زیرباند آنهایی انتخاب شده اند که با یک تصویر خاص (در اینجا تصویر انیشتین) بهترین سازگاری را داشته باشد. با وجود اینکه نتیجه حاصل از نویز سفید و حتی از هر تصویری که از مدل طیفی استخراج شود (شکل ۳-۹) دارای ساختارهای بیشتری است ولی بازهم چندان شبیه یک عکس نیست.



شکل ۳-۲۰: یک تصویر نمونه که از مدل حاشیه ای موجک استخراج شده است، پارامترهای توزیع زیرباند طوری انتخاب شده است که با تصویر انیشتین (شکل ۳-۲۱) تطبیق داشته باشد.

مدل حاشیه ای موجک می تواند با توسعه خود روی یک پایه فراکامل موجک بهبود یابد. به زبان دقیقتر، ژو^{۱۸۰} و همکاران نشان داده اند (Zhu 98) که با استفاده از نسخه ای از نظریه برش نمای فوریه^{۱۸۱} که برای بازسازی تصاویر

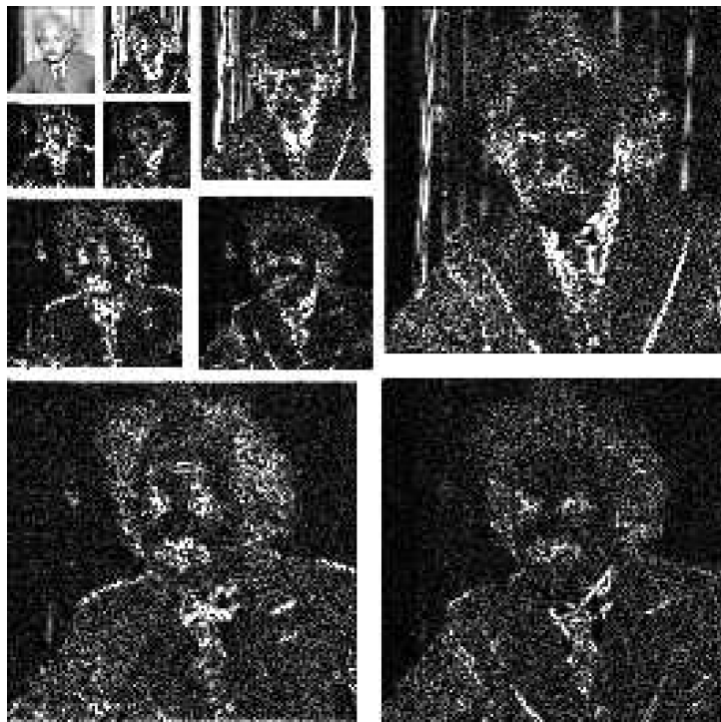
¹⁷⁹ kernel functions

¹⁸⁰ Zhou et al

توموگرافی استفاده می شود، این تعداد زیاد توابع حاشیه ای برای محدود کردن یکتای تابع توزیع با ابعاد زیاد کافی است. این ایده برای ساخت مدل‌های کارایی برای بازنمایی بافت و سنتز آن استفاده شده است (Heegar 95, Zhu 97, Portilla 2000).

ضعف این مدل اینست که خصوصیات آماری توأم به صورت ضمنی درون مجموع آمارگان حاشیه ای تعریف شده است لذا مطالعه مستقیم آنها یا بکارگیری آنها برای تولید راه‌حلهایی برای کاربردهای پردازش تصویر دشوار است. برای راحت کردن مسئله، توصیف‌های آماری توأم بکار گرفته شد.

مدلهای توأم موجک: اصلی‌ترین دلیل نمایش ضعیف تصویر در شکل ۳-۲۰ اینست که ضرایب تبدیل موجک مستقل نیستند. به تجربه ثابت شده است که ضرایب حاصل از تجزیه موجک متعامد یگه^{۱۸۲} تصاویر تقریباً ناهمبسته هستند (یعنی کواریانس آنها صفر است). ولی این جمله تنها در مورد وابستگی‌های مرتبه دوم آنها صحیح است و بسادگی می‌توان دید که وابستگی‌های آماری مهمی از مراتب بالاتر نیز وجود دارد. شکل ۳-۲۱ بزرگی (مقدار مطلق) ضرایب را در تجزیه موجک متعامد یگه جدایی‌پذیر چهار مرحله ای نشان می‌دهد.



شکل ۳-۲۱: بزرگی ضرایب چندمقیاسه موجک برای تصویر انیشتین. هر تصویر کوچک بزرگی ضرایب زیرباند را که از کانال‌ها با فیلتری در راستا و مقیاسی متفاوت بدست می‌آید و با فاکتوری مناسب نمونه‌گیری شده است نشان می‌دهد. ضرایبی که از لحاظ مکانی در یک باند به هم نزدیک هستند تمایل بیشتری به داشتن بزرگی‌های مشابه دارند و به علاوه ضرایبی که در راستاها یا مقیاس‌های متفاوت ولی در موقعیت نسبی مشابهی قرار دارند نیز به داشتن بزرگی‌های مشابه متمایلند.

¹⁸¹ fourier projection-slice theorem

¹⁸² orthonormal wavelet decomposition

باید توجه داشت که ضرایب دارای مقدار مطلق زیادتر درون یک زیرباند بیشتر نزدیک هم رخ می دهند و نیز می توان دید که این ضرایب دارای مقدار مطلق زیاد در موقعیت مکانی مشابهی در زیر باند مقیاسها و راستاهای مجاور رخ می دهند (Simoncelli 97, Buccigrossi 99).

دلیل منطقی برای این خوشه بندی ضرایب دارای مقدار مطلق زیاد اینست که ویژگیهای محلی و ایزوله متداول در تصاویر در حوزه موجک به صورت قرارگیری یک گروه از توابع پایه در موقعیتهای، راستاها و مقیاسهای مختلف بازنمایی می شوند. علامت و اندازه های نسبی این ضرایب با این بردارهای پایه به مکان، راستا، و مقیاس دقیق این ویژگیها بستگی دارد. اندازه این ضرایب همچنین با کنتراست این ساختار کم و زیاد می شود. بنابراین اندازه گرفتن یک ضریب بزرگ در یک مقیاس بدین معناست که ضرایب بزرگی نیز در مقیاسهای مجاور محتملترند. این خاصیت خوشه بندی در یک روش هیوربستیکی بسیار کارا درون کدکننده موجک درخت صفر پنهان^{۱۸۳} (EZW)(Shapiro 93) به کار گرفته شده است و از آن پس به نوعی در تقریباً تمام سیستمهای فشرده سازی تصویری استفاده شده است. یک تعریف صریحتر برای این روش برای بار نخست در وادی رفع نویز مطرح شده است.

مدل وابستگیهای واریانس: بیش از ربع قرن پیش، لی^{۱۸۴} (Lee 80) یک روال دو مرحله ای برای رفع نویز تصویر پیشنهاد داده بود که در آن ابتدا واریانس سیگنال محلی از روی همسایگی پیکسل دیده شده تخمین زده می شد و سپس پیکسلهای آن همسایگی با روش استاندارد کمترین مربعات خطی رفع نویز می شد. هر چند این روش در حوزه پیکسل کار می کرد، لیکن این مقاله این ایده را مطرح کرد که واریانس یک خصوصیت محلی است و باید به طور تطبیقی تخمین زده شود که در تقابل با دیدگاه مدل کلاسیک گوسی بود که واریانس را یک ثابت عمومی برای سراسر تصویر قلمداد می کرد.

رودرمن^{۱۸۵} (Ruderman 96) خصوصیات واریانس محلی مشتقات تصویر را آزمود و متوجه شد که حوزه مشتق گیری شده می تواند بوسیله نرمال سازی با یک تخمین محلی از انحراف معیار بسیار همگن شود. تا دهه ۱۹۹۰ این ایده در رفع نویز درون دامنه موجک استفاده نشد ولی پس از آن با تخمین واریانس خوشه های ضرایب موجک در مکانها، مقیاسها و/یا راستاهای نزدیک به استفاده از این واریانس تخمین زده شده برای رفع نویز خوشه توجه شد (Malfait 97, Simoncelli 97, Chang 98, Simonvelli 99, Michak 99, Wainwright 2001, Abramovich 2002).

مدل های غیر خطی مختلفی برای برطرف کردن این نوع وابستگی ارائه شده اند. به عنوان مثال شوارتز^{۱۸۶} از مکانیزم نورونی نرمال سازی با تقسیم^{۱۸۷} در رفع افزونگی استفاده کرد (Schwartz01). نرمال سازی پاسخ یک نورون هنگامی اتفاق می افتد که محرک ورودی به نورون، بزرگتر از میدان گیرندگی آن باشد و نورون های مجاور آن را نیز فعال کند. در این صورت، پاسخ نورون های مجاور باعث کاهش فعالیت نورون مرکزی می شوند. یکی از مکانیزم هایی که برای توجیه این رفتار بیان شده است، نرمال سازی با تقسیم است. در این مکانیزم، پاسخ نورون به مجموع پاسخ نورون های اطراف خود تقسیم می شود. مدل نرمال سازی با تقسیم بر اساس مدلی از تصاویر طبیعی است که رفتار نورون ها را به صورت ترکیبی از متغیرهای

¹⁸³ embedded zerotree wavelet

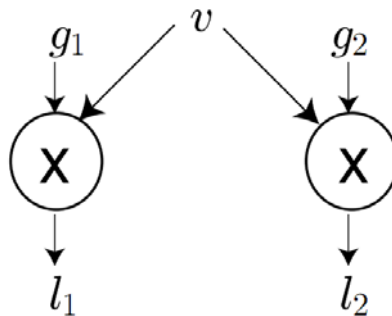
¹⁸⁴ Lee

¹⁸⁵ Ruderman

¹⁸⁶ Schwartz

¹⁸⁷ divisive normalization

نرمال که دارای واریانس مشترکی هستند بیان می‌کند (Wainwright99). به عنوان مثال، پاسخ دو فیلتر که دارای وابستگی واریانس هستند توسط مدل شکل ۳-۲۲ توجیه می‌شود. در این شکل، پاسخ دو فیلتر وابسته با l_1 و l_2 مشخص شده اند. دو متغیر مستقل گوسی به نام های g_1 و g_2 در یک متغیر تصادفی عددی به نام v ضرب می‌شوند تا پاسخ نورون های وابسته را ایجاد کنند. در این مدل، متغیر v مسئول ایجاد وابستگی بین متغیرهای l_1 و l_2 می‌باشد. در تحلیل های آماری نشان داده می‌شود که این متغیر تنها باعث وابستگی واریانس در نورون های مذکور می‌شود و بر میانگین آنها تاثیری ندارد. در مقاله (Wainwright99) نشان داده شد که اگر پاسخ نورون به واریانس آن تقسیم شود، وابستگی بین پاسخ ها از بین می‌رود. شوارتز از این واقعیت استفاده کرد و پاسخ نورون را به مربع پاسخ های نورون های اطراف تقسیم کرد تا پاسخ مستقل حاصل شود.



شکل ۳-۲۲: مدل ترکیب متغیرهای گوسی برای توجیه فتر وابستگی واریانس. در این شکل دو متغیر گوسی مستقل g_1 و g_2 در متغیر عددی v ضرب شده اند تا پاسخ وابسته نورون های l_1 و l_2 را ایجاد کنند.

در مدل شوارتز، هر نورون به وسیله‌ی اتصالات افقی با نورون های مجاور خود ارتباط دارد. این مدل نورون های $V1$ را با استفاده از تحلیل مؤلفه‌های مستقل به دست می‌آورد و سپس با استفاده از اتصالات افقی بین نورون ها، سعی در برطرف کردن وابستگی بین آنها دارد. واریانس هر نورون به صورت حاصلجمع مربعات پاسخ نورون های اطراف نشان داده می‌شود:

$$\text{var}(L_x | L_y, y \in C_x) = \sum_{y \in C_x} w_{xy}^l L_y^2 + \sigma_x^2 \quad (12-3)$$

در معادله بالا، L_x پاسخ نورون x قبل از نرمال سازی را نشان می‌دهد و C_x یک همسایگی اطراف نورون x است که نورون های آن دارای وابستگی واریانس با x هستند. وزن اتصال افقی بین نورون های x و y با w_{xy}^l نشان داده می‌شود و σ_x^2 بخشی از واریانس x است که به پاسخ نورون های اطراف بستگی ندارد. به ازای هر ورودی، مربع پاسخ نورون x به واریانس آن تقسیم می‌شود تا وابستگی از بین برود:

$$R_x = \frac{L_x^2}{\left(\sum_{y \in C_x} w_{xy}^l L_y^2\right) + \sigma_x^2} \quad (13-3)$$

در معادله ۳-۱۳، R_x پاسخ مستقل برای نورون x می باشد. پارامترهای مدل با استفاده از تخمین درستنمایی بیشینه^{۱۸۸} روی یک مجموعه از تصاویر طبیعی به دست می آیند (برای سادگی فرض شده که پاسخ نورون ها دارای توزیع نرمال می باشند):

$$\hat{w}_{xy}^l, \hat{\sigma}_x = \arg \max_{w_{xy}^l, \sigma_x} \prod_i p(L_x(i) | L_y(i), y \in C_x) \quad (۱۴-۳)$$

که در معادله بالا p تابع توزیع احتمال توام برای پاسخ نورون ها را نشان می دهد. این مدل توانست رفتارهای غیر خطی نورون های V1 مانند پاسخ مستقل از شدت روشنایی به لبه ها در یک جهت خاص را مدل کند. مدل‌های دیگری نیز بر اساس این اصل بنا شده اند که از جمله آنها می توان به (Lyu 2007) اشاره کرد که با زمینه تصادفی مارکوفی^{۱۸۹} تصویر را به زمینه هایی تقسیم کرده است و در هر یک به کمک تقسیم واریانس محلی هر زمینه بر ترکیب خطی واریانسهای سایر زمینه ها، نرمالسازی با تقسیم را انجام داده است.

آمار توأم دو فیلتر خطی وقتی با همان تصویر کانالو شوند یک وابستگی آماری واضح را بدست می دهند: وقتی که یک فیلتر به جنبه برجسته ای از تصویر پاسخی قوی می دهد، فیلتر دیگر نیز احتمالاً پاسخی قوی خواهد داد (به عنوان مثال دو فیلتر عمودی دارای تفاوت مکانی از هم به یک لبه عمودی طولانی در تصویر پاسخ می دهند). این پدیده به خاصیت خود سازماندهی^{۱۹۰} تصاویر معروف است (Turiel, Mato, Parga, & Nadal, 1998). قدرت این وابستگی با شباهت ویژگیهای فیلترهای خطی مانند موقعیت نسبی، راستا، مقیاس مکانی، فاز و غیره مشخص می شود. مختصات در توزیع توأم شرطی منعکس می شود و این توزیع شکلی مشابه پایون دارد که خود نشان دهنده وابستگی واریانس است (Buccigrossi & Simoncelli, 2001; Schwartz & Simoncelli, 1999) یا با بررسی توزیع حاشیه ای در برابر توزیع توأم می توان به این نتیجه رسید (Zetsche et al., 1993; Zetsche & Nuding, 2005). هوانگ و مامفورد^{۱۹۱} نمودار کانتور توأم را برای یک پایگاه داده بزرگ تصاویر بررسی کردند و وابستگی توأم را به صورت یک توزیع گوسی تعمیم یافته دو بعدی مدل کردند (Huang and Mumford, 1999). وابستگیها در پاسخ انواع گوناگونی از فیلترهای خطی دیده می شود که از جمله آنها می توان به موجکهای از پیش تعریف شده و فیلترهایی که برای کسب بیشترین پراکندگی یا استقلال طراحی شده اند اشاره کرد. این وابستگیها حتی وقتی پاسخ فیلترها بطور خطی ناهمبسته هستند نیز حضور دارند.

اصل واریانس تطبیقی محلی بسیار قوی است ولی یک مدل احتمالاتی کامل بدست نمی دهد. همانند مدل‌های قبل می توان مدل بسیار صریحتری از طریق بررسی آمارگان ضرایب ساخت (Simoncelli 97). سطر بالای شکل ۳-۲۳ هیستوگرام توأم چند جفت متفاوت از ضرایب موجک را نشان می دهد. همانند توزیعهای حاشیه ای، برای اینکه این هیستوگرامهای توأم که از جفتهای مختلفی از ضرایب که در جاهای مختلف تصویر پخش شده اند را بتوانیم به عنوان نمایندگان توزیع آنها در نظر بگیریم، نیاز به فرض همگنی داریم. ضرایبی که از توابع پایه مجاور باشند کانتورهایی تقریباً

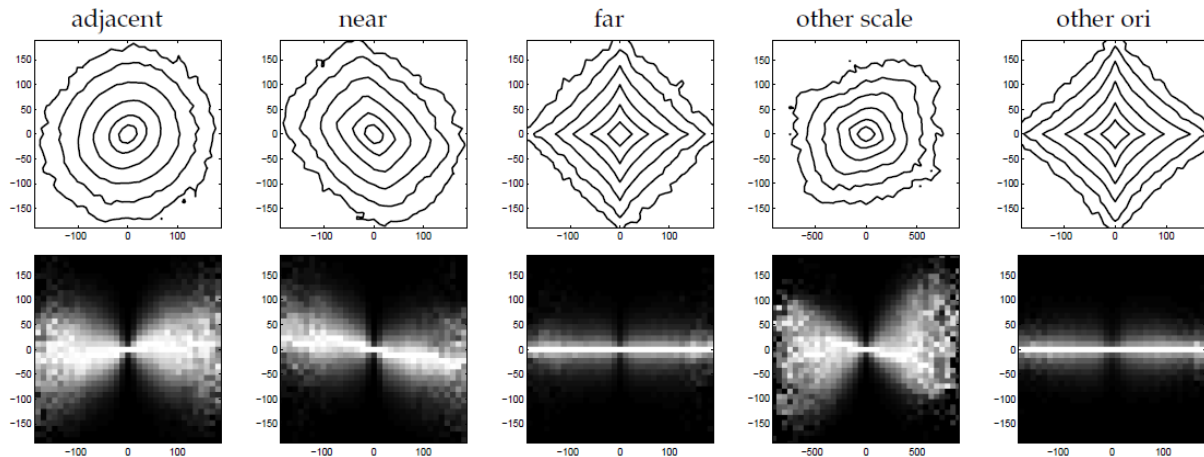
¹⁸⁸ maximum likelihood estimate

¹⁸⁹ Markov random field

¹⁹⁰ self organizing

¹⁹¹ Huang & Mumford

دایروی تولید می کنند در حالیکه مابقی آنها آشکارا در راستای محورها کشیده شده اند. زتسه^{۱۹۲} (zetsche 93) توزیعی تجربی جفتهای رُبعی^{۱۹۳} (تبدیل هیلبرت^{۱۹۴}) توابع پایه را بررسی کرد و دریافت که کانتورهای آنان نیز تقریباً دایروی است. چند پژوهشگر نیز مدل گوسی تعمیم یافته حلقوی^{۱۹۵} را به عنوان مدلی برای آمارگان توأم ضرایب نزدیک بهم موجک پیشنهاد دادند (Huang 99, Sirvastava 2002).



شکل ۳-۲۳: توزیع توأم تجربی ضرایب موجک که با جفتهای متفاوتی از توابع پایه آزموده شده اند (روی شکل واحدی از صحنه های خیابانهای نیویورک که در شکل ۳-۹ معرفی شدند). ردیف بالا توزیع توأم را به صورت نمودار کانتور نشان می دهد که خطهای آن با تناوب یکسانی از احتمال لگاریتمی نمونه برداری شده است. سه نمونه سمت چپ با زوجهایی از توابع پایه متناظرند که مقیاس و راستای یکسانی دارند ولی فاصله مکانی آنها متفاوت است. نمودار بعدی با زوجی از مقیاسهای مجاور متناظر است (با راستایی یکسان و موقعیت مکانی تقریباً مشابه) و نمودار راست با زوجی در راستاهای همود بر هم (ولی مقیاسی یکسان و موقعیت مکانی تقریباً مشابه) متناظر است. ردیف پایین توزیع شرطی متناظر را نشان می دهد. در این نمودارها روشنایی با فرکانس رخداد متناسب است و هر ستون به صورت مستقل نرمالیزه شده است تا از تمام بازه شدت روشنایی استفاده کند.

هیستوگرامهای توأمی که در سطر اول شکل ۳-۲۳ مشاهده می شوند این مسئله را که آیا ضرایب مستقل هستند یا خیر را روشن نمی کنند. برای روشن کردن این مسئله، سطر پایین هیستوگرامهای شرطی همان داده ها را نشان می دهد. x_2 را متناظر با ضرایب توزیع (محور عمودی) و x_1 را ضرایب شرطی (محور افقی) در نظر می گیریم. هیستوگرام حاصل چند وجه مهم از ارتباط بین دو ضریب را نشان می دهد. اولاً، امید ریاضی^{۱۹۶} x_2 روی تمام مقادیر x_1 برابر صفر است که بیانگر ناهمبستگی تقریباً کامل آنها (از دید آمارگان مرتبه دوم) است. دوم اینکه واریانس هیستوگرام شرطی x_2 آشکارا به مقدار x_1 وابسته است و شدت این وابستگی به جفت ضریبی که مد نظر بوده است بستگی دارد. بنابراین با وجود اینکه x_2 و x_1 ناهمبسته اند بازهم وابستگی آماری از خود نشان می دهند!

¹⁹² Zetsche

¹⁹³ quadrature pairs

¹⁹⁴ Hilbert transform

¹⁹⁵ circular generalized Gaussian

¹⁹⁶ expected value

افزونگی در تصاویر طبیعی شامل واریانس پاسخ فیلترهای خطی مانند گابور و موجک می شود و یک نگاشت خطی توانایی آشکار ساختن آن را ندارد. مدل‌هایی که برای رفع این نقیصه توسعه یافته اند فرض می کنند که وابستگی تنها بین پاسخ فیلترهای خطی در یک همسایگی محدود از آن فیلتر وجود دارد. در تمام این مدل ها، فرض شده که پاسخ فیلترهای خطی در یک همسایگی محدود به یکدیگر وابسته هستند. محدوده این همسایگی معمولاً در مدل ها مشخص نشده است و هر مدل اندازه ی دلخواهی برای آن در نظر گرفته است (مالمیر). اندازه همسایگی وابستگی در تمام تصاویر یکسان نیست که علت آن تا حدودی توسط وابستگی زمینه قابل توجه است.

یکی از مطالعات جدیدی که خصوصیات آماری تصاویر را مورد بررسی قرار داده، وابستگی بین واریانس ها را بر اساس زمینه¹⁹⁷ تصویر تشریح می کند (Karklin05). در این تحقیق، پاسخ فیلترهای خطی گابور روی مجموعه ای از تصاویر طبیعی مورد آزمایش قرار گرفتند. برای تصاویر خاصی، پاسخ فیلترها از شرط مستقل بودن انحراف پیدا می کرد و توزیع آن مطابق با توزیع فرض شده نبود. شکل ۳-۱۹ پاسخ نمونه فیلترها را برای یک تصویر خاص نشان می دهد. همانطور که ملاحظه می شود، پاسخ فیلترها برای محدوده خاصی از تصویر که دارای بافت هم جهت با فیلتر می باشد، با توزیع پیش فرض مطابق نیست. در این مقاله، این خصوصیت از پاسخ فیلترها به زمینه تصویر نسبت داده شده است به طوری که زمینه های متفاوت الگوهای متفاوتی از وابستگی را در تصاویر به وجود می آورند. این وابستگی ها بسته به اندازه زمینه بین پاسخ فیلترهای مختلف برقرار است.

فرم هیستوگرامهایی که در شکل ۳-۲۳ دیده می شود به طور اعجاب آوری برای بسیاری از تصاویر پایدار است. به علاوه شکل کیفی این روابط آماری برای جفتهای ضرابی موجود در موقعیتهای مکانی مجاور و راستاهای مجاور صادق است. با نگاهی به ضرابی که فاصله بیشتری دارند (چه از نظر مکانی و چه از نظر مقیاس) این وابستگی کمتر به چشم می خورد و به ذهن القا می کند که شاید فرض مارکوف مناسب باشد. از سوی دیگر ضرایب زیرباند دارای توزیع حاشیه ای غیر گوسی هستند و زوجهای ضرایب، یا خوشه های محلی، تقریباً کروی هستند. زیرباندهای تصویر دارای تعداد کمی ضرایب خیلی بزرگ هستند (که می توان آنها را به راحتی از نویز تفکیک کرد) و این ضرایب معمولاً در نزدیکی هم قرار دارند این کانتورهای حلقوی (یا بیضوی)، یا همان وابستگی بین بزرگی ضرایب همانند رفتار توزیعهای حاشیه ای متناظر می توانند با استفاده یک حوزه تصادفی با واریانس که با توجه به مکان نوسان می کند¹⁹⁸ مدل شوند. مثال مناسبی برای اینکار از ضرب یک بردار گوسی در یک ضریب اسکالر پنهان بدست می آید که به مدل مخلوط مقیاس گوسی¹⁹⁹ (GSM) مشهور است. این توزیعها زیرمجموعه مهمی از توزیعهای متقارن بیضوی را بازنمایی می کند که شامل آنهاست که به عنوان تابعی از نرم ربعی بردار تصادفی محسوب می شوند. این نوع از مدلها که در حوزه تصادفی تعبیه شده اند در زمینه پردازش گفتار بسیار سودمند واقع شده اند (Brehm 87). یک مجموعه مرتبط از مدلها به مدلهای خود رگرسیون واریانس-متغیر شرطی²⁰⁰ (ARCH) مانند (Bollersley 94) معروفند و برای بسیاری از سیگنالهای حقیقی که دچار نوسانات ناگهانی و پس از آن دوره های نسبتاً

¹⁹⁷ context

¹⁹⁸ این خاصیت به heteroscedasticity معروف است.

¹⁹⁹ gaussian scale mixture model

²⁰⁰ autoregressive conditional heteroskedastic model

آرامی هستند (مانند قیمت‌های بازار سهام) مفید واقع شده اند. فیزیکدانانی که خصوصیات آشوب را بررسی می کردند نیز متوجه رفتارهای مشابهی شده اند (98 Turial).

مدلهایی مانند GSM از معماریهای انعطاف پذیری استفاده می کنند ولی به تبدیل خطی ثابتی مانند تجزیه چندمقیاسه موجک اتکا دارند و به کمک آنها وابستگی بین ضرایب خود را مدل می کنند. مدل GSM حاصل نگرش متفاوتی به خاصیت خودسازماندهی تصاویر است (Wainwright et al., 2001; Wainwright & Simoncelli, 2000) که از معرفی آن بیش از سه دهه می گذرد (Andrews and Mallows 1974). مدل مخلوط مقیاس گوسی (Andrews, Buccigrossi and Simoncelli, 2001; and Mallows, 1974; Wainwright et al., 2001) و برخی از مشتقات آن (Buccigrossi and Simoncelli, 2001; Romberg et al., 1999) ضرایب تبدیلیهای خطی (مانند هرم چندمقیاسه موجک) را به صورت حاصلضرب متغیرهای گوسی مستقل و متغیرهای همزن²⁰¹ که متقابلاً وابسته هستند، توصیف می کنند. وابستگی بین متغیرهای همزن در طول یک ساختار از پیش تعیین شده (معمولاً درختی از ضرایب موجک) انتشار می یابد ولی قدرت این وابستگی می تواند از داده ها یاد گرفته شود. این مدلها نشان داده اند که می توانند فشردن سازی تصویر (Buccigrossi and Simoncelli, 1999) و رفع نویز آنرا (Wainwright et al., 2001; Portilla et al., 2003) بهبود بخشند که به طور ضمنی نوید پیدا کردن توصیف بهتری از آمارگان تصویر را می دهد. این مدلها برای توجیه وابستگیهای آماری مشاهده شده در تصاویر طبیعی طراحی شده اند ولی در عین حال می توانند برای شرح دادن خصوصیت‌های غیرخطی نورونهای حسی نیز استفاده شوند و بر این نکته که بازنمایی نورونی برای آمارگان صحنه های طبیعی بهینه سازی شده اند، صحنه بگذارد. برای مثال این مدلها تخمینهای محلی واریانس را برای ورودیهای فیلترهای خطی حساب می کنند. نرمال کردن خروجی فیلترها کد مستقلتری بدست می دهد که بهتر از مدل‌های خطی ساده با پاسخهای نورونی مطابق است، و در ضمن رفتارهای پیچیده سلولهای ساده نظیر اشباع پاسخ کنتراست²⁰²، بازداری فعالیت به ازای راستاهای متقاطع، و فرونشانی ماسک²⁰³ را نیز از خود بروز می دهد (Schwartz and Simoncelli, 2001).

این مدل از دو بخش تشکیل شده است: یک بخش گوسی چندمتغیره \vec{g} که در یک متغیر تصادفی اسکالر مثبت v ضرب شده است. مؤلفه دوم، \vec{g} ، مؤلفه گوسی را بطور کارایی مقیاس می کند و یک مخلوط \vec{l} را طبق رابطه زیر می سازد که توزیع آن نیز در رابطه آمده است:

$$\vec{l} = v\vec{g}$$

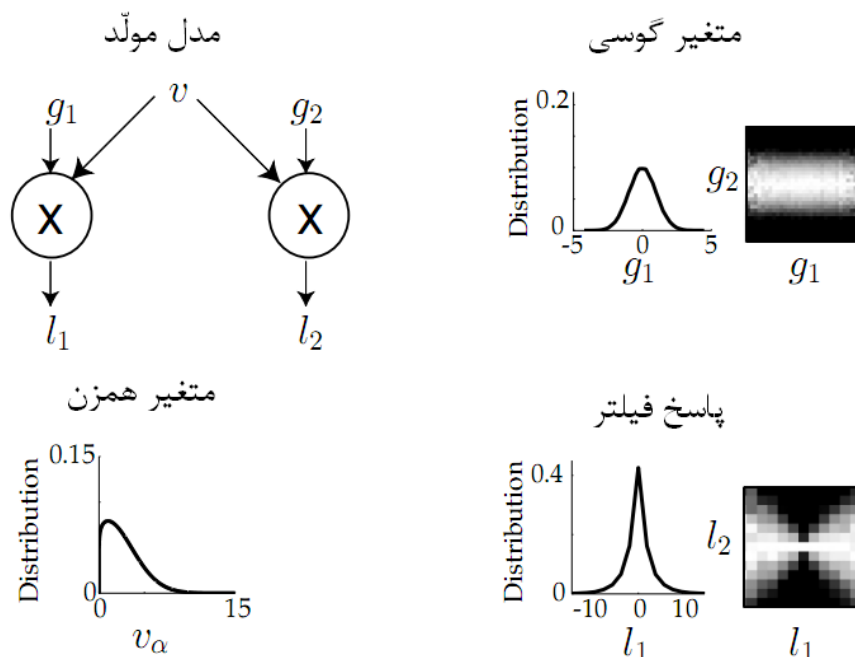
$$p(\vec{l}) = \int \frac{1}{(2\pi)^2 \left|v^2 \Sigma\right|^{\frac{1}{2}}} \exp\left(-\frac{\vec{l}^T \Sigma^{-1} \vec{l}}{2v^2}\right) p(v) dv \quad (15-3)$$

²⁰¹ mixer variable

²⁰² contrast response saturation

²⁰³ mask suppression

که m تعداد فیلترها (ابعاد)، Σ ماتریس کواریانس، و متغیر v با توجه به توزیع $p(v)$ مخلوط شده است و این رابطه در یک همسایگی محاسبه می شود که متغیر v در آن معتبر است^{۲۰۴}. در کاربرد آن برای تصاویر طبیعی (Wainwright & Simoncelli, 2000) هر l_i به عنوان پاسخ یک فیلتر خطی وقتی که روی یک تکه تصویر اعمال می شود، تلقی می شود. به این متغیر اسکالر برای اشتباه نشدن با مقیاسهای موجک گاهی متغیر همزن^{۲۰۵} نیز گفته می شود. شکل ۳-۲۴ (بالا-چپ) یک مدل مولد دوبعدی ساده را نشان می دهد که در آن l_1 و l_2 توسط متغیر همزن مشترک v تولید می شوند. شکلهای ۳-۲۴ (بالا-راست) و ۳-۲۴ (پایین-چپ) نیز تابع توزیع حاشیه ای و شرطی توأم گوسی متغیرهای گوسی و همزن را برای داده هایی که از مدل ساخته می شوند، نشان می دهد.



شکل ۳-۲۴ (بالا-چپ): مدل مولد برای یک GSM دوبعدی. هر پاسخ فیلتر، l_1 و l_2 با ضرب متغیر گوسی خود، g_1 و g_2 ، در متغیر همزن مشترک v تولید می شود (بالا-راست) آمارگان توأم حاشیه ای و شرطی مؤلفه های گوسی مدل GSM (نمودار پایین شکل). برای آمارگان توأم شرطی، شدت روشنمایی با شمارش هر ظرف^{۲۰۶} هیستوگرام بدست می آید و هر ستون برای رسیدن به تمام بازه شدت روشنایی مستقلاً نرمال می شود. (پایین-چپ) توزیع حاشیه ای متغیر همزن مدل GSM. متغیر همزن طبق تعریف باید مثبت باشد و در اینجا توزیع رایلخ^{۲۰۷} با پارامتر $a=0.1$ برای آن برگزیده شده است. در هر حال توزیع دقیق متغیر همزن برای بدست آوردن خصوصیات آماری پاسخهای فیلتر که در تصویر پایین-راست نشان داده می شود چندان مهم نیست. (پایین-راست) آمارگان توأم حاشیه ای و شرطی پاسخهای فیلتر تولید شده.

مدل GSM یک توجیه بالا به پایین برای دو مشخه پایین به بالای آمارگان تصاویر طبیعی که قبلاً به آنها اشاره شد بدست می دهد: تابع توزیع حاشیه ای یک فیلتر با کورتوسیس بالا و آمارگان شرطی توأم دو فیلتر خطی که یک متغیر

^{۲۰۴} البته این مدل دارای نمایشهای دیگر مانند $\vec{x} = \sqrt{z}\vec{u}$ است که در کارکرد مدل تاثیری ندارد.

²⁰⁵ Mixer variable

²⁰⁶ bin

²⁰⁷ Rayleigh distribution

همزن مشترک دارند (Wainwright & Simoncelli, 2000; Wainwright et al., 2001). شکل ۳-۲۴ (بالا-چپ) توزیع حاشیه ای و شرطی توأم پاسخ دو فیلتر l_1 و l_2 را بر مبنای داده های ساخته شده در شکل ۳-۲۴ (بالا-راست) و ۳-۲۴ (پایین-چپ) نشان می دهد.

مدل GSM رابطه نزدیکی با روشهای پایین به بالا آمارگان تصویر و بازنمایی قشر مخ دارد. اولاً مدل‌های کدگذاری پراکنده و بازنمایی میدان گیرندگی قشر مخ معمولاً از خصوصیات زبرگوسی توزیعهای حاشیه ای پاسخهای فیلتر استفاده می کنند که به طور طبیعی در مدل مولد GSM بوجود می آید (شکل ۳-۲۴ پایین-راست). ثانیاً GSMها توجیهای برای مختصات فیلتر پیشنهاد می دهند از آن قبیل می توان به چارچوب حباب^{۲۰۸} (Hyvärinen, Hurri, & Vayrynen, 2003) اشاره کرد. وقتی که در مدل GSM پاسخ فیلترها از یک متغیر همزن مشترک استفاده می کند (شکل ۳-۲۴ پایین-راست) مختصات بوجود می آید. ثالثاً برخی از چارچوبهای پایین به بالا مستقیماً از نسخه هایی از دو مؤلفه GSM استفاده می کنند. مثلاً مدل‌های آمارگان تصویر و کنترل بهره قشر مخ^{۲۰۹} (Schwartz & Simoncelli, 2001) به نرمال کردن با تقسیم مؤلفه های خروجی می پردازند که دارای مشخصاتی است که از لحاظ آمارگان حاشیه ای و توأم یادآور مؤلفه گوسی GSM است (Wainwright & Simoncelli, 2000) (شکل ۳-۲۴ بالا-راست). در پژوهشی Ruderman & Bialek (94) نشان داده شده است که پیکسل‌های یک تصویر (نه حاصل کانولوشن پاسخ فیلتر با تصویر) را می توان به صورت حاصلضرب انحراف معیار محلی و یک مؤلفه تقریباً گوسی تجزیه کرد. در مجموع مدل GSM راه جالبی برای مجتمع کردن تعدادی از روشهای موثر آماری پیشنهاد می کند.

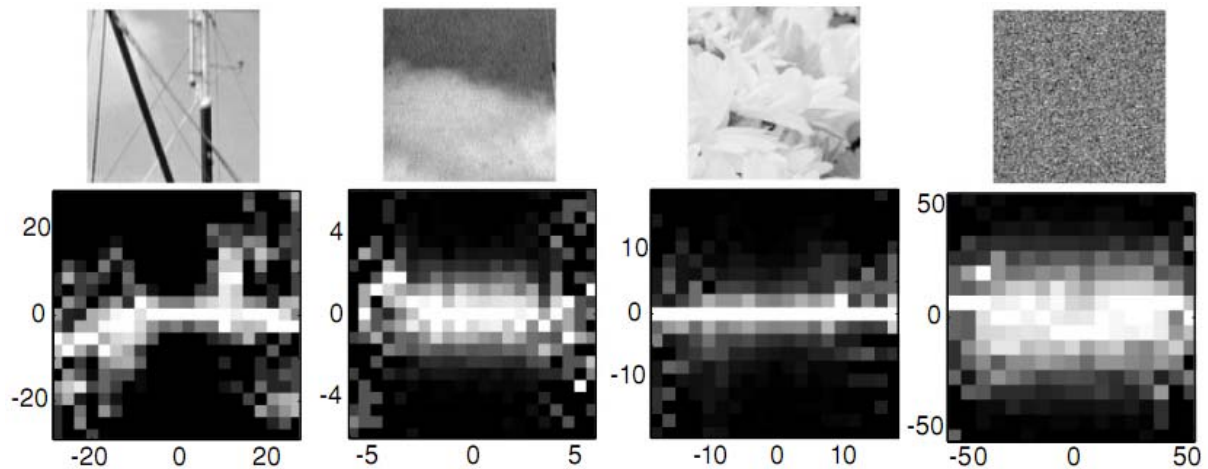
در فرمولاسیون اصلی GSM، یک همزن برای یک مجموعه از متغیرهای گوسی وجود دارد و لذا وابستگی پاپیون شکل آنها همگن است. با اینحال پاسخ یک بازه از فیلترهای خطی به تکه های تصویری توسط ناهمگنی در درجه وابستگی آماری آنها مشخص می شود. در مقاله ای (Wainwright & Simoncelli 2000) یک چینش از پیش مشخص شده سلسله مراتبی درختی در نظر گرفته شده است (و قاعده‌تاً متغیر همزن طوری تولید شده است که به درخت وابسته باشد). با اینحال برای بازه متنوعی از فیلترهای خطی و انواع گوناگونی از کلاسهای صحنه ها، باید چینش سلسله مراتبی از روی مثالها یاد گرفته شود. به علاوه از آنجایی که اشیا مختلف وابستگیهای خاص خود را القا می کنند، برای تکه های مختلف تصویر پینشهای مختلفی مناسب خواهد بود. مثلاً برای یک جفت فیلتر، قدرت وابستگی شرطی توأم ممکن است برای تکه های مختلف تصویر فرق کند. (شکل ۳-۲۵). این امر گویای اینست که بر پایه تکه های تصویر، همزنهای متفاوتی باید باید به فیلترهای مختلف انتساب شوند.

مدلهای مولد GSM یک راه ساده برای پارامتری کردن وابستگی آماری واریانس فعالیت فیلترهای لایه اول در اختیار می گذارند به طوری که برخی از خصوصیت‌های کلیدی پایین به بالای آماری تصاویر را اخذ می کند. با همه این اوصاف این مدل مبتنی بر یک بازنمایی خطی ثابت است (که برای ساختار تصاویر طبیعی بهینه سازی نشده است) و نیز به توصیف وابستگیهای جفتی بین خروجی فیلترهای خطی محدود است (Karklin 2007). با وجود اینکه GSM وابستگی فیلترهای

²⁰⁸ bubbles framework

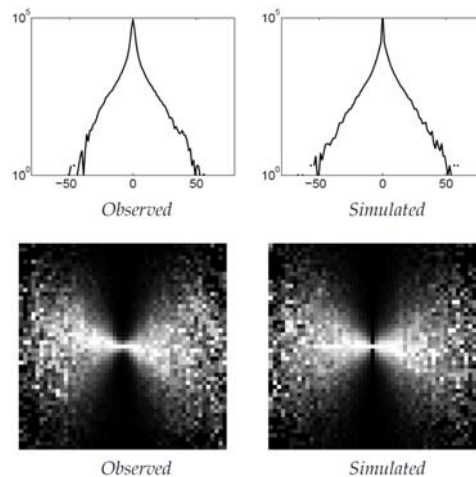
²⁰⁹ cortical gain control

خطی را فرموله می کند، اما این وابستگیها به تنهایی الگوی وابستگی بین فیلترها را نشان نمی دهند که این امر مانع اصلی کاربرد پایه های بالا به پایین برای مدلهای یادگیری سلسله مراتبی پایین به بالاست.



شکل ۳-۲۵: آمارگان توأم شرطی برای تکه های تصویری متفاوت و نویز سفید. آمارگان برای یک جفت فیلتر عمودی که از لحاظ مکانی همپوشانی ندارند. تکه های تصویری 100×100 پیکسل هستند. ، شدت روشنمایی با شمارش هر ظرف هیستوگرام بدست می آید و هر ستون برای رسیدن به تمام بازه شدت روشنایی مستقلاً نرمال می شود.

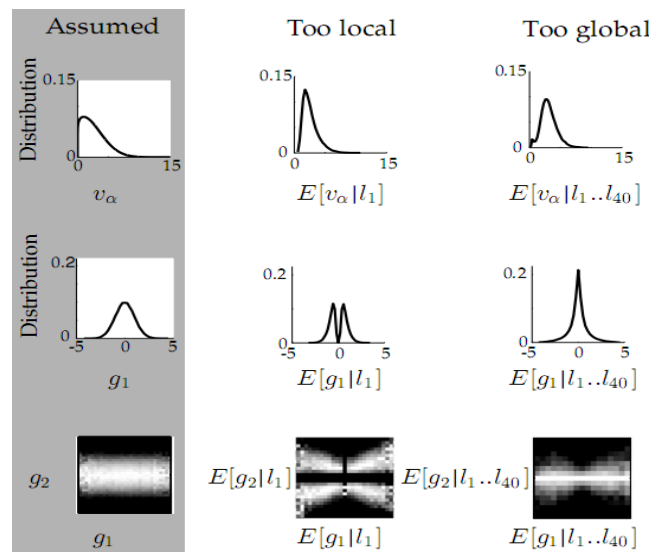
برای رفع این نقصها، توسعه هایی برای مدل GSM پیشنهاد شده است. برخی از مدلهای جدید تصویر، ضرایب موجک درون هر همسایگی محلی را با استفاده از یک GSM توصیف می کنند که رفتار شدیداً زبرگوسی توزیعیهای حاشیه ای ضرایب موجک تصاویر طبیعی و نیز همبستگی آنها در بزرگی محلی را مدل می کند (شکل ۳-۲۶).



شکل ۳-۲۶: مقایسه آمارگان ضرایبی که از یک زیرباند تصویر نمونه اخذ شده اند (چپ) با آنهایی که توسط تحریک یک مدل محلی GSM تولید شده اند (راست). پارامترهای مدل (ماتریس کواریانس و احتمال اولیه متغیر همزن) توسط پیشینه سازی درستنمایی ضرایب زیرباند تخمین زده شده اند. (بالا) هیستوگرام لگاریتمی توزیع حاشیه ای (پایین) هیستوگرام شرطی دو ضریب همسایه مکانی. شدت روشنایی با فرکانس رخداد متناسب است و در هر ستون برای رسیدن به تمام بازه شدت روشنایی مستقلاً نرمال می شود.

به عنوان مثال بارانیوک²¹⁰ و همکارانش یک متغیر همزن دو حالتی مخفی را برای مشخص کردن دو مدل رفتاری متناظر با نواحی دارای بافتها و ویژگیهای مسطح و یا با کنتراست کم استفاده کردند (Cruse 98, Romberg 99). مقاله های دیگری فرض کردند که واریانس محلی توسط یک متغیر پیوسته برای متغیر همزن هدایت می شود (LoPresto 97, Wainwright 2000, Mihcak 99, Wainwright 2001, Portilla 2003). برخی از مدل های GSM برای تصاویر با متغیرهای متغیر همزن v به گونه ای رفتار کردند که گویی متغیرهایی مستقلند که این فرض حتی در مورد همسایگی هایی از ضرایب که همپوشانی نیز داشتند استفاده شده است (LoPresto 99, Wainwright 2000, Portilla 2003). مدل های پیچیده تر وابستگی های بین متغیرهای همزن را توصیف کرده اند (Crouse 98, Romberg 99, Wainwright 2001).

یکی از توسعه های این مدل (Schwartz 2006) الگوهای وابستگی بین فیلترهای خطی را یاد می گیرد و نهایتاً یک بازنمایی سلسله مراتبی را ارائه می دهد. ایده اصلی این مقاله از (Karklin 2003a) آمده است و (لگاریتم) مقدار همزن برای هر فیلتر به صورت ترکیب خطی مقدارهای تعداد کمی از مؤلفه های همزن زیربنایی تولید می کند. در مقاله (Schwartz 2006) مسئله به این صورت مدل می شود که تعدادی متغیر همزن $\vec{v} = (v_\alpha, v_\beta, \dots)$ با فیلترهای خطی طوری خوشه بندی شوند که هر چند فیلتر خطی یک متغیر همزن را بطور مشترک استفاده کند. این شرح یک صورت مسئله انتساب²¹¹ را القا می کند که در آن فیلترهای خطی l_i و همزنهای v_j با هم جفت می شوند.



شکل ۳-۲۷: تخمین محلی و عمومی داه های تصویری (ستون چپ) توزیعهای فرض شده متغیرهای گوسی و همزن. (ستون وسط و راست) تخمینهایی بر اساس تعداد متفاوتی از پاسخهای فیلتر (چه یک فیلتر که خیلی محلی است و چه ۴۰ فیلتر شامل دو راستا در یک ناحیه ۳۸×۳۸ پیکسلی که خیلی عمومی است) (سطر بالا) توزیع تخمینی یکی از متغیرهای همزن. توجه شود که متغیرهای همزن طبق تعریف مثبت هستند. (سطر وسط) توزیع تخمینی یکی از متغیرهای گوسی (سطر پایین) آمارگان توأم تخمین توزیع شرطی متغیرهای گوسی.

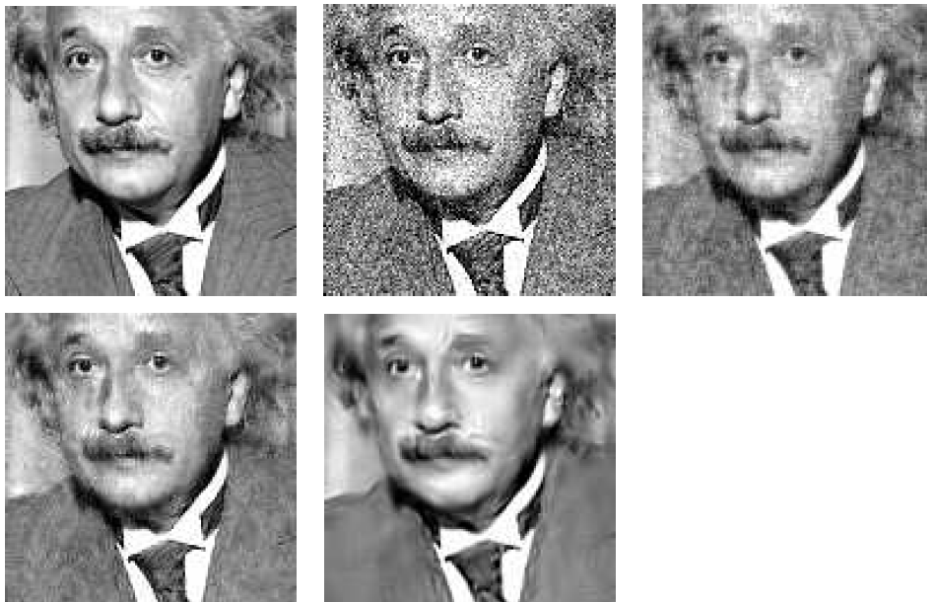
²¹⁰ Baraniuk

²¹¹ assignment problem

در این مقاله از تابعهای احتمالی رایلینخ دارای همپوشانی برای همزن ها استفاده شده است. روشهای انتساب دیگری برای سایر مدل‌های تصویری استفاده شده است که از آن قبیل می‌توان به مدلسازی پویای درخت^{۲۱۲} (Williams & Adams, 1999; Adams & Williams, 2003) و شبکه های اعتبار^{۲۱۳} (Hinton, Ghahramani, & Teh, 1999) اشاره کرد. شکل ۳-۲۷ نشان می‌دهد که چرا حل کردن مسئله انتساب مهم است.

ویژگیهای کیفی مدل نشان داده است که مدل وابستگی چندان زیادی به شکل دقیق $p(v)$ ندارد (Schwartz, 2006). استفاده از خانواده های دیگری از متغیرهای همزن مانند جذر توزیع گاما (Wainwright et al., 2001)، لگاریتم توزیع نرمال (Portilla et al., 2001)، و تابع احتمال اولیه جفری^{۲۱۴} (Portilla et al., 2003) دلیلی بر این مدعاست.

ساختار گوسی زیرین مدل GSM به این مدل اجازه می‌دهد که به مسائلی مانند رفع نویز نیز اعمال شود. تخمین زن از آنچه که برای مدل گوسی و مدل حاشیه ای موجه گفته شده بود پیچیده تر است ولی روی بازه گسترده ای از تصاویر و میزان نویز بهبود چشمگیری دارد (شکل ۳-۲۸). تخمین زن GSM هم از لحاظ آنچه به چشم می‌آید و هم از لحاظ میانگین خطای مربعی (Portilla 2003) به طور مشهودی بهتر عمل می‌کند.



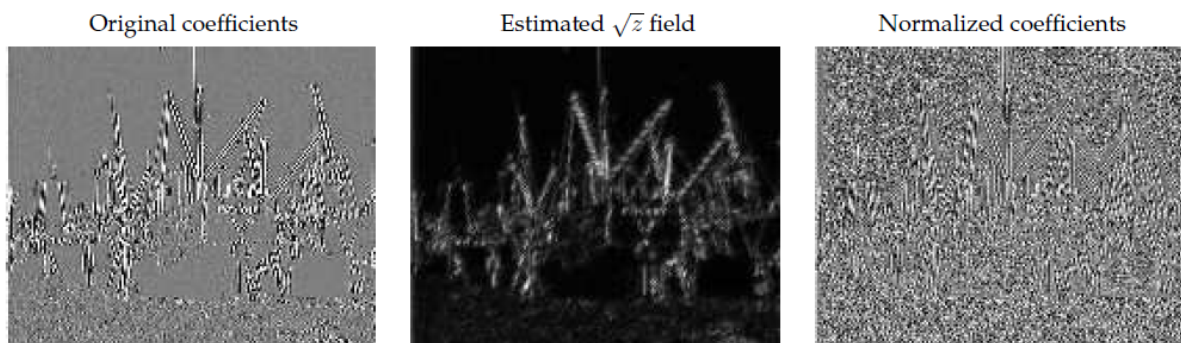
شکل ۳-۲۸: مثالهایی از رفع نویز. (بالا چپ) تصویر اصلی (انیشتین) (بالا وسط) تصویر آلوده شده به نویز سفید گوسی با $\sigma=21.4$ و نرخ قله سیگنال به نویز برابر $PSNR=22.06$. (بالا راست) رفع نویز با مدل حاشیه ای گوسی $PSNR=27.87$ (پایین چپ) رفع نویز با مدل حاشیه ای گوسی تعمیم یافته $PSNR=29.24$ (پایین راست) رفع نویز با مدل مخلوط مقیاس گوسی $PSNR=30.86$. تمام روشها در یک حوزه فراکامل موجه پیاده سازی شده اند. در هر حالت واریانس نویز معلوم فرض شده اند و فرایند رفع نویز به کمک کمینه خطای مربعی بیزی انجام شده است (یعنی میانگین توزیع احتمال ثانویه). پارامترهای مدل با بیشینه سازی درستنمایی داده ها تخمین زده شده است.

²¹² Dynamical tree modeling

²¹³ Credibility network

²¹⁴ Jeffry's prior

همانند دو مدلی که در بخشهای قبل به آن اشاره شد، نشانه هایی وجود دارد که بیانگر اینست که حتی توسعه های مدل GSM برای اخذ کامل ساختار تصاویر معمولی کافی نیست. برای نشان دادن این مدعا، باید خاطر نشان کرد که نرمال کردن هر ضریب با (جذر) واریانس تخمین زده آن باید یک حوزه نویز سفید گوسی را ایجاد کند (Wainwright 2000). شکل ۳-۲۹ این فرایند را مشخص می کند و مثالی از زیرباند موجک، حوزه واریانس تخمین زده شده و ضرایب نرمال شده را نشان می دهد. با وجود حذف افزونگی های تصویر و ساختارهای مرتبط با آنها، دو نوع ساختار مهم باقی مانده است. نخست، با وجود اینکه ضرایب نرمال سازی شده قطعاً به حوزه همگن نزدیکترند ولی علامت ضرایب همچنان ساختار مهمی را نشان می دهد. و دوم اینکه حوزه واریانس خودش هم از همگنی دور است زیرا بیشتر مقادیر مهم روی کانتورهای یک بُعدی متمرکز شده اند (Simoncelli 2005).



شکل ۳-۲۹: مثالی از زیرباند موجک (چپ)، جذر حوزه واریانس (وسط) و زیرباند نرمالسازی شده (چپ)

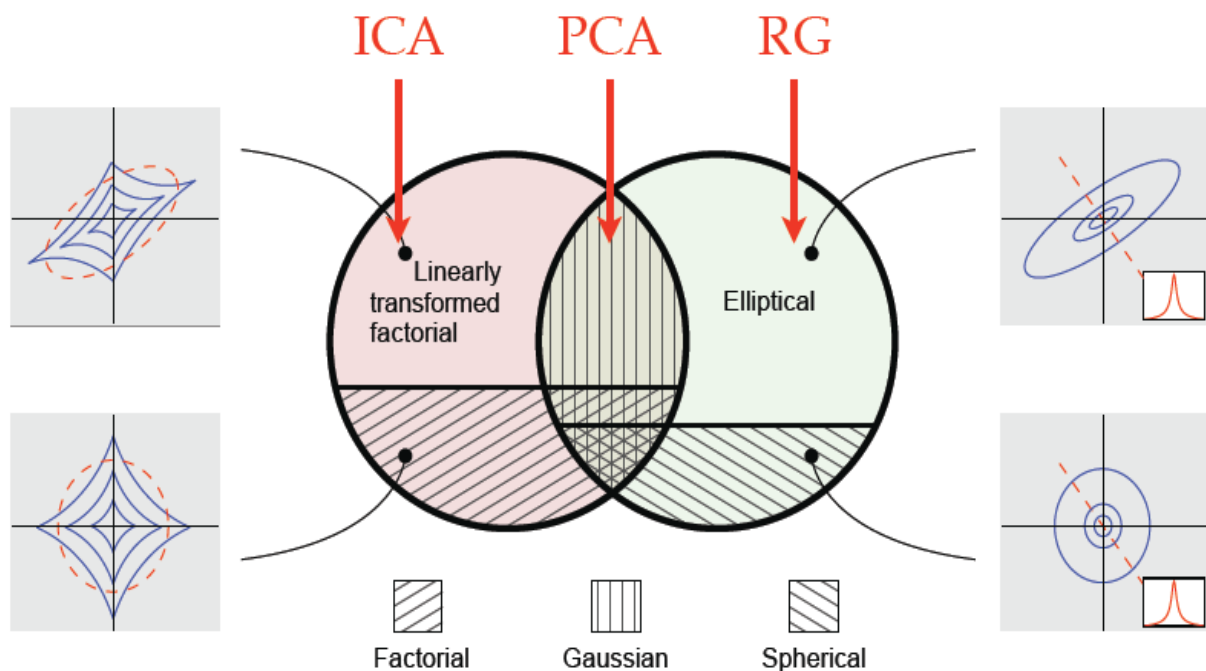
یکی از مشکلات دیگر از آنجا ناشی می شود که همسایگی یک فیلتر نه تنها باید توسط پهنای مکانی آن مشخص شود، بلکه باید به مشخصات ویژگی های آن، محتوای تصویر و کلاس تصویر نیز توجه داشته باشد. به عنوان مثال برای تکه هایی با بافت عمودی در یک تصویر از لبه و یک تصویر از چهره باید همسایگیهای متفاوتی برای فیلتر در نظر گرفت. بنابراین درجات متفاوتی از چینشهای عمومی یا محلی خاص مورد نیاز خواهد بود. از آنجایی که پیش از دریافت تصویر هیچ اطلاعات پیشینی درباره چگونگی تعیین گروههای متغیر همزن وجود ندارد باید انتساب آنها را از یک مجموعه تصویر نمونه یاد گرفت. به علاوه ممکن است برای تصاویر مختلف به گروه بندیهای مختلفی از متغیرهای همزن احتیاج باشد (Schwartz 2006).

۳-۱-۴-۲- روند پیشرفت

با گذشت نیم قرن از مدل گوسی/فوریه، اواخر دهه ۱۹۸۰ و دهه ۱۹۹۰ پیشرفت ناگهانی و شایان توجهی در دیدگاه داشته است که از تاثیر (الف) تجزیه تصویر چندمقیاسه (ب) مشاهدات و توصیفات غیرگوسی از تصویر و (ج) مدل های آماری سازگار با واریانس مبتنی بر متغیرهای نهفته ناشی شده است. بهبودهایی که در اثر ظهور این ایده ها در کاربردهای پردازش تصویر حاصل شده است مداوم و بزرگ بوده است ولی پیاده کردن کامل این ایده ها و پالایش آنها همچنان در حال تکوین است.

باید دقت داشت که روند پیشرفت بدلیل ذات علمی خود هر روز صحت مدلها و نظریه های موجود را به چالش می کشد. به عنوان مثال گفته شده است که پایه های PCA و ICA تقریباً برای هر منبعی قابل محاسبه است. ولی باید توجه

داشت که این دو روش تنها زمانی حذف گارانتی‌ها را تضمین می‌کنند که مدل منبع آنها صحیح باشد (منبع جدایی پذیر با تبدیل خطی^{۲۱۵})؛ که البته در همین شرایط هم که به نظر می‌آید این متدولوژیها پاسخهای جالبی تولید می‌کنند، مؤلفه‌های بازنمایی تولید شده از استقلال آماری فاصله زیادی دارند. این شرایط در جای خود اینگونه نشان می‌دهد که تبدیلهای ICA که شامل توابع پایه محلی جهتدار هستند، مشابه میدان گیرندگی نورونهای قشر مخ هستند. همانطور که اشاره شد، افزونگی در این روش اندکی بیشتر از روش PCA و سایر فیلترهای باندگذر کاسته می‌شود، ولی بازهم وابستگیهای مرتبه بالایی از خود نشان می‌دهد. حال اگر شرایطی در نظر گرفته شود که کاهش افزونگی برای کلاسهایی از مدل‌های منبع انجام شود که توزیعهای متقارن بیضوی^{۲۱۶} (EDS) نام دارند (شکل ۳-۳۰)، مشاهده می‌شود که تبدیلهای خطی هیچ تاثیری بر روی وابستگیهای بالاتر از مرتبه دوم ندارند و لذا روش ICA هیچ برتری نسبت به PCA ندارد (Lyu 2008d).



شکل ۳-۳۰: نمودار و رابطه بین مدل‌های توزیع. دو دایره توزیعهای جدایی پذیر با تبدیلات خطی (آنگونه که ICA فرض می‌کند) و توزیعهای متقارن بیضوی (ESDها) را نشان می‌دهد. اشتراک این دو کلاس مجموعه تمام توزیعهای گوسی است. توزیعهای جدایی پذیر زیرمجموعه ای از توزیعهای جدایی پذیر با تبدیلات خطی هستند و توزیعهای متقارن کروی زیرمجموعه ای از توزیعهای متقارن بیضوی هستند.

توزیع بردار تصادفی $\vec{x} \in \mathbb{R}^d$ با میانگین صفر بیضوی است اگر به فرم زیر باشد:

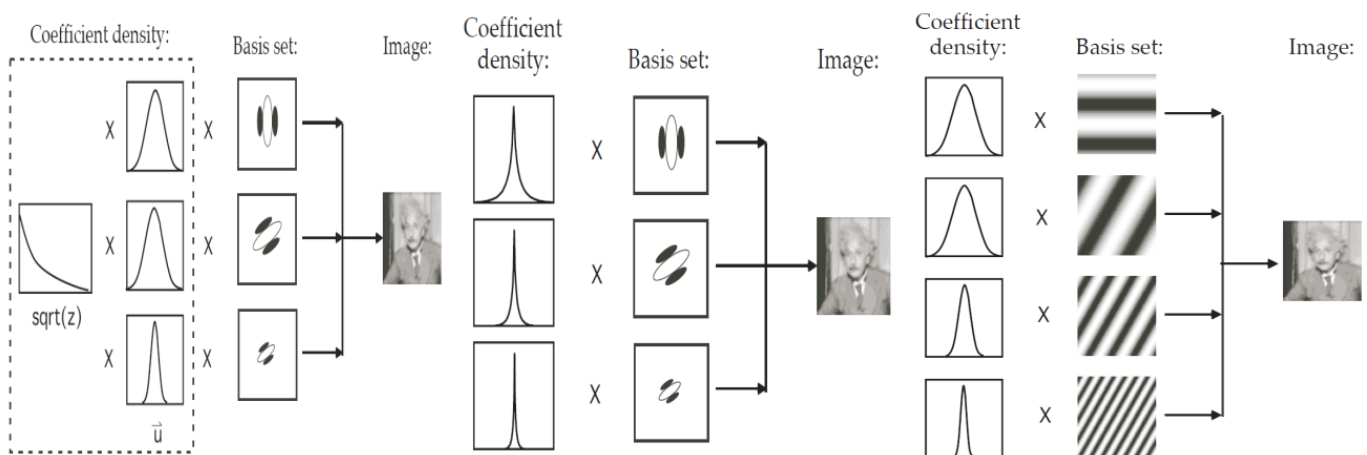
$$p(\vec{x}) = \frac{1}{\alpha |\Sigma|^{\frac{1}{2}}} f\left(-\frac{1}{2} \vec{x}^T \Sigma^{-1} \vec{x}\right) \quad (۱۶-۳)$$

²¹⁵ Linearly transformed factorial source

²¹⁶ elliptically symmetric densities

که در آن Σ یک ماتریس مثبت-معین باشد، $f(\cdot)$ تابع مولدی باشد که شرط $\int_0^\infty f(-r^2/2)r^{d-1}dr < \infty$ و $f(\cdot) \geq 0$ را ارضا کند، و ثابت نرمال کننده طوری انتخاب شود که انتگرال توزیع برابر یک باشد. مشخصات تعریف کننده یک ESD طوری است که محدوده های توزیع ثابت آن بیضی هایی باشد که Σ تعریف می کند. در حالت خاص وقتی که Σ مضربی از ماتریس همانی باشد، محدوده های توزیع ابرکره هستند و توزیع به نام توزیع متقارن کروی^{۲۱۷} (SSD) نامیده می شود. اگر \vec{x} دارای آمارگان مرتبه دوم متناهی باشد، Σ ضریبی از ماتریس کواریانس آنست که به طور ضمنی بیانگر اینست که یک ESD با یک عملگر PCA یا وایتنینگ قابل تبدیل به SSD است. مقاله (Lyu 2008d) برای این کلاس از مدل های منبع تصویر، روش جایگزین غیرخطی برای حذف افزونگی های مراتب بالاتر پیشنهاد می کند و روش خود را گوسی سازی شعاعی^{۲۱۸} می نامد.

شکل ۳-۳۱ خلاصه ای از مدل های گوسی، مدل های پایه های خطی، و مدل های توزیع توأم موجک را نشان می دهد. نسخه هایی از مدل GSM که در این بخش آورده شده است، چه از نظر شکل دهی به توزیع ضرایب و چه از نظر کیفیت عملکرد آنها در کاربردهای پردازش تصویر، از به روزترین^{۲۱۹} روشهاست. برخی مسائل وجود دارد که حائز اهمیت بالایی برای توسعه چنین مدل هایی است. نخست تلاش برخی از محققان در بررسی روشهایی برای توصیف قاعده مندیهها در حوزه واریانس محلی است. این روشها شامل حوزه های تصادفی مکانی^{۲۲۰} (Hyvarinen 2003, Karklin 2003, Hyvarinen 2000) و مدل های درخت شکل چندمقیاسه (Romberg 99, Wainwright 2001) است. بسیاری از ساختارها در حوزه واریانس ممکن است خاصیت هایی شبیه ویژگی های ناپیوسته مانند لبه ها، خطوط، و گوشه ها داشته باشد. درباره این ساختارها مطالب زیادی در ادبیات بینایی رایانه ای وجود دارد ولی تهیه مدلی که هم صریح و هم انعطاف پذیر باشد دشوار است. مطالعات دیگری نیز روی قاعده مندیه های هندسی که از امتداد کانتور یا مرزها بست می آید صورت گرفته است (Sigman 2001, Elder 2002, Giesler 2001, Hoyer 2002, Zhu 2003).



شکل ۳-۳۱: خلاصه ای از مدل های گوسی (راست)، مدل های پایه های خطی (وسط)، و مدل های توزیع توأم موجک (چپ)

²¹⁷ spherically symmetric density

²¹⁸ radial gaussianization

²¹⁹ state-of-the-art

²²⁰ spatial random fields

این ساختارها و سایر ساختارهای موجود در تصاویر بدون شک در مدل‌های آماری آینده جایی برای خود دست و پا خواهند کرد و کاربردهای پردازش تصویر را بیش از پیش بهبود خواهند داد.

۳-۱-۴-۳- ساختار سلسله مراتبی

همانگونه که اشاره شد مغز انسان با یک ساختار سلسله مراتبی داده‌های ورودی را پردازش می‌کند و هر لایه از لایه‌های قبلی اطلاعات بیشتری را استخراج می‌کند و سیگنال‌های پردازش شده را به لایه‌های بعدی خود تحویل می‌دهد. در مدل‌های پایین به بالا نیز این ساختار حفظ می‌شود و طراحان مدل سعی می‌کنند با طراحی تابع فعالیت نورونها، میدان گیرندگی آنها و شکل بهم بندی و اتصال نورونها، به پدیده‌های مشاهده شده در آزمایشات نوروبیولوژیکی نزدیک شوند و بتوانند توجیهی برای رفتارهای مشاهده شده نورونها یافته، سایر رفتارهای تکی و جمعیتی نورونها را نیز پیش بینی کنند. با در نظر داشتن این اهداف یعنی نزدیکی به ساختار مغز به عنوان مدل هدف، نزدیک شدن به مدل‌های پایین به بالا برای استفاده از مزایای آنها، و جبران ضعفهایی از مدلها که پیشتر بیان شد، برخی از پژوهشگران دست به افزودن لایه‌هایی به مدل‌های موجود زده اند و سازمان سلسله مراتبی را به خدمت گرفته اند. در ادامه به ایده‌هایی سلسله مراتبی پرداخته می‌شود که اخیراً جای خود را در ادبیات فهم تصاویر طبیعی باز کرده اند و در توجیه لایه‌های بالاتر قشر خاکستری مخ به توفیق‌هایی دست یافته اند.

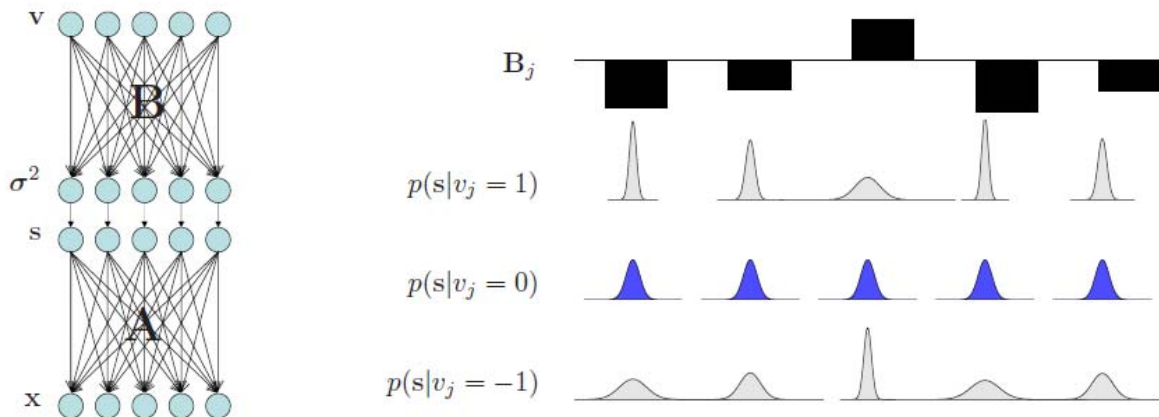
نخستین ایده ای که معرفی می‌شود، در حوزه ICA اضافه کردن لایه دومی به لایه اول فیلترها بود که واریانس فعالیت‌های لایه اول را مشخص می‌کرد (Hoyer, 2000a, 2000b; Hoyer & Hyv^{arinen}, 2002; Karklin & Lewicki, 2003a, 2005; Park & Lee, 2005). کارکلین و لویکی (Karklin and Lewicki 2003a, 2005) مدلی را پیشنهاد کرده اند که در آن واریانس هر واحد در لایه اول از ترکیب جمع مجموعه ای از واحدهای تابع پایه واریانس در لایه دوم تشکیل شده است. روش شوارتز برای مدل GSM (Schwartz 2006) را می‌توان نسخه ای از این روش انگاشت که بجای ترکیب همکاری کننده، واحدهای لایه دوم دارای ترکیبی ترکیبی رقابتی دارند. در پایان نیز مدل تحلیل وابستگی‌های کواریانسی معرفی می‌شود که شالوده اصلی شبیه سازی نرونهای این مقاله را در بر دارد.

سلسله مراتب در مدل تحلیل مؤلفه مستقل: مقاله ای که در سال ۲۰۰۰ منتشر شد شرح کاری بود که سعی داشت تحلیل مؤلفه‌های مستقل را گسترش دهد و گزینندگی نرون‌های پیچیده را شبیه سازی کند (Hyvarinen 2000). ایده اصلی در توسعه این مدل، وابستگی بین ضرایب مؤلفه‌های مستقل تصاویر طبیعی بود. در مدل ICA زیرفضا (Hyv^{arinen} and Hoyer, 2000) ضرایب خطی دیگر مستقل فرض نمی‌شوند و بجای آن گروههایی از توابع پایه باهم همسایگی (زیرفضاها) یی را می‌سازند که در هر کدام انرژی ضرایب با هم همبسته است و به همین دلیل فعالیت در آن یا بسیار زیاد و یا نزدیک به صفر است. ایده اصلی در توسعه این مدل، وابستگی بین ضرایب مؤلفه‌های مستقل تصاویر طبیعی بود. پاسخ مؤلفه‌های مستقل به دسته‌هایی تقسیم می‌شد که پاسخ‌های درون هر دسته به یکدیگر وابسته و پاسخ‌های دسته‌های متفاوت از یکدیگر مستقل بودند. هر کدام از این دسته‌ها زیرفضای مستقل ویژگیها نامیده شد. برای شبیه سازی خصوصیات نرون‌های پیچیده مربع ضرایب مؤلفه‌های مستقل با یکدیگر ترکیب شدند تا پاسخ نرون‌های پیچیده را شبیه سازی کنند. برای شبیه سازی رفتار غیرخطی نرون‌های پیچیده، این مدل از مربع پاسخ فیلترهای خطی استفاده می‌کند و یک ترکیب خطی از این پاسخ‌ها را به عنوان پاسخ نرون پیچیده مدل می‌کند. سپس توابع پایه خطی برای بیشینه کردن استقلال نرُم

برداری هر همسایگی (بجای استقلال تک تک ضرایب) تطبیق می یابند. با توجه به این فرض که متغیرهای درون یک زیرفضا دارای توان همبسته ای هستند، این ساختار وابستگی میان بزرگی ضرایب را می گیرد و به مدل اجازه می دهد که زیرفضاهای بهینه را برای توصیف وابستگی بیابد.

در این مدل توابع پایه در لایه اول قرار دارند و نمایندگان یک همسایگی آنها که در لایه بالاتر قرار گرفته است و حاصل ترکیب خطی مربعات آنهاست برای کسب بیشینه استقلال مورد بهینه سازی قرار می گیرد و تغییرات آنها به لایه نخست منعکس می شود.

سلسله مراتب در مدل وابستگیهای واریانس: یکی از نقاط ضعف تحلیل مؤلفه های مستقل این است که مؤلفه های حاصل، محدود به خصوصیات نورو ن های قشر ابتدایی بینایی می باشند که خصوصیتی نسبتاً ساده از تصاویر هستند. این روش قابلیت گسترش به خصوصیات سطح بالاتر را ندارد زیرا ضرایب این مؤلفه ها را مستقل در نظر می گیرد. کارکلین در سال ۲۰۰۳ با استفاده از وابستگی بین ضرایب این مؤلفه ها توانست مدلی سلسله مراتبی بسازد که توانایی یادگیری خصوصیات سطح بالا را از تصاویر دارد (Karklin 2003). ایده اصلی در این مقاله استفاده از وابستگی واریانس بین مؤلفه های مستقل بود. برای مدل کردن وابستگی واریانس، یک لایه بر روی مدل مؤلفه های مستقل اضافه شد که هدف آن تعیین واریانس برای ضرایب مؤلفه های مستقل بود (شکل ۳-۳۲). در لایه جدید، هر مؤلفه یک عامل واریانس نامیده شد که با یک ترکیب خطی، واریانس ضرایب مؤلفه های مستقل را تعیین می کرد.



شکل ۳-۳۲: (چپ) یک بازنمایی گرافیکی از مدل سلسله مراتبی برای وابستگی واریانس. واریانس ضرایب s تابعی از متغیرهای نهفته مرتبه بالاتر v هستند. پارامترهای مدل B تبدیلی خطی از v به $\log \sigma^2$ را مشخص می کنند و پارامترهای A تبدیلی از s به x را مشخص می کنند (در این مرحله نویز اضافه گوسی نیز می تواند اضافه شود، $x=As+\epsilon$ ، ولی در شکل آورده نشده است) (راست) نمایشی از تاثیر آزمین ستون B بر توزیع توأم $p(s)$ با فرض آنکه تمام متغیرهای مرتبه بالای دیگر v_k ($k \neq j$) برابر صفر باشد. برای روشن شدن مطلب توزیعهای شرطی $p(s|B, v)$ گوسی انتخاب شده اند ولی انواع دیگری از پارامتری کردن نیز قابل قبول است.

در این مدل، ضرایب مؤلفه های مستقل از یک توزیع گوسی تعمیم یافته (معادله ۳-۱۱) در نظر گرفته شده اند:

$$p(\bar{s}) = N(0, \lambda, q) = Z \exp\left(-\left|\frac{\bar{s}}{\lambda}\right|^q\right) \quad (16-3)$$

در این معادله، q وزن دنباله های توزیع و λ واریانس توزیع را مشخص می کند. در تحقیقاتی که روی تصاویر طبیعی انجام شد مشخص گردید که توزیع ضرایب له دلیل وابستگی واریانس و مفهوم، از معادله بالا منحرف می شود. کارکلین این تغییر را با کنترل λ مدل می کند. در مدل ارائه شده، واریانس هر عنصر از ضرایب مؤلفه های مستقل از یک ترکیب خطی از مؤلفه های واریانس به دست می آید:

$$\lambda = \exp([B\bar{v}]) \quad (17-3)$$

در معادله فوق، B ماتریس توابع پایه واریانس و \bar{v} ضرایب این توابع را نشان می دهند. در این مدل، توزیع احتمال \bar{v} از نوع گوسی با میانگین و واریانس ثابت در نظر گرفته شد. وابستگی بین ضرایب مؤلفه های مستقل در این مدل با ترکیب خطی مجموعه ای از مؤلفه های واریانس پیاده سازی می شود. در حقیقت آنچه باعث وابستگی بین مؤلفه های \bar{s} در معادله ۱۶-۳ می شود، ماتریس B است که ترکیب مؤلفه های واریانس را مشخص می کند.

برای تخمین ماتریس B از تحلیل درستیابی بیشینه استفاده می شود. در این تحلیل، به ازای یک مجموعه از تصاویر طبیعی، ماتریس B به نحوی تعیین می شود که احتمال مشاهده داده ها بیشینه شود:

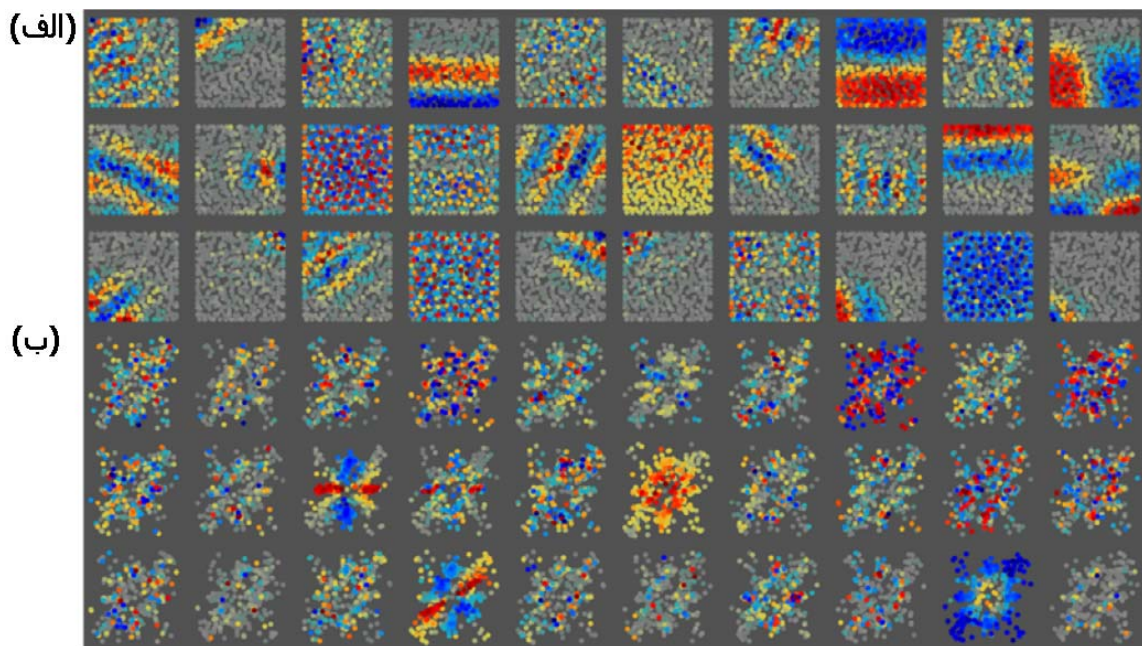
$$p(x_1, x_2, \dots, x_n) = \prod_i p(x_i | A, B) \quad (18-3)$$

در معادله ۱۸-۳ فرض بر این است که تصاویر موجود x_i از یکدیگر مستقل هستند. در این معادله A ماتریس مؤلفه های مستقل را نشان می دهد که به صورت جداگانه تعیین شده است. شکل ۳-۳۳ مجموعه ای از مؤلفه های واریانس به دست آمده توسط این مدل را نمایش می دهد.

اهمیت این مدل در به دست آوردن خصوصیات سطح بالا از تصاویر طبیعی با استفاده از مفهوم وابستگی واریانس می باشد. مدل های خطی مانند تحلیل مؤلفه های مستقل با نادیده گرفتن قسمت مهمی از وابستگی های موجود در تصاویر طبیعی، قابلیت گسترش به خصوصیات سطح بالا را ندارند. این امر به دلیل سادگی مدل بیان شده در این روش می باشد. مدل کارکلین از این جهت حائز اهمیت است که نشان می دهد مفهوم وابستگی در تصاویر طبیعی محدود به دامنه ای خاص در تصویر نمی باشد.

بر خلاف سایر مدل های تحلیل تصاویر، مدل های واریانس (Karklin 2003a) و کواریانس کارکلین (Karklin 2008) (که در ادامه همین قسمت به آن اشاره خواهد شد) ساختارهای آماری غیرخطی را با استفاده از بازنمایی توزیع آنها کدگذاری می کند که این رهیافت احتمالاً در مغز نیز استفاده می شود. مدل های پیشنهاد شده همچنین نسبت به روش های یادگیری بدون ناظر استاندارد تخمین توزیعی بهتری از تصاویر طبیعی بدست می دهند. با وجود آنکه داده ها روی داده های

نورونی تطبیق داده نشده اند، تعدادی از خاصیت‌های کلاسیک نورونهای ناحیه V1، و نیز اثرات ظریفی که در V2 و V4 دیده می‌شوند را بروز می‌دهند که یک توصیف کارکردی برای این رفتارهای نورونی را ارائه می‌کند.



شکل ۳-۳۳: مجموعه‌ای از مؤلفه‌های واریانس پایه که از تصاویر طبیعی استخراج شده‌اند. هر مؤلفه یک الگوی مشخص برای انحراف از شرط مستقل بودن را نشان می‌دهد. مؤلفه‌ها بر اساس (الف) موقعیت مکانی و (ب) جهت مؤلفه‌های مستقل متناظرشان مرتب شده‌اند.

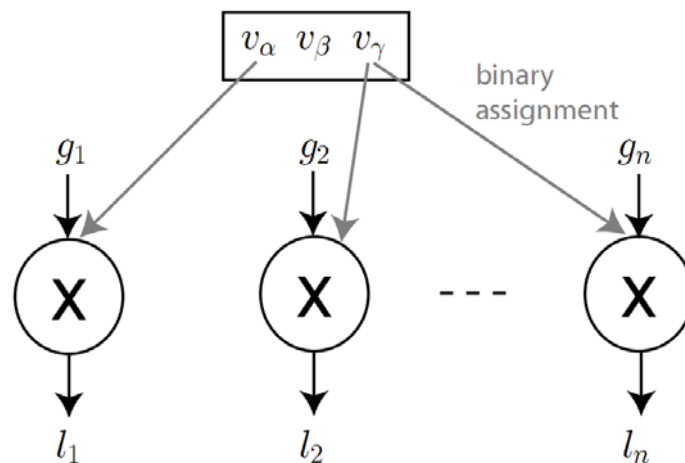
در این مدل توابع پایه در لایه اول قرار دارند که از ترکیب خطی آنها با ضرایبی که ماتریس A در بردارد تصویر نهایی ساخته می‌شود. در بین این توابع پایه، وابستگی واریانسی به عنوان تابعی از متغیر نهفته \vec{v} فرض شده است. لایه دوم این مدل متغیرهای نهفته و ارتباط آنها برای تولید تک تک واریانسها را مدل می‌کند. در این مدل محتمل‌ترین مجموعه متغیرهای \vec{v} و ماتریس مربوط به ترکیب آنها، B ، از استنتاج بدست می‌آید و سپس مؤلفه‌های مستقل از روی آنها در لایه اول ساخته می‌شود.

سلسله مراتب در مدل مخلوط مقیاسهای گوسی: یکی از مدل‌هایی که مدل کارکلین را گسترش داد، مقاله ای بود که در ۲۰۰۶ منتشر شد (Schwartz06). در این مقاله مسئله به این صورت مدل می‌شود: تعدادی متغیر همزن $(v_\alpha, v_\beta, \dots) = \vec{v}$ با فیلترهای خطی طوری خوشه بندی شوند که هر چند فیلتر خطی، یک متغیر همزن را بطور مشترک استفاده کند. این شرح یک صورت مسئله انتساب را القا می‌کند که در آن فیلترهای خطی l_i و همزنهای v_i با هم جفت می‌شوند. استنتاج کردن این مسئله انتساب دقیقاً همان فرایندی است که سطوح سلسله مراتب را در مدل آماری استنتاج می‌کند. با وجود اینکه این مدل از مدل اصلی GSM پیچیده تر است ولی استنتاج آنها با کمک روشهای پیشینه سازی امید ریاضی (EM)^{۲۲۱} و نمونه برداری مونت کارلو زنجیره مارکوف^{۲۲۲} (به عبارت دقیقتر، روش نمونه برداری گیبس) انجام شده است.

²²¹ expectation maximization

در این مقاله، که بر مبنای مدل مؤلفه های مشترک واریانس می باشد، یک مجموعه از عوامل واریانس (متغیرهای همزن در GSM) بر لایه ای از ضرایب مؤلفه های مستقل (توزیعهای گوسی چند متغیره در GSM) اضافه شده اند. این مدل نیز بر اساس وابستگی واریانس بین پاسخ مؤلفه های مستقل می باشد. تفاوت این مدل با مدل کارکلین در نحوه انتساب مؤلفه های واریانس به ضرایب مؤلفه های مستقل می باشد. در مدل کارکلین، ضرایب واریانس با استفاده از ماتریس مؤلفه های واریانس با یکدیگر ترکیب می شدند تا واریانس ضرایب مؤلفه های مستقل را تعیین کنند ولی در مدل شوارتز، ضرایب واریانس با یکدیگر رقابت می کنند و برنده رقابت به مؤلفه ها در لایه ای پایین تر نسبت داده می شود.

به ازای هر ضریب واریانس و مجموعه ای از ضرایب مؤلفه های مستقل، یک مجموعه احتمال محاسبه می شود که احتمال نسبت دادن ضریب واریانس به مؤلفه های مستقل را نشان می دهد. پس از محاسبه احتمال به ازای تمام ضرایب واریانس، ضریب با بیشترین مقدار احتمال به هر ضریب مؤلفه مستقل نسبت داده می شود. شکل ۳-۳۴ این مدل را شرح می دهد.



شکل ۳-۳۴: مدل انتساب احتمالی یک مدل GSM چند بعدی. پاسخ فیلترهای $l = \{l_1, \dots, l_n\}$ از ضرب متغیرهای گوسی $g = \{g_1, \dots, g_n\}$ در متغیرهای $\{v_\alpha, \dots, v_\mu\}$ که فرض می شود که $\mu < n$ می توان هر مخلوط l_i را به عنوان پاسخ فیلتر خطی وقتی که روی یک تکه تصویر خاص اعمال می شود در نظر گرفت. مسئله انتساب برای تولید پاسخ فیلتر l_i می پرسد که کدام متغیر همزن v_j باید به کدام متغیر گوسی g_i نسبت داده شود. مجموعه متغیرهای همزن $v_\alpha, v_\beta, v_\gamma$ در مستطیل سیاه رنگ مشخص شده است. پیکانهای خاکستری رنگ انتساب باینری را نشان می دهند. l_1 با همزن v_α تولید می شود و l_2 و l_n با همزن مشترک v_γ تولید می شوند.

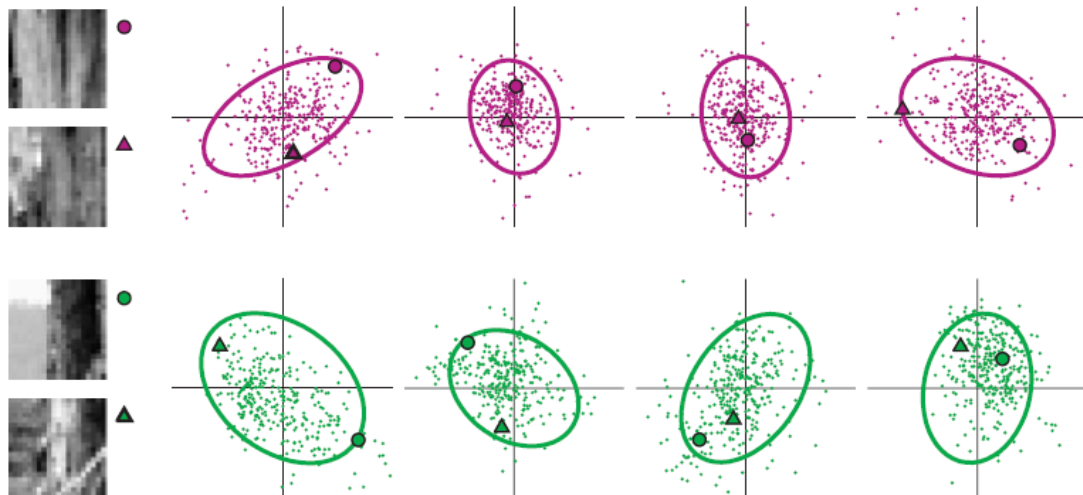
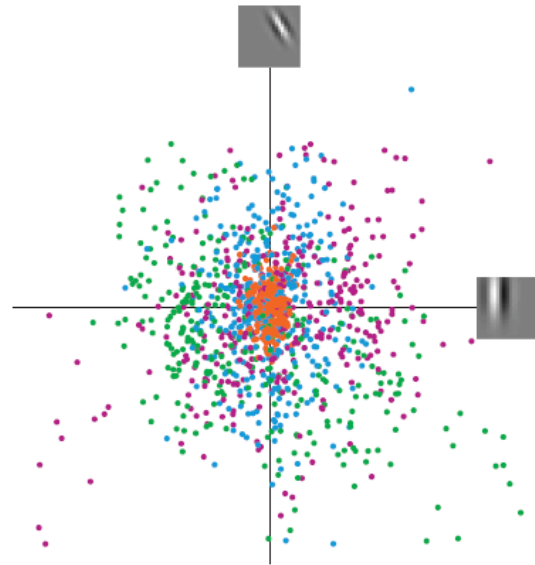
در این مدل توابع پایه گوسی در لایه اول قرار دارند که از ترکیب خطی آنها با متغیرهای همزن \bar{v} پاسخ فیلترها ساخته می شود. در بین این توابع پایه، وابستگی واریانسی به عنوان تابعی از متغیر نهفته \bar{v} فرض شده است به طوری که چندتا از توابع پایه متغیر همزن مشترکی را استفاده می کنند. لایه دوم این مدل متغیرهای همزن (نهفته) را مدل می کند. در این مدل محتمل ترین انتساب یک عضو مجموعه متغیرهای همزن به هریک از توابع پایه، از استنتاج بدست می آید و سپس پاسخ فیلترها از روی آنها در لایه اول ساخته می شود.

سلسله مراتب در مدل وابستگیهای کواریانس: هنگامی که تصویری از یک صحنه پیچیده طبیعی را نگاه می‌کنیم، در چندین مکان مختلف (مثلاً تنه یک درخت و یا در لبه‌های آن) به نقاط تمرکز^{۲۲۳} خاصی بر می‌خوریم که حتی با وجود اینکه تک تک تصاویری که روی شبکه می‌افتد ذاتاً دارای تنوع زیادی است، درکی منطقی از ساختار زیرین (بافت پوست یا کانتور لبه درخت) بدست می‌دهد. شکل ۳-۳۵ مسئله‌ای را که مغز ما به راحتی حل کرده است را شرح می‌دهد: نواحی تصویر که از لحاظ ادراکی متمایزند الگوهای پاسخی را می‌سازند که همپوشانی زیادی دارند و نمی‌توان آنها را با هیچ بازنمایی خطی سطح پایین از هم تمیز داد. پس چه نوع محاسباتی برای بدست آوردن تعمیم مناسب برای محرک تصاویر طبیعی لازم است؟

نورونهای بینایی اولیه را معمولاً به صورت کشف‌کننده‌های خطی بینایی توصیف می‌کنند (Hubel, Movshon 78, 62) و مدلهایی که حول این ایده تولید شده‌اند با دقت بالایی می‌توانند رفتار نورونها را، از شبکه (Chichilinsky 01) تا سلولهای ساده در قشر بینایی مخ (Carandini 97) اخذ کنند. ولی همانگونه که شکل ۳-۳۵ نشان می‌دهد، نه تک تک ویژگیها و نه تبدیلات خطی نمی‌توانند تصاویری با یک ساختار را از تصاویر دیگر جدا کند. در لایه‌های بعد سیستم بینایی احتمالاً ویژگیهای مجردتری محاسبه می‌شوند ولی دانش ما در این مورد فعلاً محدود است. در لایه V1، سلولهای پیچیده در مکانهای مختلفی به لبه پاسخ می‌دهند (Hubel 62) ولی مدل‌های کلاسیک این سلولها (Movshon 78b, Adelson 85) از تشریح برخی از پدیده‌های غیرخطی مشاهده شده در این لایه، مانند بازنشانی اطرافیان و بازداری فعالیت به ازای راستاهای متقاطع عاجزند (Bonds 89, Jones 2002, Cavanagh 2002). از آن مهمتر اینست که هیچ توجیه کارکردی برای نقش این رفتارها در ادراک صحنه‌های طبیعی نیز بدست نمی‌دهند. در لایه‌های بالاتر بینایی نظیر V2 و V4، نورونها نسبت به خصوصیات تصویر مانند مکان و مقیاس بیشتر نامتغیرند (Kabotake Gallant 96, Connor 2007, Hedge 2000) و به نظر می‌رسد که شکل یا بافت را کدگذاری کنند (Gallant 96, Connor 2007). برای تعمیم کارای این نورونها، مدار نورونی باید یک بازنمایی تولید که از یکسو برای پراکندگی وسیع صحنه‌های طبیعی از یک نوع (مثلاً بافت یا کانتور) مشابه باشد و از سوی دیگر برای پراکندگی وسیعتر تمام صحنه‌های طبیعی دیگر متمایز باشد. کارهای نظری پیشین نشان داده است که نورونها در بخش اولیه قشر خاکستری مخ یک کد بهینه تشکیل می‌دهند که با آمارگان تصاویر طبیعی سازگار است (Olshausen 96, Bell 97) اما هیچ اطلاعاتی درباره این نمی‌دهد که چگونه نورونها روی تنوع ذاتی این عناصر تصویر تعمیم می‌یابند.

روش حاضر توسعه‌ای بر روش کدگذاری کاراست و پیشنهاد می‌دهد که بازنمایی توزیعهای بیشمار آمارگانی که نواحی محلی تصویر را معین می‌کنند، وجه مهمی از محاسبات بصری است. نورونهای لایه‌های بعد، به جای اینکه شدت روشنایی پیکسل‌های یک تکه تصویر از بافت یا لبه را کدگذاری کنند، توزیع تصویر (یعنی بازه و الگوی تنوع شدت روشنایی پیکسلها یا ویژگیهای تصویر) را که با تصویر ورودی همخوانی بیشتری دارد کدگذاری می‌کند. این نحوه کدگذاری باعث می‌شود که بازنمایی نورونی روی نقاط تمرکز قابل تعمیم باشد و خاصیت‌های مجردتری از تصویر را منتقل کند. نشان داده شده است که مدلی که حول این هدف محاسباتی طراحی می‌شود و برای صحنه‌های طبیعی بهینه‌سازی می‌شود، خواص غیرخطی سلولهای پیچیده و نورونهای لایه‌های بالاتر را نیز در بر می‌گیرد و لذا یک تفسیر کارکردی از این پدیده‌ها را ارائه می‌کند.

223 fixation



شکل ۳-۳۵: الگوهای آماری، نواحی مختلف صحنه های طبیعی را متمایز می کند. (بالا چپ) یک صحنه طبیعی با چهار ناحیه مشخص شده (برگرفته از (Doi 2005)). (بالا راست) نمودار پراکندگی نشان دهنده الگوی پراکندگی زوجهای باینده های فیلتر خطی (فیلترهای خطی راستدار) برای ۲۰۳۲۰ تکه تصویر نمونه گیری شده از چهار ناحیه مشخص می باشد. خروجی نواحی مختلف دارای همپوشانی زیادی است و مبین آنست که ویژگیهای خطی برای تمایز آنها هیچ راهی ندارند. (پایین) هر ستون خروجی توأم جفت متفاوتی از باینده های خطی هستند که به نواحی که لبه درخت و پوست درخت قرار دارند مربوطند (اولین ستون همان ویژگیهایی هستند که در تصویر بالا-راست آورده شده اند). همبستگی در هر قسمت می تواند با یک توزیع گوسی و کواریانس آن توصیف شود (بیضی ها). تفاوت در توزیعهای بین سطرها مشخصات الگوهای درون همبستگیها را آشکار می کند که وقتی ویژگیهای بیشتری مد نظر قرار بگیرند واضحتر نیز می شود. این الگوها می توانند برای تعمیم درون نواحی استفاده شوند، در حالی که همچنان بتوانند آنها را تمایز بدهند. به عنوان مثال در هر ناحیه دو تکه تصویر مشخص شده و با مثلث و دایره نمایش داده شده است. با وجود آنکه جفتهای تصویر به وضوح تفاوت زیادی دارند، هر تصویر با توزیع ناحیه تصویر خود سازگار است. با بررسی مقابله ای توزیع در ابعاد مختلف، استنتاج نوع تکه های تصویر ممکن است، حتی با وجود این امر که تکه های مختلف تصویر روی بعضی از ویژگیها افکنش مشابهی داشته باشند.

اساساً تعمیم عبارتست از تشخیص مشخصات مشترک یک کلاس از روی نمونه های آن. هدف مدل ارائه شده نیز یادگیری توزیع آماری ایست که مشخصات نواحی محلی تصویر را در بر بگیرد (مانند آنچه در شکل ۳-۳۵ آمده است) و آنها را از روی تکه های منفرد تصویر تشخیص دهد. حال باید پرسید که کدام قاعده مندپهای آماری برای اینکار مناسب است؟ همانطور که شکل ۳-۳۵ القا می کند، توزیعهای تصاویری که از لحاظ مفهومی مشابهند، الگوهای سازگاری را با درجه ای از تنوع در یک بُعد نشان می دهد و همچنین همبستگی (مستقیم یا معکوس) زیادی را بین برخی از ابعاد ویژگیها به نمایش می گذارد.

اگرچه افکنش این الگوها روی دو بُعد، همانطور که در مثالها دیده می شود، به نظر ماهرانه و زیرکانه می رسد، توزیع کامل چند متغیره با صدها بُعد نیز امضاهای آماری برجسته ای تولید می کند که توسط سیستم بینایی قابل استخراج است. برای اینکه مشخص شود که این مدل چگونه تعمیم می یابد، نخست باید مشخص کرد که چگونه بازنمایی توزیع نواحی محلی تصویر را انجام می دهد. یک راه ساده برای خلاصه سازی الگوی همبستگی برای نوع تصویر داده شده ماتریس کواریانس داده هاست. یک کد نورونی برای این ساختار می تواند با شماره گذاری مجموعه کواریانسهای مشاهده شده و انتساب یک نورون به هر یک از آنها ساخته شود ولی هاین راه حل دارای دو ایراد است: نخست اینکه کلاس تصاویر محلی از پیش معلوم نیست، و دوم اینکه با در نظر گرفتن تعداد محدود نورون در سیستم بینایی، بازنمایی تمام انواع تصاویر امکان پذیر نیست حتی اگر تنها ترکیبات ممکن مرزهای تصاویر را هم در نظر بگیریم. به جای این راه حل یک کد پراکنده در نظر گرفته شده است که در آن فعالیت سطح بندی شده جمعیت نورونها برای توصیف زنجیره الگوهای بالقوه کواریانس استفاده می شوند.

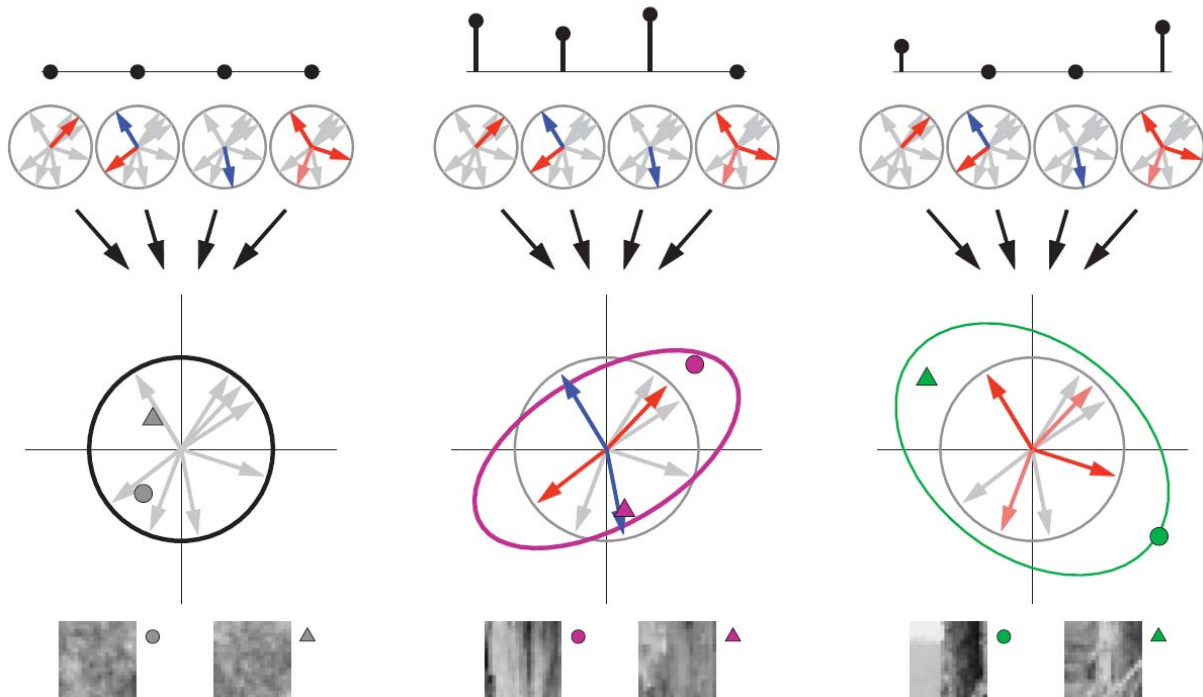
این مدل کدگذاری توزیع به صورت شماتیک در شکل ۳-۳۶ آورده شده است. این مدل همبستگی موجود در نواحی محلی تصویر را با یک توزیع گوسی چند متغیره دارای میانگین ثابت صفر و کواریانسی که تابع فعالیت نورنهاست بازنمایی می کند. این تعریف ساده آماری از سویی انعطاف پذیری در اخذ زنجیره توزیعهای تصاویر طبیعی و از سوی دیگر سادگی محاسباتی را برای تخمین راحت پارامترها به ارمان می آورد. این مدل از دو سری پارامتر برای توصیف همبستگی را توزیع تصاویر استفاده می کند.

سری اول بردارهای \vec{b}_k (پیکانههای درون دایره خاکستری) هستند که ویژگیهایی از تصویر را مشخص می کند که توزیع تصویر در راستای آنها می توان از توزیع استاندارد^{۲۲۴} (دایره سیاه رنگ) کشیده تر یا فشرده تر باشد. این بردارها بین تمام نورونهای این مدل (که با چهار دایره خاکستری که مجموعه پیکانههای یکسانی درون آنهاست نشان داده شده اند) مشترکند. از آنجا که این بردارها اجباراً با محورهای ابعاد ورودی همراستا نیستند، تغییر دگرگونی در راستای یک بردار ممکن است متناظر با تغییر همبسته متغیرهای همبسته در چندین بُعد باشد.

نورونهای مدل، z_j ؛ تغییرات در این جهتها را با وزنهای w_{jk} نشان می دهند: هر نورون مجموعه وزنهای متفاوتی دارند که متناظر با انقباض یا انبساط در راستای ویژگی \vec{b}_k است. وزن مثبت (قرمز رنگ) بدین معناست که نورونها به محرکهای بیشتری در آن راستا پاسخ می دهند، وزن منفی (آبی رنگ) به معنای بازه محدودتری از محرکهاست، و وزنهای نزدیک به صفر (خاکستری رنگ) نمایانگر خنثی بودن نورون نسبت به آن راستاست. فعالیت ترکیب شده همه نورونها شکل

²²⁴ canonical distribution

توزیع کدگذاری شده را نشان می دهد (بیضی های توزیع گوسی). با داشتن یک نقطه تمرکز -یک تصویر ورودی- مدل یک بازنمایی نورونی (همان توزیع تصویر) را محاسبه می کند که محتملترین تشریح را از تصویر ورودی بدست بدهد. این مدل می تواند روی نواحی مختلف تصویر تعمیم یابد اگر بازنمایی استنتاج شده در طول یک ناحیه مشابه باشد (مثلاً برای جفت تکه های تصویر شکل ۳-۳۵).



شکل ۳-۳۶: مدل کدگذاری توزیع. بجای کدگذاری مقادیر دقیق پیکسل تصویر ورودی (پایین)، مدل های پیشنهادی برای هر تصویر محتملترین توزیع درون آن (بیضی) را استنتاج می کند. الگوهای فعالسازی برای نورونهای مدل در بالای هر ستون نشان داده شده است. نبودن فعالیت به معنای فقدان ساختار تصویری مورد نظر نورون است (ستون چپ) که توزیع کانونی است و آمارگان تمام تصاویر طبیعی را منعکس می کند (دایره سیاه). افزایش فعالیت نورونی به معنای انحراف از این توزیع کانونی است و الگوهای آماری را در نواحی محلی تصویر نشان می دهد (ستونهای وسط و راست، نمادها و تکه های تصویری مشابه شکل ۳-۳۵) در هر ستون الگوی فعالسازی برای هر دو ورودی یکسان است.

با تطبیق پارامترهای w_{jk} و \bar{b}_k مدل به داده، می توان کاراترین راه را برای استفاده از نورونهای محدود برای توصیف طیف وسیعی از توزیعهای مشاهده شده در تصاویر طبیعی بدست آورد. باید توجه داشت که با وجود اینکه هدف این روش بدست آوردن یک بازنمایی پایدار از تمام تکه های یک ناحیه تصویر است، هیچ فرضی در مورد محلّی بودن صورت نگرفته است (کدگذاری برای هر تکه تصویر مستقلاً صورت گرفته است) و این وظیفه مدل است که یک بازنمایی فشرده را از تمام تکه ها یادگیری کند و خودبخود کشف کند که کدام بازنمایی خواص آماری از یک نوع را به اشتراک می گذارد.

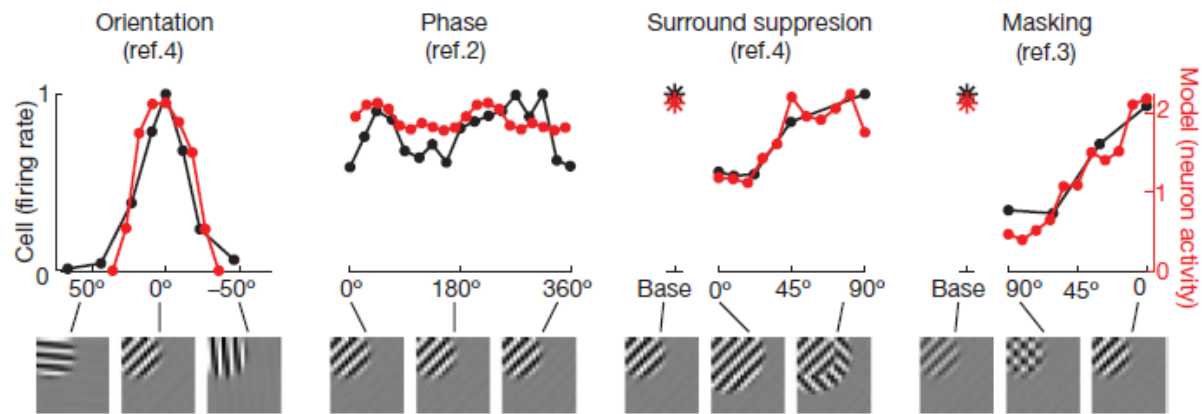
اگر طبق فرضیه ها نورونهای بخش بینایی قشر خاکستری مخ الگوهای همبستگیهای نواحی محلی را کدگذاری کند و خصوصاً با آمارگان صحنه های طبیعی سازگار شده باشد، انتظار می رود که بازنمایی یادگیری شده توسط مدل، خواص

نورونهای بینایی را منعکس کند. به همین منظور در مقاله (Karklin 2008) مدل روی مجموعه بزرگی از تصاویر طبیعی آموزش داده شده است و پارامترهای بدست آمده و نیز خصوصیات پاسخ نورونهای مدل به تصاویر طبیعی بررسی شده است.

بردارهای \vec{b}_k جهت‌های انبساط و انقباض متداول در شکل توزیع تصویر را کدگذاری می‌کنند. همانند تکه‌های تصویر، هر یک از آنها مانند یک ویژگی لبه مانند جهتدار و محلی هستند. مجموعه کامل آنها پهنای یک تکه تصویر را فرش می‌کند و تمام بازه راستاها و فرکانسهای مکانی تصاویر طبیعی را می‌پیماید و این ویژگیهای تصویر راستادار میان گذر با تصاویر بهینه برای تحریک سلولهای ساده بخش تصویری اولیه قشر خاکستری مخ سازگارند (Jones 87, van Hateren 98). چنانکه اشاره شد قبلاً بازنمایی‌های مشابهی با استفاده از مدلهای آمارگان خطی که کارایی کدگذاری تصویر را بهینه می‌کردند بدست آمده بود (Olshausen 96, Bell 97). در مدلی که در اینجا به آن اشاره شد، باوجود اینکه ویژگیها صریحاً برای بازسازی تصویر استفاده نمی‌شود ولی در عوض به بهبود توزیع کدشده کمک می‌کند (پیکانها در شکل ۳-۳۶) بنابراین در حالیکه تفسیرهای سنتی از کدگذاری اولیه حسگری اینگونه است که باید بازسازی مناسبی از محرک را بدست دهد، نتیجه این روش تفسیری به آن اضافه می‌کند: این کدگذاری باید تنوع در توزیعهای تصویر را نیز منتقل کند و به قسمتهای بینایی که در مسیر داده‌ها پس از دیگری قرار گرفته اند امکان ارائه بازنمایی مجردتری را می‌دهد.

دومین مجموعه پارامترها، وزنهای w_{jk} هستند که نقش هر نورون را در شکل دادن به توزیع تصویر کدشده مشخص می‌کنند. به جز پاسخ دادن به تعداد معدودی از ویژگیهای تصویری برانگیخته‌کننده و یا فرونشاندنده، تعداد زیادی از نورونها برای تشکیل یک الگوی تنوع که در توزیع مخصوص تصویر وجود دارد شرکت می‌کنند. با وجود این اهمیت کارکردی این زیرواحدها اصلاح ساختار آماری توزیع کدگذاری شده است، این نورونها همچنین ویژگیهای محرکی که این نورون بیش از همه به آن حساس است را منعکس می‌کند. باید توجه داشت که یک نورون مدل، به جای آنکه با فعال شدن خود از حضور بهترین محرک خود خبر بدهد، توسط تمام تصاویر این توزیع فعال می‌شود ولی بسته به محرکی که در فضای تصویر موجود است، احتمال کمی وجود دارد که از فعالیت نورون جلوگیری شود.

برای مقایسه رفتار نورون مدل با سلول نورونی که در قشر بینایی مخ قرار دارد، باید پاسخ آنرا با محرکی که در آزمایشات کلاسیک فیزیولوژی استفاده می‌شود، پنجره‌های سینوسی، آزمو، پارامترهای مدل پس از آموزش با تصاویر طبیعی ثابت می‌شوند و پاسخ نورونی روی مجموعه‌ای از الگوهای که مرکز آنها در محدوده بینایی است و پاسخ بیشینه را بدست داده اند محاسبه می‌شود. این نورون خاص مدل خواص متنوعی از خود نشان می‌دهد که در نورونهای پیچیده V1 سلولهای V2 وجود دارند؛ خواص نظیر نامتغیری به فاز، تنظیم با راستا و ویژگیهای فرونشانی پیچیده (شکل ۳-۳۷). زیرمجموعه بزرگی از این جمعیت خواص مشابهی را نشان می‌دهد، در حالیکه برخی دیگر از آنها الگوهای پیچیده تر (مانند الگوهای که در لایه‌های بالاتر بینایی مانند V2 و V4 مشاهده شده اند) را کدگذاری می‌کنند. باید تاکید کرد که ویژگیهای تصویری که در نورونها شکل گرفته است بدون هیچ پیش فرضی درباره ساختار تصویری که توسط نورونها کدگذاری شده اند و نیز بدون تطبیق دادن یک مدل بر داده‌های آزمایشات فیزیولوژی بدست آمده اند. خصوصاً اینکه این مدل ویژگیهای تصویری کدگذاری شده را محدود به محلی بودن و راستادار بودن نمی‌کند و از پیش نیز چگونگی ترکیب زیرواحدها در الگوهای که هر نورون نشان می‌دهد را نیز تعیین نمی‌کند.

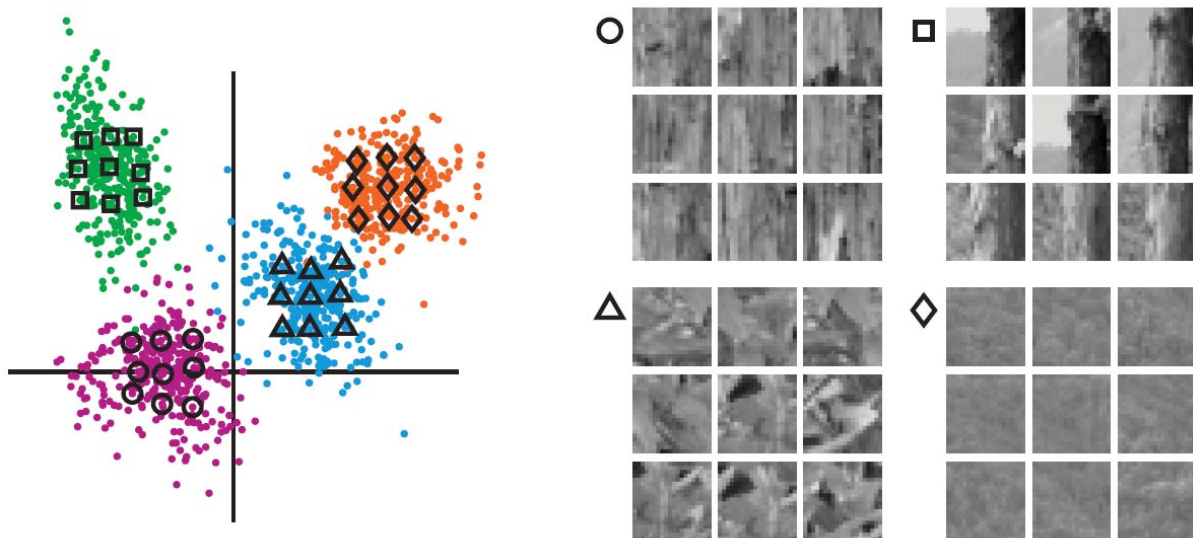


شکل ۳-۳۷: هنگامی که پنجره های سینوسی به نرون مدل داده شود، از خود وجوه مشترک زیادی با سلولهای پیچیده V1 نشان می دهد. این پاسخ با راستای پنجره بخوبی تنظیم می شود ولی نسبت به فاز آن نامتغیر است. افزودن پنجره به نواحی اطراف (سومین نمودار از سمت چپ، 0°) باعث فرونشانی نسبت به پاسخ مبنا به یک پنجره می شود (ستاره) ولی این فرونشانی با راستای محرک نواحی اطراف تنظیم می شود و هنگامی که زاویه آن عمود بر زاویه مرجع نرون باشد (90°)، پاسخ در ضعیفترین حالت خود قرار دارد. ماسک کردن با یک پنجره عمود که روی محرک اصلی قرار می گیرد باعث فرونشانی پاسخ می شود (نمودار راست، 90°) ولی این فرونشانی نیز وابسته به راستای می باشد. پاسخ تمام نرونها مدل در یک مقیاس ترسیم شده اند (محور قرمز)؛ نرخ آتش سلولها در هر نمودار برای داشتن بیشینه ۱ نرماسازی شده اند؛ و راستای مرجع برای نرون مدل و سلولهای درون نمودار به 0° شیفت داده شده اند.

حال به روشی که مدل از جمعیت نرونها استفاده می کند تا تصویر را بازنمایی کند پرداخته می شود. اگر مدل بتواند روی تنوع زیادی که در تصاویر طبیعی وجود دارد تعمیم یابد، آنگاه تکه های تصویری که در سراسر فضای اصلی ویژگیهای خطی پراکنده اند باید در فضای بازنمایی مدل دارای خوشه بندی محکمی باشند. این امر را می توان با ننگشت بازنمایی مدل از مجموعه تصاویر روی دو بُعد (همانند فضای تصویر در شکل ۳-۳۵) نشان داد که اینکار در شکل ۳-۳۸ انجام شده است. همانگونه که انتظار می رفت، با کدگذاری توزیع تصویر بجای کدگذاری ویژگیهای خاصی درون یک تصویر، نرونها مدل توانسته اند که تصاویری که از لحاظ ادراکی مشابهند را با بازنمایی مشابهی ارائه دهند و انواع مختلف تصویر را نیز تفکیک کنند.

یکی از محدودیتهای این چارچوب آماری اینست که الگوریتم صریح و سراسری برای کدگذاری نرونی بدست نمی دهد. با این وجود می توان استنتاج درون مدل را با یک محاسبه سر راست تخمین زد: یک نرون مربعات پاسخهای تعداد زیادی از ویژگیهای \bar{b}_k تصویر را جمع می کند و الگو را با وزنها w_{jk} همبسته می کند. این محاسبات را می توان به چشم یک تعمیم از مدل استاندارد سلولهای پیچیده انگاشت که در آن هر سلول پیچیده مربع خروجی دو سلول ساده را به عنوان ورودی می پذیرد (Movshon 78, Adelson 85, van Hateren 98, Heeger 92). از سوی دیگر نرونها مدل می توانند ورودیهای متعددی بگیرند و ویژگیهای خطی خودبخود یاد گرفته شوند. آزمایشات نشان داده است که تعداد بهینه ویژگیهای ورودی برای تصاویر مختلف متفاوت است و این امر در مورد راههای تجمیع ویژگیها نیز صادق است. این پیش بینیها با تحلیلهای زیرناحیه های کارکردی سلولهای پیچیده V1 نیز همخوانی دارد (Chen 2007, Rust 2005). به علاوه برخی از نرونها مدل الگوهای مکانی پیچیده تری را تولید می کردند که پاسخ نرونی به تنوع غنی تری از تصاویری که واقعاً

توسط متخصصان فیزیولوژی آزمایش شده بود را پیش بینی می کرد. آزمایشهایی که چنین محرکهایی را در بر داشته باشند می توانند درستی این مدل را رد یا قبول کنند.



شکل ۳-۳۸: تعمیم روی تنوع تصاویر طبیعی. (چپ) در مقایسه با افکنشهای خطی (شکل ۳-۳۳ بالا-راست) یک افکنش دوبعدی از بازنمایی مدل (فعالیت ۱۵۰ نورون مدل) خوشه‌هایی را نشان می‌دهد که بخوبی از یکدیگر جدا شده‌اند. (راست) هر گروه 3×3 از تصاویر با گروهی از نمادها که در تصویر چپ قرار داده شده‌اند متناظرند. علیرغم تنوع ظاهر لبه‌ها و بافتها، بازنمایی مدل از تصاویر در هر ناحیه تعمیم یافته است و در عین حال تمایز بین آنها را حفظ کرده است.

پدیده‌های غیرخطی‌ای که توسط نورونهای مدل بروز می‌کند (شکل ۳-۳۷) قبلاً نیز در مدل‌های سلولهای پیچیده گنجانده شده بودند (Cavanagh 2002, Carandini 97, Heeger 92, Heeger 96). اکثر این مقالات به تطبیق مدل‌های ریاضی روی داده‌های نورونی اختصاص یافته‌اند (Cavanagh 2002, Carandini 97, Heeger 92, Cadieu 2007) و بنابراین تفسیری کارکردی از این خواص مشاهده شده به دست نمی‌دهند. سایر مدل‌ها نیز چنانکه اشاره شد، به دنبال اهداف محاسباتی خاصی مانند استقلال آماری (Schwartz 2001, Hyvarinen 2001)، پایداری بازنمایی در طول زمان (Berkes 2005, Hurri 2003)، و یا نامتغیری به ازای موقعیت و مقیاس (Riesenhuber 99) بوده‌اند. با اینحال این مدل‌ها نیز صراحتاً به مسئله تعمیم نپرداخته‌اند که در مقاله (Karklin 2008) توسط استنتاج توزیع آماری که محتمل‌ترین توصیف تصویر را بدست می‌دهد، این هدف تامین شده است. یک مزیت مهم این روش اینست که بجای فرض کردن نامتغیری (و یا حساسیت) به تعداد اندکی از پارامترهای محرک مانند موقعیت و مقیاس، مدل مجموعه‌ای به مراتب عمومی‌تر از ویژگیها را یاد می‌گیرد که توسط ساختارهای آماری تصاویر طبیعی بدست می‌آیند. اگر نورونهای لایه‌های بالاتر بینایی بر اساس این آمارگان تعمیم یافته باشند، باید در مقابل برخی از ابعاد ورودی نامتغیر باشند و پاسخ آنها به تصاویر طبیعی باید ساختار آماری مشترکی را در نواحی محلی منعکس کند. بنابراین این مدل یک راه کمی برای تفسیر پاسخ نورونها به محرکهای پیشرفته است که به قاعده مندی آماری آنها باز می‌گردد.

۳-۲- مروری بر روشهای آشکارسازی چهره

۳-۲-۱- مقدمه

تحقیقات پر شتاب در زمینه پردازش چهره بر مبنای این فرض است که اطلاعات گرانبهای مانند هویت کاربر، وضعیت و منظور او از تصاویر قابل استخراج است و کامپیوتر می تواند در برابر آن (مثلاً دیدن حالت چهره فرد) عکس العمل مقتضی را نشان دهد. در سالهای اخیر چهره و شناسایی حالات چهره نسبت به آنچه سالیان قبل از آن توسط فیزیک-روانشناسان، عصب شناسان و مهندسان درباره آن صورت گرفته بود پیشرفت بیشتری داشته است. بسیاری از نمایشهای تحقیقات و کاربردهای تجاری بر پایه همین تحقیقات تهیه شده است. اولین گام هر سیستم پردازش چهره پیدا کردن مکانی است که چهره در تصویر قرار دارد. بهر حال تشخیص چهره از یک تصویر چالشهای بزرگی را در بر دارد که به دلیل تنوع در مقیاس، موقعیت، راستا (چپ-راست، دوران داده شده)، و وضعیت (تمام رخ، نیم رخ) رخ می دهد. حالات چهره، انسداد، و شرایط نورپردازی ظاهر کلی چهره را تغییر می دهد.

یکی از تعاریف آشکارسازی چهره اینگونه است: با دادن یک تصویر دلخواه، هدف تشخیص چهره تعیین اینست که آیا هیچ چهره ای در تصویر وجود دارد یا خیر، و اگر وجود دارد مکان و گستردگی هر چهره را بازگرداند (Yang 2002). چالشهای مرتبط با تشخیص چهره می تواند به دسته های زیر تقسیم شود:

وضعیت: تصاویر چهره ها بدلیل وضعیت نسبی چهره-دوربین (روبرو، ۴۵ درجه، نیمرخ، بالا به پایین) تفاوت می کند، و برخی از ویژگیهای چهره مانند یک چشم یا بینی ممکن است قسمتی یا تماماً پوشانده شوند.

حضور یا عدم حضور مؤلفه های ساختاری: ویژگیهای چهره مانند ریش، سبیل، و عینک ممکن است وجود داشته باشند و در بین همین مؤلفه ها نیز تنوع زیادی به دلیل شکل، اندازه، و رنگ وجود دارد.

حالات چهره: ظاهر چهره ها تحت تاثیر مستقیم حالات احساسی چهره است.

انسداد^{۲۲۵}: بخشی از چهره ممکن است توسط اشیا دیگر پوشانده شده باشد. در یک تصویر از گروهی از افراد، برخی از چهره ها ممکن است توسط چهره ها دیگر مسدود شده باشد.

راستای تصویر: تصاویر چهره کاملاً با دوران دوربین حول محور نوری دوربین تغییر می کند.

شرایط تصویربرداری: هنگامی که تصویر تشکیل می شود، عواملی مانند نور (طیف، توزیع منبع و شدت آن) و شاخص های دوربین (پاسخ سنسور، عدسی ها) می تواند روی ظاهر یک چهره موثر باشد.

مسائل متعددی موجود است که قرابت زیادی با آشکارسازی چهره دارد. مکانیابی چهره^{۲۲۶} سعی می کند تا وضعیت مکانی یک چهره را تعیین کند. این مسئله در واقع ساده سازی شده مسئله تشخیص چهره با فرض اینست که تصویر ورودی دربردارنده تنها یک چهره است. هدف مسئله تشخیص ویژگیهای چهره^{۲۲۷}، همان تشخیص وجود و مکان ویژگیها مانند چشمها، بینی، حفره های بینی، ابرو، دهان، لبها، گوشها، و غیره با فرض وجود تنها یک چهره در تصویر است. شناسایی/تعیین هویت چهره^{۲۲۸} تصویر ورودی (پروپ) را در یک پایگاه داده ها (گالری) جستجو می کند و تطابقیهای احتمالی را گزارش می کند. هدف تصدیق هویت چهره^{۲۲۹}، تصدیق هویت ادعا شده فردی است که در تصویر حضور دارد. روشهای ردگیری چهره^{۲۳۰} دائماً مکان و احتمالاً راستای یک چهره را در دنباله تصاویر به صورت بلادرنگ تخمین می زند. تشخیص حالت چهره^{۲۳۱} به شناسایی حالتیهای درونی انسان(مثلاً ناراحت، خوشحال، منزجر،...) می پردازد.

آشکار است که آشکارسازی چهره اولین گام برای هر سیستم خودکاری است که قصد دارند کارهای فوق را انجام دهند. شایان ذکر است که در بسیاری از مقالات از عبارت آشکارسازی چهره استفاده شده است در حالی که در مطالعه عملی آنها تنها یک چهره در تصویر موجود است. این بخش بین دو مقوله آشکارسازی چهره و مکانیابی چهره تفاوت قائل شده است چرا که دومی مسئله ساده شده اولی است. در ضمن تمرکز روی روشهای آشکارسازی چهره خواهد بود و نه ردگیری آن.

از میان روشهای آشکارسازی چهره، آنهایی که بر مبنای الگوریتمهای یادگیری هستند اخیراً توجه زیادی را به خود جلب کرده اند و نتایج بسیار خوبی را عرضه داشته اند.

یک مسئله مربوط و مهم اینست که کارایی یک روش چگونه ارزیابی می شود. معیارهای رایج در ادبیات این زمینه نرخ تشخیص و نرخ اعلامهای نادرست است. معیارهای سنجش دیگری نیز برای ارزیابی الگوریتم تهیه شده است که از جمله آنها می توان به زمان یادگیری، زمان اجرا، تعداد نمونه های مورد نیاز برای آموزش، و نسبت بین تشخیصهای درست و اعلامهای اشتباه است. وقتی که محققان از تعاریف متفاوتی برای تعریف نرخهای تشخیص درست و اعلامهای اشتباه استفاده کنند، ارزیابی مشکل و مشکلتر می شود. در این پژوهش نرخ تشخیص عبارتند از: نسبت بین تعداد چهره هایی که به درستی آشکارسازی شده اند و تعداد چهره هایی که توسط انسان تعیین شده است. یک ناحیه از تصویر که به عنوان چهره شناخته شده است، درست محسوب می شود اگر ناحیه تصویر بیش از کسر خاصی از چهره را در تصویر پوشش دهد. به طور کلی تشخیص دهنده ها دو دسته خطا دارند: اشتباهاً منفی^{۲۳۲} که در آن چهره ها از قلم می افتند و در نتیجه نرخ تشخیص تنزل پیدا می کند و اشتباهاً مثبت^{۲۳۳} که در آن ناحیه از تصویر چهره در نظر گرفته می شود در حالی که اینچنین نیست. یک ارزیابی عادلانه باید این دو عامل را به حساب آورد زیرا با تنظیم پارامترهای روشی می توان نرخ تشخیص را افزایش داد در حالی که تشخیص های اشتباه نیز افزایش می یابد.

²²⁶ face localization

²²⁷ facial feature detection

²²⁸ face recognition/identification

²²⁹ face authentication

²³⁰ face tracking

²³¹ facial expression recognition

²³² false negative

²³³ false positive

مقالات بسیاری درباره آشکارسازی چهره به رشته تحریر در آمده است و این مقوله سهم زیادی از بحث تشخیص اشیا را در بینایی ماشین تشکیل می دهد. تقریباً تمام روشهای مبتنی بر مدل و ظاهر برای تشخیص اشیا سه بعدی، به تشخیص اشیای صلب محدود است در حالیکه تلاش می کند شناسایی را در طیف وسیعی از موقعیتهای دوربین و شرایط نورپردازی به صورت موثری انجام دهد.

تشخیص چهره می تواند به صورت یک مسئله شناسایی دو کلاسه دیده شود که در آن یک ناحیه تصویر به یکی از دو دسته «چهره» و «غیر چهره» تخصیص می یابد. طبعاً، آشکارسازی چهره یکی از معدود تلاشهایی است که سعی می کند از خود تصویر و نه بازنمایی تجریدی آن یک کلاس از اشیا را تشخیص دهد که خود دارای تنوع بین کلاسی بالایی است (که قبلاً اشاره شد). همچنین یکی از معدود کلاسهایی از اشیا است که در آن تنوع توسط مجموعه داده های آموزشی بزرگ پوشش داده شده است و بنابراین برخی از روشهای تشخیص قابل اعمال به طیف گسترده تری از مسائل شناسایی هستند. آشکارسازی چهره همچنین چالشهای جالبی به مسئله شناسایی الگو و تکنیکهای یادگیری ذیل خود می افزاید. هنگامی که یک تصویر خام یا فیلتر شده به عنوان ورودی به دسته بندی کننده اعمال می شوند، ابعاد فضای ویژگیها بسیار بزرگ است (یعنی همان تعداد پیکسلها در داده های آموزشی نرمال شده). کلاسهای چهره و غیر چهره با توابع توزیع چندقله ای بیان می شوند و مرزهای تصمیم گیری احتمالاً غیرخطی هستند. برای کارایی، هر دسته بندی کننده باید بتواند از تعداد محدودی داده آموزشی یاد بگیرد یا وقتی با این داده های آموزشی ابعاد بالا کار می کند کارایی خود را حفظ کند.

در ادامه این بخش، مروری بر روشهای موجود تشخیص چهره از روی یک تصویر ثابت رنگی یا خاکستری فام انجام می گیرد. روشهای یاد شده را می توان در چهار دسته قرار داد. برخی از روشها نیز کاملاً از روی مرزهای این دسته ها عبور می کند که به آنها اشاره خواهد شد.

۱. **روشهای مبتنی بر دانش^{۲۳۴}**: این روشها دانش انسانی را درباره ای اینکه چه چیز یک چهره را تشکیل می دهد، درون یک پایگاه قوانین کدگذاری می کند. عموماً این قوانین روابط بین ویژگیهای چهره را مد نظر قرار می دهد. هدف عمده این دسته مکانیابی چهره است.
۲. **روشهای نامتغیر با ویژگی^{۲۳۵}**: هدف این الگوریتمها پیدا کردن ویژگیهای ساختاری ایست که حتی با وجود تغییرات وضعیت، دیدگاه، و شرایط نورپردازی نیز ثابت بمانند و از آنها برای مکانیابی صورت استفاده شود. این روشها نیز عمدتاً برای مکانیابی چهره استفاده می شوند.
۳. **روشهای تطبیق کلیشه^{۲۳۶}**: چندین الگوی استاندارد چهره برای توصیف کل چهره یا ویژگیهای آن به طور مجزا ذخیره شده است. همبستگی بین تصویر ورودی و الگوهای ذخیره شده محاسبه می شود و مورد استفاده روال آشکارسازی قرار می گیرد. این روشها هم برای مکانیابی و هم برای تشخیص چهره استفاده می شود.

²³⁴ knowledge-based methods

²³⁵ feature invariant approaches

²³⁶ template matching methods

۴. روشهای مبتنی بر ظاهر^{۲۳۷}: در مقابل روشهای تطبیق کلیشه، مدلها(یا کلیشه ها) از روی مجموعه تصاویر ورودی یاد گرفته می شود که باید تنوع نمایشهای گوناگون ظاهر چهره را اخذ کند. این مدلها یادگیری شده عموماً برای آشکارسازی چهره استفاده می شود.

در ادامه به انگیزه و خط مشی کلی هر دسته پرداخته می شود و بدنبال آن یک مرور اجمالی از روشهای خاص آن و مزایا و معایب آنها ارائه خواهد شد.

۳-۲-۲- روشهای مبتنی بر دانش

در این گرایش، روشهای آشکارسازی چهره بر اساس قوانینی انجام می پذیرد که از دانش پژوهنده از چهره افراد نشأت می گیرد. توصیف ویژگیهای چهره و روابط آنها با قوانین ساده کار آسانی است. به عنوان مثال یک چهره معمولاً با دو چشم متقارن، یک بینی، و یک دهان در تصاویر ظاهر می شود. ارتباط بین این ویژگیها می تواند با فاصله و موقعیت نسبی آنها بیان شود. نخست ویژگیهای چهره در تصویر باید استخراج شود، سپس چهره های کاندید بر اساس قوانین وضع شده انتخاب می شوند. یک فرایند تصدیق^{۲۳۸} نیز معمولاً برای کاهش تشخیص های نادرست بکار گرفته می شود.

یکی از مشکلات این روش، دشواری ترجمه دانش انسانی به قوانین خوش تعریف است. از طرفی اگر قوانین سخت گیرانه و دارای جزئیات باشند ممکن است منجر به حذف برخی از کاندیدهای چهره بودن شوند که نتوانسته اند تمام شرایط را برآورده سازند. از سوی دیگر اگر قوانین خیلی هم عمومی باشند ممکن است منجر به پاسخهای اشتباهاً مثبت زیادی شوند.

علاوه بر این گسترش این گرایش برای آشکارسازی چهره در وضعیتهای مختلف دشوار است چرا که بشمار آوردن تک تک حالات ممکن چالشی بزرگ است. لیکن باید در نظر داشت که الگوریتمهای اکتشافی^{۲۳۹} درباره چهره برای تشخیص چهره تمام رخ^{۲۴۰} در پس زمینه های ساده (غیر درهم ریخته^{۲۴۱}) بخوبی عمل می کند.

یکی از روشهای این گرایش از یک روش سلسله مراتبی مبتنی بر دانش برای آشکارسازی چهره استفاده می کند. این سیستم از سه لایه از قوانین استفاده می کند. در بالاترین سطح، تمام کاندیدهای چهره بودن با اسکن تصویر توسط یک پنجره و اعمال یک دسته از قوانین به هر مکان پیدا می شوند. قوانین در سطحهای بالاتر توصیفات کلی و اجمالی از این هستند که یک چهره چه شکلی است و هر چه سطح پایینتر می رود بر جزئیات ویژگیهای چهره بیشتر تاکید می کند. همانطور که در مثال شکل ۳-۳۹ نشان داده شده است، یک سلسله مراتب از وضوحهای مختلف (هرم دقت) با میانگین گیری و نمونه برداری تولید می شود.

²³⁷ appearance-based methods

²³⁸ verification

²³⁹ heuristics

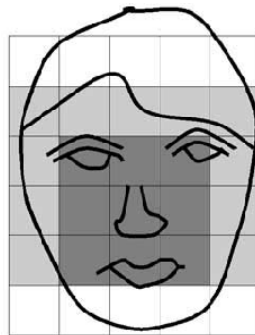
²⁴⁰ frontal face

²⁴¹ uncluttered



شکل ۳-۳۹: (چپ) $n = 1$ تصویر اصلی (دوم) $n = 4$ (سوم) $n = 8$ (راست) $n = 16$. تصویر اصلی و تصاویر متناظر آن با دقت پایینتر. هر سلول مربعی شامل $n \times n$ پیکسل است که شدت آن برابر میانگین شدت روشنایی پیکسلهای آن سلول است.

مثالهایی از قوانینی که در دقتهای پایینتر بکار گرفته می شوند عبارتند از: (الف) بخش مرکزی صورت (که در تصویر ۳-۳۹ تیره تر سایه خورده است) دارای چهار سلول با شدت روشنایی یکنواخت است. (ب) بخش گرد بالای صورت (که در تصویر ۳-۴۰ روشنتر سایه خورده است) دارای شدت روشنایی پایه ای یکنواختی است. (ج) تفاوت آشکاری بین سطوح خاکستری متوسط بخش میانی و بخش گرد بالایی وجود دارد.



شکل ۳-۴۰: یک چهره معمول استفاده شده در روشهای بالا به پایین مبتنی بر دانش: قوانین درباره ویژگیهای نواحی چهره بر مبنای دانش انسانی (مثلاً توزیع شدت روشنایی و تفاوتها) وضع می شوند.

جستجو بدنبال کاندیدهای چهره بودن در تصویر دارای کمترین دقت (سطح یک) صورت می گیرد و آنها برای پردازش بیشتر در دقتهای بهتر استفاده می شوند. در سطح دوم، یکنواخت سازی هیستوگرام محلی^{۲۴۲} روی کاندیدهای چهره بودن که به سطح دوم آمده اند صورت می گیرد و سپس لبه یابی روی آن انجام می شود. نواحی کاندیدهای باقیمانده در سطح سوم با مجموعه قوانین دیگری که به ویژگیهای چهره مانند چشمها و دهان پاسخ می دهند آزموده می شوند. یک ویژگی جالب این روش استفاده از راهبرد درشت-به-ظریف^{۲۴۳} یا کانون توجه^{۲۴۴} برای کاهش حجم محاسبات است. با وجود اینکه این روش نرخ تشخیص بالایی را بدست نمی دهد (۵۰٪ صورت از ۶۰٪ صورت با ۲۸ اعلام اشتباهاً مثبت)، ایده استفاده از سلسله مراتب چند دقتی و قوانین راهنمای جستجو در برخی کارهای بعدی هم استفاده شده است. یکی از این مقالات از یک

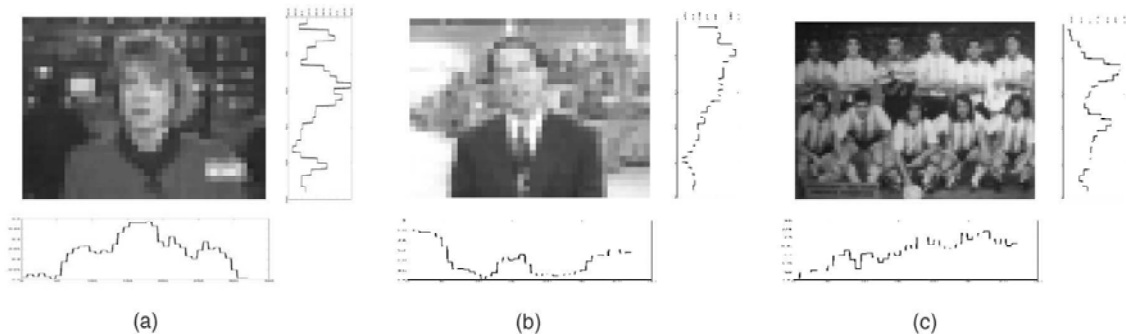
²⁴² local histogram equalization

²⁴³ coarse-to-fine

²⁴⁴ focus-of-attention

مکانیابی برپایه قوانین استفاده می کند که در آن ویژگیهای چهره توسط یک روش افکنش که بطور موفقی مرز چهره را پیدا می کند، تعیین مکان می شوند.

اگر $I(x,y)$ مقادیر شدت روشنایی یک تصویر $m \times n$ در نقطه (x,y) باشد، نماهای افقی و عمودی آن به صورت $VI(x) = \sum_{y=1}^n I(x,y)$ و $HI(x) = \sum_{y=1}^n I(x,y)$ تعریف می شوند. ابتدا نمای افقی تصویر ورودی تهیه می شود و سپس دو کمینه محلی که توسط تغییرات ناگهانی در شدت روشنایی HI را به عنوان سمت چپ و سمت راست سر در نظر گرفته می شود. به طور مشابه نمای عمودی تصویر نیز تهیه می شود و کمینه های محلی به عنوان مکان لبها، نوک بینی، و چشمها در نظر گرفته می شود. این ویژگیهای کشف شده یک کاندیدا را تشکیل می دهند. شکل ۳-۴۰ چپ یک مثال را نشان می دهند که در آن مرزهای چهره متناظر با کمینه ای است که در آن تغییرات ناگهانی شدت روشنایی رخ می دهد. متعاقب آن قوانین تشخیص ابرو/چشم، بینی/سوراخ بینی، و دهان برای تصدیق این کاندیداها بکار گرفته می شود.



شکل ۳-۴۱: (چپ) و (وسط) $n = 8$ (راست) $n = 4$. نماهای افقی و عمودی. تشخیص یک چهره با جستجو بدنبال دره های نمودار نما ممکن است. ولی این روش در یافتن چهره در پس زمینه پیچیده و یا چند چهره با مشکل مواجه می شود.

شکل ۳-۴۱ وسط نمایانگر مثالی است که تشخیص چهره در پس زمینه پیچیده آن توسط نماهای افقی و عمودی دشوار است. به علاوه این روش نمی تواند چند چهره را در یک تصویر مانند ۳-۴۱ راست تشخیص بدهد. اساساً روش افکنش زمانی کاراست که قاب پنجره که روی آن کار می شود در جای مناسبی قرار گیرد تا دچار اشتباه نشود.

۳-۲-۳- روشهای مبتنی بر ویژگی

در مقابل روش بالا به پایین مبتنی بر دانش، محققان سعی در یافتن ویژگیهای نامتغیر چهره کرده اند. فرض نهفته در این روش بر این مشاهده تکیه دارد که انسان بدون دشواری چهره ها و اشیا را در وضعیتهای گوناگون مکانی و نوری تشخیص می دهد و بنابراین خصوصیات یا ویژگیهایی وجود دارد که در برابر این دگرگونیها نامتغیر باشد. روشهای متعددی پیشنهاد شده اند که ابتدا ویژگیهای چهره را تشخیص می دهند و سپس از روی آن حضور چهره را نتیجه می گیرند. ویژگیهای چهره مانند ابروها، چشمها، بینی، دهان، و خط رویش مو اغلب توسط تشخیص دهنده های خط استخراج می شوند. بر پایه این ویژگیهای استخراج شده، یک مدل آماری ساخته می شود تا ارتباط آنها را توصیف و وجود چهره را تصدیق کند.

یک مشکل این الگوریتمهای وابسته به ویژگی اینست که ویژگیهای تصویر بشدت تحت تاثیر نورپردازی، نویز، و انسداد قرار می گیرند. محدوده ویژگیها می تواند برای چهره ها ضعیف شود در حالی که سایه ها می توانند منجر به لبه های قوی ای شوند که با وجود این مسائل الگوریتمهای دسته بندی مفهومی بی استفاده خواهند بود.

۳-۲-۱- ویژگی های چهره

یک روش مکانیابی برای تقطیع یک چهره از پس زمینه درهم ریخته برای شناسایی چهره به صورت زیر است: این روش از یک نقشه لبه ها (لبه یاب کنی) و الگوریتم اکتشافی برای حذف و گروه بندی لبه ها استفاده می کند تا در نهایت تنها لبه های روی کانتور چهره باقی بماند. سپس یک بیضی روی مرز سر و پس زمینه برازش می شود. این روش به دقت ۸۰ درصد دست یافته است.

در مقاله ای دیگر بجای استفاده از لبه ها، از روش ساده برای تشخیص صورت به کمک حبابها و رگه ها^{۲۴۵} (دنباله ای از لبه های همسو) استفاده شده است. این مدل چهره از دو حباب تیره و سه حباب روشن برای بازنمایی چشمها، گونه ها، و بینی استفاده می کند. مدل همچنین از رگه ها برای نمایش پیرامون چهره، ابروها، و لبها استفاده می کند. دو تنظیم مثلثی برای کد کردن رابطه مکانی بین حبابها استفاده شده است. یک تصویر لاپلاسین دقت پایین برای تسهیل یافتن حبابها تولید می شود. سپس تصویر برای یافتن موجودیتهای مثلثی پویش شده تا به عنوان کاندید در نظر گرفته شود. اگر یک رگه حول یک کاندید تشخیص داده شود وجود چهره اعلام می شود.

در مقاله دیگری یک راه برای مکانیابی ویژگی های چهره و چهره ها در تصاویر خاکستری فام تهیه شده است. پس از فیلتر باند گذر، عملگرهای مورفولوژی برای بسط نواحی دارای شدت روشنایی بالا و اشکال معین استفاده می شوند (مثلاً چشمها). هیستوگرام تصویر پردازش شده معمولاً یک قله برجسته را نشان می دهد. بر اساس این قله و پهنای آن، مقادیر آستانه تطبیقی انتخاب می شوند تا دو تصویر دوسطحی (باینری) بسازند. در هر دوی این تصاویر مؤلفه های همبند تشخیص داده می شوند تا محلهای کاندید برای ویژگیهای چهره را نشان دهند. پس از آن ترکیب این محلها توسط یک دسته بندی کننده ارزیابی می شود تا تعیین کند که آیا چهره ای موجود هست یا خیر. در این مقاله نحوه اعمال عملگرهای مورفولوژی و نیز نحوه ترکیب ویژگیهای کاندید چهره برای یافتن یک چهره روشن نشده است.

در پژوهش دیگری از روش احتمالاتی برای یافتن یک صورت در صحنه دارای درهم ریختگی استفاده شده است. این روش بر پایه تشخیص دهنده های محلی ویژگی و تطابق تصادفی گراف^{۲۴۶} است. انگیزه این مقاله، فرموله کردن مسئله مکانیابی چهره به صورت یک مسئله جستجو است که در آن هدف پیدا کردن ترتیبی از ویژگیهای خاص چهره است که بیشترین احتمال قرار داشتن در یک الگوی چهره را دارند (یعنی ترکیب ساخته شده از آن ویژگیها به احتمال زیاد یک چهره خواهد بود). پنج ویژگی (دو چشم، دو حفره بینی، و تقاطع بینی/لب) برای توصیف یک چهره نوعی استفاده شده است. برای هر جفت از ویژگیهای چهره از یک نوع (مثلاً جفت چشم چپ و چشم راست) فاصله نسبی محاسبه می شود و مدل آن روی مجموع تصاویر به صورت توزیع گوسی بدست می آید. یک کلیشه چهره با میانگین گیری از پاسخ فیلترهای مشتق گوسی در

²⁴⁵ blobs and streaks

²⁴⁶ random graph matching

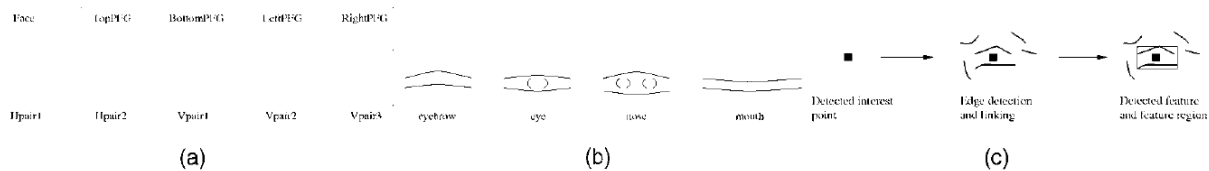
راستاها و مقیاسهای گوناگون (در پیکسلهای درون ویژگی چهره) برای تعدادی از چهره های موجود در پایگاه داده تعرف می شود. با دادن یک تصویر آزمایشی، ویژگیهای کاندید چهره توسط تصابق پاسخ فیلتر در هر پیکسل با بردار کلیشه شناسایی می شود (که ذاتاً مشابه عمل همبستگی است). دو کاندیدی برتر ویژگیها که قویترین پاسخ را دارند برای جستجو بدنبال بقیه ویژگیهای چهره انتخاب می شوند. از آنجایی که ویژگیهای چهره نمی توانند در هر موقعیت دلخواه قرار بگیرند، موقعیت مورد انتظار ویژگی دیگر توسط مدل آماری فاصله متقابل تخمین زده می شود. افزون بر این کوواریانس تخمینها قابل محاسبه است. بنابراین موقعیت مورد انتظار ویژگی ها با احتمال بالایی می تواند تخمین زده شود. سپس مجموعه های ستاره ای^{۲۴۷} میتواند از کاندیدهایی که در مکانهای مناسب قرار دارند تشکیل شده و مجموعه ستاره ایی که بیشترین شباهت به چهره را دارد تعیین شود. پیدا کردن بهترین مجموعه ستاره ای به صورت یک مسئله تطابق تصادفی گراف فرموله شده است که در آن گره های گراف متناظر با ویژگیهای چهره و یالهای آن متناظر با فاصله بین ویژگیهای مختلف است. رتبه یک مجموعه ستاره ای بر اساس یک تابع توزیع احتمال صورت می گیرد که احتمال مربوط بودن مجموعه ستاره ای به چهره را در برابر احتمال تولید ترکیب از مکانیسم -غیر چهره- دیگری نشان می دهد. در این روش زمانی ادعا می شود که چهره به درستی تشخیص داده شده که سه ویژگی یا بیشتر را روی صورت بیابد و می تواند به دقت مکانیابی ۸۶ درصد دست یابد.

بجای استفاده از فاصله متقابل برای توصیف رابطه بین ویژگیهای چهره در مجموعه ستاره ای، می توان از مدل دیگری برای چهره نیز استفاده کرد. در یک مقاله، بازنمایی و رتبه بندی مجموعه های ستاره ای به مدد نظریه احتمالات شکل صورت گرفته است. آمار شکل، یک تابع توزیع احتمال توأم روی n نقطه ویژگی است که برای ویژگی نام با (X_i, Y_i) و با فرض اینکه نقاط ویژگی اصلی روی صفحه ای با توزیع $2n$ بعدی توزیع نرمال است، نشان داده می شود. این مقاله از همان روش پیشینه درستنمایی برای تعیین کاندید چهره استفاده کرده است. یک مزیت این روشها اینست که چهره های که بخشی از آنها پوشانده شده است را می تواند تشخیص دهد. با اینحال روشن نیست که آیا این روشها می توانند طوری تغییر کنند که چند چهره را به طور کارایی در یک صحنه تشخیص دهند.

در یک پژوهش دیگر روشی مبتنی بر ویژگی ارائه شده است که از حجم زیادی از شواهد بدست آمده از تصویر و شواهد متن آنها استفاده می کند. در نخستین گام یک فیلتر مشتق دوم گوسی به نسبت سه بر یک بر تصویر خام اعمال می شود. نقاط مهم در پیشینه های محلی پاسخ فیلتر تشخیص داده می شود که نمایانگر مکانهای احتمالی ویژگیهای چهره است. در گام دوم لبه در اطراف نقاط مهم آزموده می شود و نقاط را به نواحی جداگانه ای گروه بندی می کند. این گروه بندی ادراکی لبه ها بر پایه مجاورت و شباهت آنها از لحاظ راستا و قدرت است. اندازه گیری خواص ناحیه از قبیل طول لبه، قدرت لبه، و واریانس شدت روشنایی محاسبه شده و در بردار ویژگی ذخیره می شود. میانگین و ماتریس کوواریانس هر بردار ویژگی چهره از روی داده های آموزشی ویژگیهای چهره محاسبه می شود. یک ناحیه تصویر می تواند یک کاندید قابل قبول ویژگی چهره باشد اگر فاصله مالهالانوبیس بین بردار ویژگیهای متناظر آن از آستانه ای کمتر باشد. ویژگی های برجسته خورده براساس دانش مدل -که می داند که آنها نسبت به هم در کجاها باید دیده شوند- باهم گروه بندی می شوند. هر ویژگی چهره و گروه سپس با استفاده از شبکه بیزی ارزیابی می شود. یک وجه جالب این روش اینست که این روش می تواند چهره ها را در راستاها و مکانهای تشخیص دهد کارایی آن برای تصاویر چهره هایی با ابعاد، راستاها و دیدگاههای مختلف به ۸۵ درصد

247 constellation

هم رسیده است. با این وجود نرخ تشخیص اشتباهاً مثبت ۲۸ درصدی آن و پیاده سازی آن برای چهره های بزرگتر از 60×60 پیکسل از ضعفهای این روش است. بعدها این روش با کمک مدل‌های کان‌تور پویا بهبود یافت. شکل ۳-۴۲ شمایی از این روش تشخیص چهره مبتنی بر ویژگی را نشان می دهد.



شکل ۳-۴۲: (چپ) مدل چهره به صورت یک صفحه با شش ویژگی چهره همراستا (ابروها، چشمها، بینی و دهان) (وسط) هر ویژگی چهره به صورت یک جفت لبه همراستا مدل می شود. (راست) فرایند انتخاب ویژگی با نقاط مهم آغاز شده، که در پی آن تشخیص لبه و الحاق صورت می گیرد و با یک مدل آماری آزمایش می شود.

در یک پژوهش الهام گرفته از بیولوژی، آشکارسازی چهره با توجه به مدل استخراج ویژگی شبکیه و حرکات نوسانی ریز چشم صورت پذیرفته است. الگوریتم ارائه شده روی نقشه آشکار سازی^{۲۴۸} ناحیه مورد علاقه^{۲۴۹} کار می کند که در آن لاتیس شبکیه از روی سلولهای عقده ای درشت سلول در سیستم بینایی انسان مدل می شود. در فاز اول یک پوشش دانه درشت از تصویر بر اساس پاسخ فیلتر میدان گیرندگی مکان چهره را تخمین می زند. هر ناحیه دریافت از تعدادی نورون تشکیل شده است که با فیلترهای گوسی در راستاهای مختلف پیاده سازی شده است. در فاز دوم نقشه آشکارسازی با پوشش جزئی تر تصویر پالایش می شود تا چهره پیدا شود.

نویسندگان مقاله ای دیگر روشی براساس مورفولوژی ایجاد کرده اند تا قطعه هایی را که شبیه چشم می نامند برای تشخیص چهره استخراج کنند. آنها معتقدند که چشمها و ابروها مهمترین و پایدارترین ویژگیهای چهره انسان هستند و به همین دلیل برای تشخیص گزینه خوبی هستند. آنها قطعات شبیه چشم را به صورت لبه های روی پیرامون چشمها تعریف کرده اند. ابتدا عملیات مورفولوژی مانند بستن^{۲۵۰}، تفاضل مقطوع^{۲۵۱}، و آستانه گیری^{۲۵۲} اعمال می شود تا پیکسلهایی که در آن شدت روشنایی تغییر زیادی می کند استخراج شوند. به این پیکسلها در این روش پیکسلهای شبیه چشم اطلاق می شود. سپس یک فرایند برچسب گذاری برای تولید قطعات شبیه چشم استفاده می شود. این قطعات برای راهنمایی جستجو به سمت نواحی بالقوه چهره استفاده می شوند که این نواحی دارای ترکیب موقعیت هندسی چشمها، بینی، ابروها و دهان است. نواحی کاندید چهره سپس با یک شبکه عصبی تصدیق می شوند.

²⁴⁸ conspicuity map
²⁴⁹ region of interest
²⁵⁰ closing
²⁵¹ clipped difference
²⁵² thresholding

در یکی دیگر از مقالات روشی برای تشخیص شکل معرفی شده و برای آشکارسازی چهره های تمام رخ در تصاویر ثابت استفاده شده است. تشخیص از دو گام تشکیل شده است: تمرکز و دسته بندی بر اساس شدت روشنایی. تمرکز کردن بر اساس چینش فضایی تکه های لبه است که از تفاضل شدت روشنایی استخراج می شود. یک دسته مهم از این چینشها به ازای تبدیلات گونگون نوری و هندسی تعریف شده است. از یک مجموعه آموزشی از تصاویر چهره، متداولترین چینش فضایی لبه های چهره در برابر پس زمینه به کمک یادگیری مفهومی انتخاب می شود. بعدها الگوریتم CART برای ساخت یک درخت دسته تصمیم از روی تصاویر آموزشی و مجموعه ای از داده هایی از پس زمینه هایی کلی که اشتباهاً مثبت ارزیابی می شوند استفاده شد. با داشتن یک تصویر آزمایش، ناحیه های مورد نظر توسط چینش فضایی شکستگی لبه ها شناسایی می شوند. هر ناحیه سپس از روی درخت CART ساخته شده، به یکی از دو کلاس چهره و یا پس زمینه دسته بندی می شوند.

۳-۲-۲-۳-۲-۳ بافت

چهره انسان دارای بافت متمایزی است که می تواند آنها را از سایر اشیا جدا کند. این مقاله روشی را ارائه کرده است که وجود چهره را از روی بافتهای شبیه چهره (SGLD) استنتاج می کند. این بافت به کمک ویژگیهای آماری مرتبه دوم روی زیر تصویرهای 16×16 پیکسل محاسبه می شود. سه نوع ویژگی در این روش محسوب می شود: پوست، مو، و سایر بافتهای. این روش از شبکه های عصبی همبستگی سری^{۲۵۳} برای دسته بندی با ناظر و شبکه های خود سازمانده و ویژگی کوهونن^{۲۵۴} برای تشکیل خوشه های کلاسهای مختلف بافت استفاده می کند. برای استنتاج از برچسب بافتهای، روش رأی گیری از وجود بافتهای مو و صورت استفاده شده است.

مقاله دیگری نیز از مدل SGLD برای آشکارسازی چهره استفاده کرده است. اطلاعات رنگ نیز در مدل بافت چهره دخیل بوده است. با استفاده از مدل بافت چهره یک روش پویس تصاویر رنگی بدنبال چهره پیشنهاد شده است که در بخشهای نارنجی رنگ از جمله نواحی صورتها تقویت می شوند. یک مزیت این روش اینست که می تواند چهره هایی که رو به بالا نیستند یا دارای ویژگیهایی مانند ریش و عینک هستند را تشخیص دهد.

۳-۳-۲-۳-۳-۳ رنگ پوست

رنگ پوست انسان ویژگی ایست که استفاده شده و اثبات شده است که در کاربردهای زیادی از تشخیص چهره تا ردیابی دست کاراست. با وجود اینکه انسانهای مختلف رنگ پوست متفاوتی دارند، مطالعات زیادی نشان می دهد که این تفاوت عمدتاً در شدت روشنایی آنهاست و نه در خلوص رنگی آنها^{۲۵۵}. چند فضای رنگی مختلف از قبیل RGB، RGB، نرمال شده، HSV یا HSI، YCrCb، YIQ، YIQ، YES، CIE XYZ، و CIE LUV برای برچسب گذاری پیکلهای پوست استفاده شده است. روشهای زیادی پیشنهاد شده اند تا یک مدل رنگ پوست را بسازند. ساده ترین روش، تعریف یک ناحیه از پیکلهای همفام پوست با استفاده از مقادیر Cr, Cb است. یعنی $R(Cr, Cb)$ از نمونه های پیکلهای رنگ پوست. با انتخاب

²⁵³ cascade correlation neural networks

²⁵⁴ Kohonen's self organizing feature map

²⁵⁵ chrominance

مناسب آستانه ها (Cr_1, Cr_2) و (Cb_1, Cb_2) ، یک پیکسل دارای تون پوست دسته بندی می شود اگر مقادیر (Cr, Cb) آن در این بازه ها باشد.

در مقاله دیگری از هیستوگرام (r, g) یعنی $h(r, g)$ در فضای رنگی RGB استفاده شده است تا احتمال بدست آوردن یک بردار RGB خاص که پیکسل از پوست نشان می دهد، بدست آید. به عبارت دیگر یک پیکسل عضو رنگ پوست دسته بندی می شود اگر از آستانه ای بالاتر باشد که این آستانه بصورت تجربی از هیستوگرام مثالها بدست می آید.

مقاله ای دیگر روش تکرار شونده شناسایی پوست را معرفی کرده است که از تقاطع هیستوگرامها در فضای رنگی HSV بدست می آید. یک تکه اولیه از پیکسلهای رنگ پوست که به دانه های کنترل معروف شده است، توسط کاربر معرفی شده، به عنوان مقدار اولیه الگوریتم تکرار شونده استفاده می شود. برای تشخیص نواحی رنگ پوست، این روش روی تصویر حرکت می کند، یک تکه در هر تصویر، و هیستوگرام آنرا با هیستوگرام کنترلی مقایسه می کند. تقاطع هیستوگرامها برای این مقایسه استفاده می شود و اگر امتیاز تطابق یا تعداد نمونه های مشترک (تقاطع) از آستانه ای بیشتر باشد، این تکه به عنوان رنگ پوست دسته بندی می شود. مقاله دیگری یک گزاره رنگی در فضای رنگی HSV را برای جدا کردن نواحی پوست از پس زمینه تعریف می کند.

در مقابل روشهای غیر پارامتری که در بالا گفته شد، توابع توزیع گوسی و یا یک تابع توزیع توأم گوسی معمولاً برای مدل کردن رنگ پوست استفاده می شود. پارامترها در توزیع گوسی تک قله ای معمولاً توسط الگوریتم بیشینه درستنمایی تخمین زده می شوند. انگیزه استفاده از توزیع توأم گوسی از این مشاهده بر می آید که هیستوگرام رنگ پوست افراد با قومیتهای گوناگون یک تابع تک قله گوسی را نمی سازند، بلکه بیشتر چند قله ای هستند. پارامترهای توزیع توأم گوسی معمولاً توسط یک الگوریتم بیشینه کردن امید ریاضی تخمین زده می شود.

در یکی از تحقیقات گسترده، که در آن نزدیک یک میلیارد پیکسل تون رنگ پوست برچسب گذاری شده جمع آوری شده است (در فضای رنگی نرمال شده RGB)، با مقایسه کارایی روشهای هیستوگرام و توزیع توأم برای تشخیص پوست، برتری مدلهای هیستوگرام از لحاظ دقت و هزینه محاسباتی نشان داده شده است.

اطلاعات رنگ یک ابزار کارا برای شناسایی نواحی چهره و ویژگیهای خاص چهره است به شرط آنکه مدل رنگ پوست به طرز مناسبی برای شرایط مختلف نوری محیط سازگاری داده شده باشد. بهرحال این مدلهای رنگ پوست در جایی که طیف رنگی منبع نوری تغییر فاحشی کند، کارا نیست. به عبارت دیگر، شمایل رنگی اغلب به دلیل تغییرات پس زمینه و حتی پیش زمینه ناپایدار است. بنابراین مسئله ثبات رنگ²⁵⁶ از دیدگاه مدلهای مبتنی بر فیزیک فرموله شده است و چند روش برای استفاده از رنگ پوست تحت شرایط مختلف نورپردازی پیشنهاد شده است. یکی از این روشها مدل ترکیبی تطبیقی برای رنگ ارائه داده است که تحت شرایط نورپردازی متفاوت به ردیابی چهره می پردازد. بجای استفاده از مدل رنگ بر پایه ثبات رنگ، از یک مدل تصادفی برای تخمین برخط توزیع رنگ شیء استفاده می کند و آنرا برای سازگاری با دیدگاهها و نورپردازی های مختلف تطبیق می دهد. نتایج اولیه نشان می دهد که این سیستم در شرایط متفاوت نورپردازی می تواند چهره را ردگیری کند. بهرحال این روش نمی تواند برای پیدا کردن چهره در تنها یک تصویر ثابت استفاده شود.

²⁵⁶ color constancy

معمولاً رنگ پوست به تنهایی برای تشخیص و یا ردگیری چهره کافی نیست. لذا سیستمهای مدولاری بوجود آمده اند که از تلفیق تحلیل شکل، تقطیع رنگ، و اطلاعات حرکت برای یافتن و ردگیری سر و چهره در دنباله تصاویر استفاده می شوند.

۳-۲-۳-۴- چند ویژگی

اخیراً بسیاری از روشها چند ویژگی را با هم ترکیب می کنند تا چهره ها را تشخیص دهند یا مکانیابی کنند. بسیاری از آنها از ویژگیهای سراسری نظیر رنگ پوست، اندازه و شکل برای پیدا کردن کاندیدهای صورت بودن و تصدیق این کاندیدها با کمک ویژگیهای محلی و دارای جزئیات مانند ابروها، بینی، و مو استفاده می کنند. یک خط مشی معمول با پیدا کردن نواحی شبیه پوست همانگونه که پیشتر اشاره شد شروع می شود. سپس پیکسلهای شبیه پوست با تحلیل مؤلفه های همبند و یا الگوریتمهای خوشه بندی با یکدیگر گروه بندی می شوند. اگر شکل یک مؤلفه همبند بیضوی یا تخم مرغی باشد، این ناحیه یک کاندید محسوب می شود. نهایتاً از ویژگیهای محلی برای تصدیق استفاده می شود.

برخی از روشهای دیگر از مجموعه های متفاوتی از ویژگیها استفاده می کنند. از جمله این روشها، مقاله ای است که از نظریه فازی برای یافتن چهره ها در تصاویر رنگی استفاده می کند. این روش از دو مدل فازی برای توصیف توزیع رنگ پوست و مو در فضای رنگی CIE XYZ بهره می برد. پنج مدل شکل سر (یکی از روبرو و چهارتا از کنار سر) برای توصیف اجمالی ظاهر چهره در تصاویر استفاده شده است. هر مدل شکل یک الگوی دوبعدی شامل $m \times n$ سلول مربعی است که هر سلول می تواند حاوی چندین پیکسل باشد. هر سلول دارای دو خصوصیت است: نسبت پوست و نسبت مو، که نشانگر مساحت پوست (یا مساحت مو) در سلول است. در تصویر آزمایشی، هر پیکسل به عنوان مو، چهره، مو/چهره، و مو/پس زمینه بر اساس مدل توزیع دسته بندی می شود و ناحیه هایی شبیه پوست و شبیه مو تولید می کند. سپس مدل شکل سر با نواحی شبیه مو و شبیه چهره استخراج شده از تصویر تست مقایسه می شود. اگر شبیه بودند ناحیه کشف شده یک کاندید تلقی می شود. برای تصدیق، ویژگی چشم-ابرو و بینی-دهان با استفاده از لبه های افقی از کاندید استخراج می شود.

در روشی دیگر برای مکانیابی چهره و استخراج ویژگیهای صورت از شکل و رنگ استفاده می شود. ابتدا تقطیع رنگ در فضای HSV صورت می گیرد تا ناحیه های شبیه چهره مکانیابی شود. سپس به روش رشد دادن ناحیه در دقت دانه درشت، مؤلفه های همبند معین می شود. برای هر مؤلفه همبند، بهترین بیضی برازش شده به کمک گشتاورهای هندسی محاسبه می شود. مؤلفه های همبندی که با بیضی خوب تقریب زده شوند به عنوان کاندید استفاده می شوند. به دنبال این مراحل، این کاندیدها با جستجو به دنبال ویژگیها درون مؤلفه همبند تصدیق می شوند. ویژگیهایی مثل چشمها و دهان بر اساس این مشاهده که غالباً از سایر نواحی چهره تیره تر هستند استخراج می شوند.

در مقاله دیگری یک مدل گوسی رنگ چهره برای دسته بندی پیکسلهای رنگ چهره استفاده می شود. برای اینکه ویژگیهای شکل خوشه در تصویر دوسطحی مشخص شود، یک مجموعه از ۱۱ گشتاور هندسی مرتبه پایین با استفاده از

تبدیل فوریه و ملین شعاعی^{۲۵۷} استفاده می شود. برای تشخیص، یک شبکه عصبی با گشتاورهای هندسی استخراج شده آموزش داده می شود.

تقارن الگوهای چهره نیز می تواند برای مکانیابی چهره استفاده شود. دسته بندی چهره/غیر چهره با استفاده از تابع توزیع شرطی کلاسی در فضای YES صورت می شود و پس از آن برای بدست آوردن نواحی متصل هموارسازی می شوند. سپس یک کلیشه بیضی چهره برای تعیین شباهت نواحی رنگ پوست بر اساس فاصله هاوسدورف^{۲۵۸} استفاده می شود. در نهایت نقاط میانی چشم بر اساس چند تابع هزینه که از تقارن ذاتی چشم و صورت بهره می روند، پیدا می شوند. نوک بینی و مرکز دهان نیز پس از آن با بهره گیری از مرکز چشمها پیدا می شوند. یک ضعف این روش اینست که این روش تنها برای حالتی از چهره که از روبرو و در حالتی که دو چشم دیده شود به درد می خورد. روش دیگری نیز ابداع شده است که از رنگ و تقارن محلی بهره می جوید.

در مقابل روشهای مبتنی بر پیکسل یک روش مبتنی بر ساختار، رنگ، و هندسه تولید شده است. ابتدا تقطیع چند مقیاسه برای استخراج نواحی همگن در یک تصویر استفاده می شود. با استفاده از یک مدل رنگ پوست گوسی، نواحی دارای تون پوست استخراج می شود و بیضی وار گروه بندی می شود. یک چهره در صورتی تشخیص داده می شود اگر ویژگیهای چهره مانند دهان و چشمها در این بیضی وجود داشته باشد. آزمایشات نشان می دهد که این روش قابلیت تشخیص چهره در راستاهای مختلف و حتی در حضور ویژگیهایی مانند ریش و عینک را نیز دارد.

در روش دیگری یک ارائه حبابی برای استخراج توصیفی فشرده و دارای ساختار معنادار از تصاویر چندطیفی ماهواره ای پیشنهاد شده است. در هر پیکسل یک بردار ویژگی تشکیل می شود که حاصل الحاق مختصات تصویر پیکسل به مؤلفه های طیفی (یا متنی) پیکسل می باشد. سپس پیکسلها با استفاده از این بردار ویژگی خوشه بندی می شوند تا مؤلفه های چسبیده و همبند، یا حبابها را تشکیل دهند. برای تشخیص چهره ها، هر بردار ویژگی از مختصات تصویر و خلوص رنگی نرمال شده که رابطه آن به صورت زیر است، استفاده می کند. سپس یک الگوریتم اتصال برای رشد حبابها استفاده می شود و حباب پوست حاصله که اندازه و شکل آن به یک چهره معیار نزدیکتر است به عنوان چهره تشخیص داده می شود.

$$X = \left(x, y, \frac{r}{r+g+b}, \frac{g}{r+g+b} \right) \quad (19-3)$$

فاصله و رنگ نیز در روشی دیگر برای تشخیص چهره استفاده می شود. نقشه عدم توافق^{۲۵۹} محاسبه می شود و اشیا به کمک یک هیستوگرام عدم توافق از پس زمینه تقطیع می شود. برای اینکار فرض می شود که پیکسلهای پس زمینه دارای عمق یکسان هستند و از پیکسلهای اشیای پیش زمینه بسیار بیشترند. با استفاده از توزیع گوسی در فضای رنگی RGB نرمال شده، نواحی تقطیع شده با رنگ شبیه چهره به عنوان چهره طبقه بندی می شوند.

²⁵⁷ radial Mellin transform

²⁵⁸ Hausdorff distance

²⁵⁹ disparity map

۳-۲-۴- روشهای مبتنی بر تطبیق کلیشه

در تطبیق کلیشه، یک الگوی استاندارد چهره (معمولاً روبرو) از قبل بطور دستی تعریف می شود یا با یک تابع پارامتری می شود. با داشتن یک تصویر ورودی، مقادیر همبستگی با الگوی استاندارد برای پیرامون چهره، چشمها، بینی، و دهان بطور جداگانه محاسبه می شود. وجود یک چهره بر مبنای مقادیر همبستگی اعلام می شود. این روش دارای مزیت سادگی پیاده سازی است. بهرحال ثابت شده است که این روشها برای تشخیص چهره کافی نیستند زیرا نمی توانند با تغییرهایی در مقیاس، وضعیت، و شکل بطور کارایی برخورد کنند. برای کسب نامتغیر بودن با مقیاس و شکل، زیرکلیشه های دارای ابعاد و دقتهای مختلف و نیز کلیشه های انعطاف پذیر پیشنهاد شده اند.

۳-۲-۴-۱- کلیشه های از پیش تعریف شده

یکی از نخستین روشهای آشکارسازی چهره در تصاویر مختص چهره تمام رخ بوده است. در آن چند زیرکلیشه برای چشمها، بینی، و پیرامون چهره استفاده می شود تا چهره را مدل کند. هر زیر کلیشه با تکه خطهایی مشخص می شود. خطوط در تصویر ورودی بر مبنای بیشترین تغییرات گرادیان تعیین شده و سپس با زیر کلیشه تطبیق داده می شود. همبستگی بین زیرتصاویر و کلیشه های پیرامون ابتدا کاندیدهای مکان چهره را تعیین می کند. سپس در این مکانهای کاندید تطبیق دادن با سایر زیرکلیشه ها صورت می گیرد. به عبارت دیگر، فاز نخست تمرکز و توجه یا ناحیه مورد علاقه را مشخص می کند و فاز دوم جزئیات را برای تعیین حضور چهره می سنجد. ایده تمرکز توجه و زیر کلیشه ها بعدها نیز در کارهای مرتبط با آشکارسازی چهره نیز استفاده شده است.

در مقاله دیگری روشی برای مکانیابی بر اساس کلیشه چهره تمام رخ (شکل پیرامون چهره) ارائه شده است. ابتدا یک فیلتر سوبل برای استخراج لبه ها استفاده می شود. این لبه ها برای جستجوی کلیشه چهره بر مبنای چند قید گروه بندی می شوند. پس از آنکه کانتور سر پیدا شد، همین فرایند با ابعاد مختلف برای پیدا کردن ویژگیهایی نظیر چشمها، ابروها، و لبها تکرار می شود. بعدها مقاله دیگری یک روش مکانیابی با استفاده از ۴۰ کلیشه برای جستجوی ویژگیهای چهره و یک استراتژی کنترلی برای هدایت و ارزیابی تشخیص دهنده های ویژگی مبتنی بر کلیشه پیشنهاد کرد.

مقاله ای دیگر روشی دو مرحله ای برای تشخیص چهره ارائه کرده است که در آن فرضیه هایی در مورد وجود چهره بوجود می آید و آزموده می شود. یک مدل چهره به صورت ویژگیها و با لبه ها تعریف می شود. این ویژگیها انحنای سمت چپ چهره تمام رخ، سمت راست آن و خط رویش مو را توصیف می کند. تشخیص دهنده لبه مار-هیلدرت^{۲۶۰} برای تهیه نقشه لبه های تصویر ورودی استفاده می شود. سپس یک فیلتر برای حذف اشیایی که کانتور آنها احتمالاً جزو چهره نیست استفاده می شود. زوج تکه کانتورهای بر اساس نزدیکی و راستای نسبی بهم متصل می شوند. گوشه ها برای تکه سازی کانتور به ویژگیهای منحنی پیدا می شود. این ویژگیهای منحنی سپس با بررسی ویژگیهای هندسی و مقعیت نسبی در همسایگی خود برچسب گذاری می شوند. اگر مشخصات زوج ویژگیهای منحنی سازگار باشد (یعنی هر دو بتوانند متعلق به یک چهره باشند) با هم ادغام می شوند. نسبت زوج ویژگیهایی که یک لبه را تشکیل می دهند با نسبت طلایی مقایسه می شود و یک هزینه به

²⁶⁰ Marr-Hildert

لبه تخصیص داده می شود. اگر هزینه گروه سه تایی از ویژگیهای منحنی (با برجسبهای متفاوت) پایین بود، گروه به یک فرضیه تبدیل می شود هنگامی که چهره افراد در مقالات روزنامه تشخیص داده می شود، اطلاعات موازی که تعداد افراد حاضر در تصویر را نشان می دهد، از عنوان تصویر ورودی بدست می آید تا بهترین فرضیه انتخاب شود. با این حال تصاویر در این روش باید رو به بالا، تمام رخ، و بدون انسداد دید باشند. روش فوق در مقاله ای دیگر با استخراج لبه ها در دامنه موجک توسعه داده شده است.

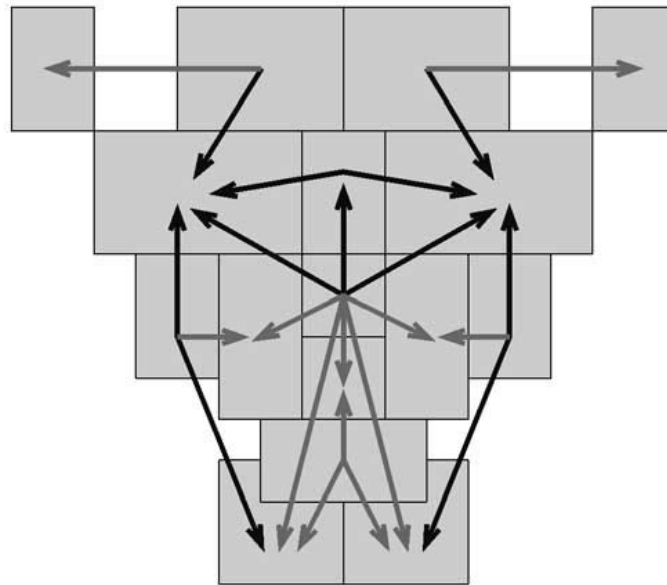
مقاله دیگری مدلی کیفی برای الگوی چهره (QMF) ارائه کرد. در QMF، هر تصویر نمونه به تعدادی بلوک تقسیم می شود، و برای هر بلوک ویژگیهای کیفی تخمین زده می شود. برای پارامتری کردن یک الگوی چهره، روشنی و لبه بودن به عنوان ویژگیهای مدل تعریف می شوند. پس از آن، این کلیشه بلوکی برای محاسبه چهره بودن در هر موقعیت هر تصویر ورودی محاسبه می شود. اگر معیار چهره بودن از یک آستانه بالاتر باشد، یک چهره تشخیص داده می شود.

شمایل^{۲۶۱}ها نیز گاهی به عنوان کلیشه ای برای مکانیابی استفاده می شوند. مجموعه ای از شمایلهای چهره به کمک تحلیل مؤلفه اصلی از نمونه های چهره استخراج می شود که در آن چهره بوسیله آرایه ای از بیتها نمایش داده می شود. این شمایلهای ویژه سپس با یک تبدیل هاف تعمیم یافته برای مکانیابی استفاده می شوند.

یک روش مکانیابی بر مبنای چند کلیشه برای مؤلفه های چهره نیز پیشنهاد شده است. این روش شامل تعداد زیادی فرضیه برای ظاهرهای احتمالی ویژگیهای چهره است. سپس یک مجموعه از فرضیه ها برای وجود چهره بر اساس فرضیه های ویژگیهای چهره به روش دمپستر-شفر تعریف می شود. با داشتن یک تصویر، تشخیص دهنده های ویژگی فاکتور اطمینان از وجود ویژگیهای چهره را حساب می کنند. فاکتورهای اطمینان با هم تلفیق می شوند تا میزان باور به وجود چهره در تصویر را مشخص کنند.

در مقاله دیگر، از مجموعه کوچکی از نامتغیرهای مکانی تصویر برای توصیف فضای الگوهای چهره بهره گرفته شده است. دیدگاهی که برای طراحی این نامتغیرها استفاده شده است اینست که با وجود اینکه تغییرات نورپردازی روشنایی نقاط مختلف چهره (مانند چشمها، گونه ها، و پیشانی) را جداگانه تغییر می دهد، روشنایی نسبی این بخشها تا حد زیادی بدون تغییر می ماند. تعیین نسبتهای روشنایی جفتی تعدادی از این نواحی و حفظ کردن راستای این نسبتها (یک ناحیه از دیگری روشنتر و یا تیره تر است) یک نامتغیری پایدار را بدست می دهد. بنابراین قاعده های روشنایی مشاهده شده به صورت یک کلیشه کد می شود که یک کلیشه فضایی دانه درشت از چهره به همراه چند زیر ناحیه مناسب است که متناظر با ویژگیهای کلیدی چهره مانند چشمها، گونه ها، و پیشانی است. قیود روشنایی بین قسمت‌های چهره با یک مجموعه مناسب از رابطه های تیره تر- روشنتر بین زیرنواحی ذخیره می شوند. یک چهره در صورتی در تصویر دیده می شود که تمام قیود تیره تر- روشنتر را ارضا کند. ایده استفاده از تفاوت شدت روشنایی بین نواحی همسایه محلی بعدها برای بازنمایی موجک برای پیاده ها، اتومبیلها و آشکارسازی چهره ها استفاده شد. این روش برای مکانیابی چهره در یک سیستم بینایی پویای روبات استفاده شده است. شکل ۳-۴۳ کلیشه بهبود یافته با ۲۳ رابطه را نشان می دهد. این رابطه های تعریف شده به ۱۱ رابطه اساسی (فلشهای تخت) و ۱۲ رابطه تایید کننده (فلشهای خط چین) تقسیم می شوند. هر پیکان در این شکل یک رابطه را نشان می دهد، که

نوک پیکان ناحیه دوم (مقسوم الیه کسر) را نشان می دهد. گفته می شود که رابطه برای کلیشه برقرار است اگر رابطه بین دو ناحیه از آستانه ای بیشتر باشد و یک چهره مکانیابی می شود اگر تعداد رابطه های اساسی و تاییدی از حدی بالاتر باشد.



شکل ۳-۴۳: یک کلیشه 16×15 برای مکانیابی تصویر. کلیشه از ۱۶ ناحیه (مستطیل‌های خاکستری) و ۲۳ رابطه (پیکانها) تشکیل شده است.

یک روش تطابق کلیشه سلسله مراتبی برای تشخیص چهره در یک مقاله پیشنهاد داده شده است. در گام اول یک تصویر ورودی از 20° تا 20° با قدمهای 5° چرخانده می شود تا چهره های چرخانده شده را پوشش دهد. یک سلسله مراتب دقت تصویر تشکیل می شود (ر.ک. شکل ۳-۳۸) و لبه ها با استفاده از عملگر لاپلاسی استخراج می شوند. کلیشه چهره از لبه هایی که توسط شش مؤلفه چهره تولید شده، تشکیل شده است: دو ابرو، دو چشم، یک بینی و یک دهان. در پایان الگوریتمهای ذهنی برای تعیین وجود چهره بکار گرفته شده است. نتایج آزمایشات نشان می دهد که این روش در تصاویری که حاوی تنها یک چهره (تمام رخ یا چرخیده) هستند بهتر از تصاویر حاوی چند چهره نتیجه می دهد.

۳-۲-۴-۲-۳ کلیشه های انعطاف پذیر^{۲۶۲}

اولین مقاله ای که از این روش استفاده کرده است، از کلیشه های انعطاف پذیر برای مدل کردن ویژگیهای چهره که با یک کلیشه اولیه ارتجاعی^{۲۶۳} مدل می شوند (مانند چشمها) استفاده کرده است. در این روش، ویژگیهای چهره با یک کلیشه پارامتری توصیف می شوند. یک تابع انرژی برای تعریف لبه های متصل، قله ها و دره های تصویر که متناظر با پارامترهای کلیشه هستند، استفاده شد. بهترین تطبیق مدل ارتجاعی با کمینه کردن تابع انرژی پارامترها بدست می آید. با وجود اینکه

²⁶² deformable templates

²⁶³ apriori elastic model

آزمایشات آنها نتایج خوبی را برای ردیابی ویژگیهای غیرصلب نشان داد، یکی از معایب این روش اینست که کلیشه انعطاف پذیر باید در نزدیکی شیء مورد نظر مقدار اولیه بگیرد.

در مقاله دیگری یک مدل آشکارسازی بر اساس مارها^{۲۶۴} و کلیشه ها تهیه شده است. یک تصویر ابتدا با یک فیلتر تار کننده کانوالو می شود و سپس یک عملگر مورفولوژیک برای تقویت لبه ها بکار گرفته می شود. یک مار تغییر یافته n پیکسلی (کوچک است) برای یافتن و حذف قطعه منحنی های کوچک استفاده می شود. هر چهره با یک بیضی تخمین زده می شود و تبدیل هاف روی باقیمانده مارها برای پیدا کردن بیضی غالب استفاده می شود. بنابراین مجموعه چهار پارامتری که بیضیها را تعریف می کنند بدست می آید و به عنوان کاندید مکان چهره استفاده می شود. برای هر یک از این کاندیدها، یک روش مشابه روش کلیشه های انعطاف پذیر برای پیدا کردن ویژگیهای جزئی تر استفاده می شود. اگر تعداد قابل توجهی از ویژگیهای چهره پیدا شوند و اگر نسبت آنها آزمون نسبت را بر اساس کلیشه چهره ارضا کند، یک چهره پیدا شده است. مقاله دیگری نیز از مارها برای مکانیابی مرز سر با کمک روشی حریصانه برای کمینه کردن تابع انرژی استفاده کرده است.

در تیپ دیگری از مقالات، یک مقاله چهره را با شکل و اطلاعات شدت روشنایی توصیف کرده است. این روش با مجموعه ای از تصاویر آموزشی آغاز به کار می کند که در آنها کانتورهای نمونه مانند مرز چشم، بینی، و گونه/چانه بطور دستی برچسب خورده اند و یک بردار از نقاط نمونه برای معرفی شکل بکار برده شده اند. آنها از یک مدل توزیع نقطه ای^{۲۶۵} برای شاخص کردن بردار در بین مجموعه ای از آنها استفاده کرده اند و از روشی مانند آنچه که برای توصیف ظاهر شکل بر اساس شدت روشنایی نرمال شده گفته شد بهره جسته اند. یک PDM شبیه چهره می تواند برای مکانیابی چهره ها در تصویر جدید با کمک جستجوی مدل فعال شکل^{۲۶۶} برای تخمین مکان شیء و پارامترهای شکل استفاده شود. بسته چهره سپس به شکل میانگین تغییر شکل می دهد و پارامترهای شکل و شدن نور استخراج می شود. این پارامترهای شکل و شدت نور می توانند برای دسته بندی استفاده شوند.

در مقاله ای دیگر از روشی مشابه برای مکانیابی چهره در تصویر استفاده می شود. ابتدا نواحی مستطیلی حاوی ویژگیهای مورد نظر روی شکل تشکیل می شود. سپس آنالیز فاکتور روی این ویژگیهای آموزشی اعمال می شود و یک تابع توزیع بدست می دهد. اگر معیارهای آماری بالاتر از آستانه ای بودند ویژگیهای کاندید تعیین می شوند و سپس توسط ASM تایید می شوند. این روش ASM در مقاله ای دیگر با دو فیلتر کالمن توسعه داده شده است تا پارامترهای مستقل از شکل شدت نور را تخمین بزند و چهره را در دنباله تصاویر ردگیری کند.

۳-۲-۵- روشهای مبتنی بر ظاهر

بر خلاف روشهای تطبیق کلیشه که کلیشه ها از پیش توسط متخصصین تعریف می شدند، کلیشه ها در روشهای مبتنی بر ظاهر از مثالهای درون تصاویر یادگیری می شوند. بطور کلی، روشهای مبتنی بر ظاهر بر تکنیکهای تحلیل آماری و یادگیری ماشین برای پیدا کردن مشخصه های تصاویر چهره/غیر چهره تکیه دارند. مشخصه های یاد گرفته شده به صورت

²⁶⁴ snakes

²⁶⁵ point distribution model (PDM)

²⁶⁶ active shape model (ASM)

مدلهای توزیع یا توابع جداکننده هستند که برای تشخیص چهره از آنها استفاده می شود. در حال حاضر به دلیل محدود کردن حجم محاسبات و کارایی تشخیص، کاهش بعد انجام می گیرد.

بسیاری از روشهای مبتنی بر ظاهر در چارچوب احتمالاتی قابل تعریف هستند. یک تصویر یا بردار ویژگیها که از یک تصویر بدست آمده است به صورت یک متغیر تصادفی \vec{x} دیده می شود و این متغیر تصادفی برای کلاسهای چهره و غیرچهره با توابع توزیع احتمال کلاسی $p(\vec{x} | \text{face})$ و $p(\vec{x} | \text{nonface})$ معرفی می شود. دسته بندی بیزی یا بیشینه احتمال وقوع برای دسته بندی یک مکان کاندید تصویر برای چهره بودن یا نبودن می تواند استفاده شود. متأسفانه به خاطر ابعاد زیاد \vec{x} پیاده سازی مستقیم دسته بندی بیزی مناسب نیست. دلیل دیگر این امر عدم آگاهی درباره اینست که آیا بطور طبیعی شکلهای پارامتر شده این دو تابع توزیع کلاسی وجود دارد یا خیر. از این جهت بخش عمده یک روش مبتنی بر ظاهر به معنی ساختن تخمین پارامتری و غیرپارامتری این دو تابع اختصاص می یابد.

یکی دیگر از خط مشیها در روشهای مبتنی بر ظاهر یافتن یک تابع جداکننده (صفحه تصمیم، آبرصفحه جداساز، تابع آستانه) بین دو کلاس چهره و غیرچهره است. معمول است که الگوهای تصویر به فضاهای دارای ابعاد پایینتر بازتاب شود و سپس یک تابع جداکننده (معمولاً بر اساس معیارهای فاصله) برای دسته بندی شکل بگیرد و یا صفحه تصمیم غیرخطی توسط شبکه عصبی چندلایه تشکیل شود.

ماشینهای بردار پشتیبان و روشهای کرنلی نیز در این زمینه پیشنهاد شده اند. این روشها به طور ضمنی الگوها را به ابعاد بالاتر فضایی می برند تا بین دو کلاس الگوی چهره و غیرچهره که در آن فضا بازتاب شده اند، صفحه ای جداکننده تشکیل دهند.

۳-۲-۵-۱- چهره های یکه^{۲۶۷}

یکی از نخستین مثالهای استفاده از چهره های یکه در تشخیص چهره توسط کوهونن صورت پذیرفت که در آن یک شبکه عصبی ساده برای تشخیص چهره تصاویر چهره نرمال شده و همراستا استفاده شد. شبکه عصبی مشخصات چهره را با تخمین زدن بردار یکه ماتریس خود همبستگی محاسبه می کند. این بردارهای یکه بعدها به چهره های یکه معروف شدند.

در مقاله ای ثابت شده است که چهره می تواند به صورت خطی توسط تعداد محدودی از تصاویر پایه کد شوند. این اثبات بر اساس تبدیل کاراونن-لووه^{۲۶۸} است که به نامهای دیگری مانند تحلیل مؤلفه اصلی و تبدیل هتلینگ^{۲۶۹} نیز معروف است. این ایده توسط دو مقاله پیشنهاد شده است. با داشتن یک مجموعه از تصاویر آموزشی n در m پیکسل که به صورت یک بردار با سایز $m \times n$ نمایش داده می شوند، بردارهای پایه که زیر فضای بهینه را پوشش می دهند طوری تعریف می شوند که خطای میانگین مربعی بین بازتاب تصویر آموزشی به این زیرفضا و تصویر اصلی کمینه شود. این مجموعه بردارهای پایه بهینه تصویر یکه خوانده می شود چرا که در حقیقت همان بردارهای یکه ماتریس کواریانس هستند که از چهره بردار شده در مرحله آموزش بدست می آید.

²⁶⁷ eigenfaces

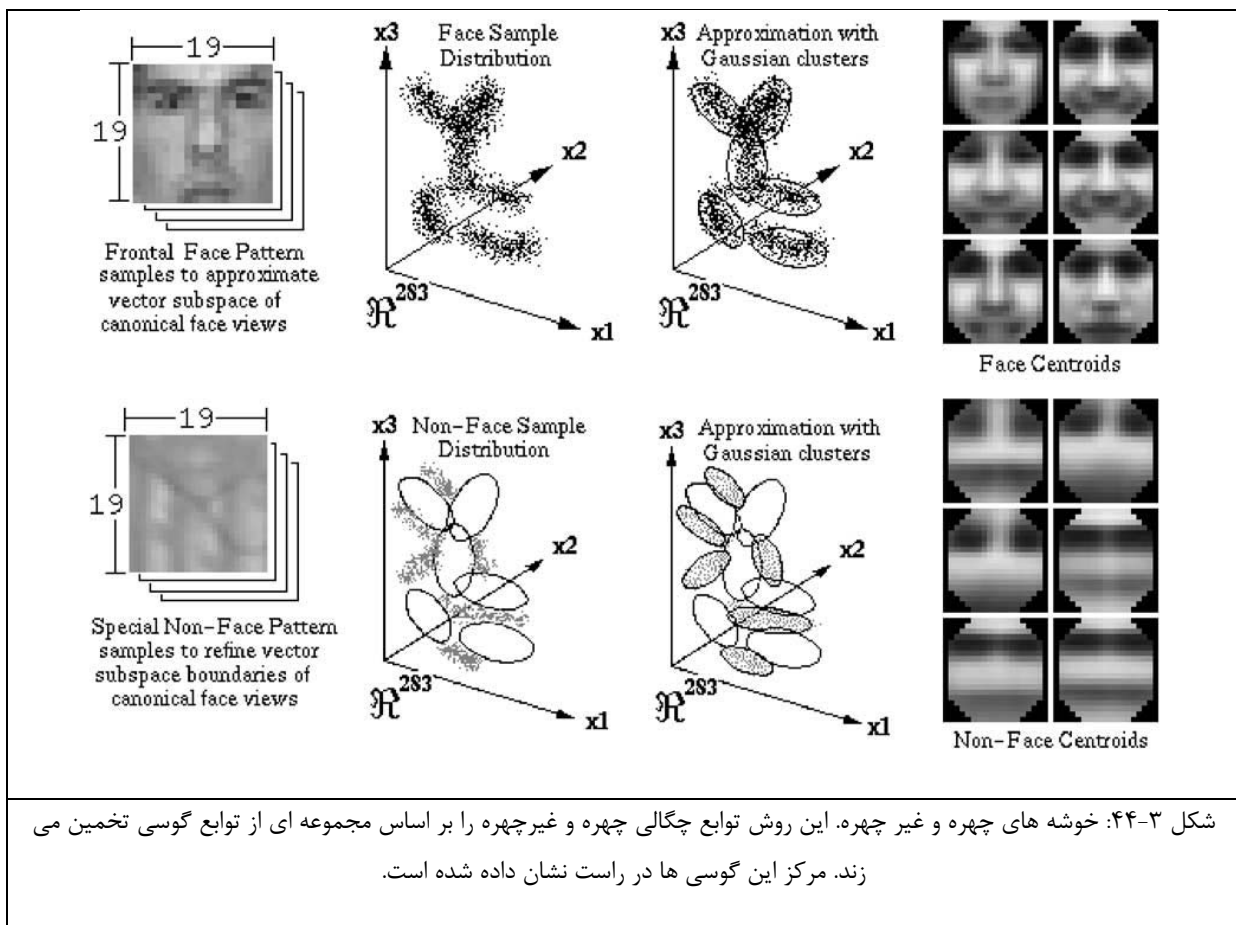
²⁶⁸ Karhunen-Loe`ve Transform

²⁶⁹ Hotelling transform

در مقاله دیگری تحلیل مؤلفه های اصلی برای آشکارسازی و شناسایی چهره بکار گرفته شده است. همانند روش قبل، تحلیل مؤلفه های اصلی روی مجموعه تصاویر آموزشی چهره اعمال شده است تا تصاویر یگه (در اینجا چهره های یگه) را تولید کند که یک زیر فضا از فضای تصویر که فضای چهره خوانده می شود جدا کند. تصاویر چهره بر روی این زیر فضا بازتاب شده و خوشه بندی می شوند. به طور مشابه تصاویر آموزشی غیر چهره نیز بر روی همان زیر فضا بازتاب شده و خوشه بندی می شوند. تصاویر چهره هنگامی که به فضای چهره بازتاب می شوند تغییر چندانی نمی کنند، در حالی که بازتاب تصاویر غیر چهره به این فضا کاملاً با تصویر اصلی متفاوت است. برای تعیین حضور یک چهره در صحنه، فاصله بین یک ناحیه تصویر و فضای چهره برای تمام نقاط تصویر محاسبه می شود. فاصله از فضای چهره به عنوان معیار چهره بودن استفاده می شود و حاصل محاسبه فاصله از فضای چهره را نقشه چهره می نامند. یک چهره می تواند از روی کمینه محلی نقشه چهره تشخیص داده شود. بسیاری از تحقیقات روی آشکارسازی چهره، شناسایی و استخراج ویژگیها از ایده تجزیه بردار یگه و خوشه بندی استفاده کرده اند.

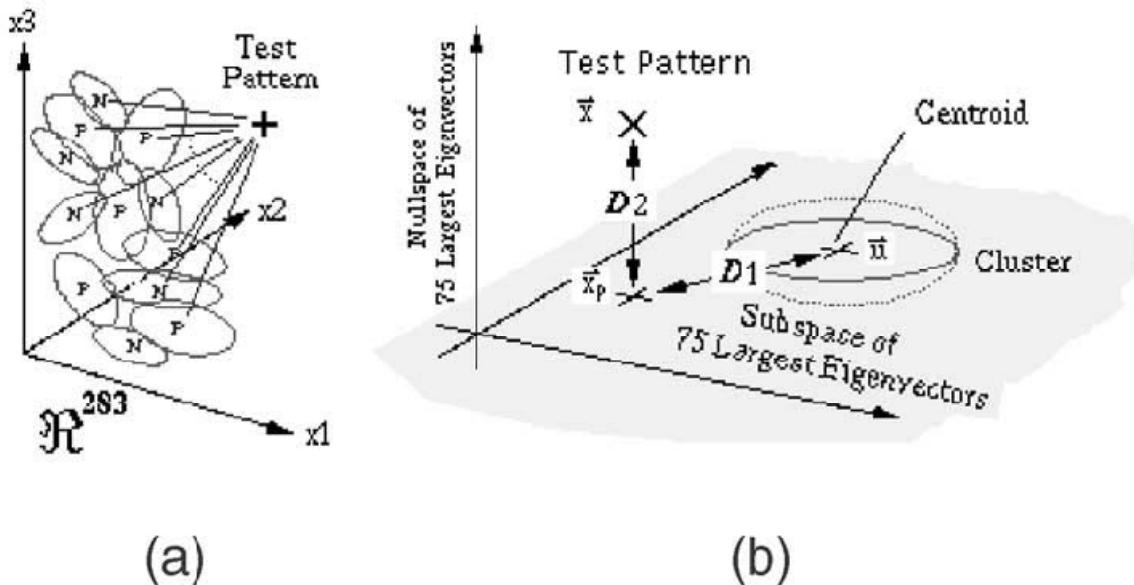
۳-۲-۵-۲- روشهای مبتنی بر توزیع

در یکی از مقالات معتبر یک سیستم مبتنی بر توزیع برای آشکارسازی چهره تهیه شده است که در آن نشان داده شده است که چگونه توزیع الگوهای تصویر برای یک کلاس شیء می تواند از نمونه های (تصاویر) مثبت و منفی کلاس استفاده کند.



این سیستم از دو مؤلفه تشکیل شده است: مدل‌های مبتنی بر توزیع الگوهای چهره/غیر چهره و یک پرسپترون چندلایه. هر مثال چهره و غیرچهره ابتدا نرمال شده به یک تصویر 19×19 پیکسلی تبدیل می‌شود و به عنوان یک بردار الگوی 361 بُعدی در نظر گرفته می‌شود. سپس الگوها به شش خوشه صورت و غیرصورت بر اساس یک الگوریتم k میانگین²⁷⁰ تغییر یافته تبدیل می‌شوند که در شکل ۳-۴۴ به تصویر کشیده شده است.

هر خوشه به صورت یک تابع گوسی چندبُعدی با یک تصویر میانگین و کواریانس نمایش داده می‌شود. شکل ۳-۴۵ معیار فاصله در این روش را نشان می‌دهد.



شکل ۳-۴۵: معیارهای فاصله. دو معیار فاصله بین الگوی تصویر ورودی و خوشه‌های نمونه‌ها حساب می‌شود.

اولین مؤلفه فاصله، فاصله ماھالانوبیس²⁷¹ نرمال شده بین الگوی تست و مرکز خوشه است که در زیرفضایی با ابعاد پایینتر که با ۷۵ بردار ویژه بزرگتر خوشه معین شده است، اندازه‌گیری می‌شود. دومین مؤلفه فاصله، فاصله اقلیدسی بین الگوی تست و بازتاب آن روی زیر فضای ۷۵ بُعدی است. این مؤلفه فاصله مسئول اختلافات الگویی است که با مؤلفه اول پوشش داده نشده بود. گام آخر استفاده از یک شبکه پرسپترون چندلایه برای دسته‌بندی الگوهای پنجره چهره از الگوهای غیرچهره است که با استفاده از دوازده جفت فاصله که بین هر خوشه چهره و غیرچهره حساب شده است انجام می‌شود. این دسته‌بندی کننده با روش پس انتشار و پایگاه داده ای با ۴۷۳۱۶ پنجره الگو آموزش می‌بیند. در آن ۴۱۵۰ نمونه مثبت و بقیه منفی هستند. توجه شود که جمع آوری الگوهای نماینده چهره آسان است ولی تولید الگوهای نماینده غیرچهره‌ها دشوارتر است. این مسئله با یک روش بوت استرپ²⁷² تخفیف داده شده است که در آن تصاویر را با انتخاب خود در طول آموزش به داده‌های آموزشی اضافه می‌کند. با شروع از یک مجموعه کوچک نمونه‌های غیرچهره در مجموعه آموزشی، دسته‌بندی کننده با این پایگاه داده از مثالها آموزش می‌بیند. سپس تشخیص دهنده چهره روی یک دنباله تصادفی از تصاویر اجرا

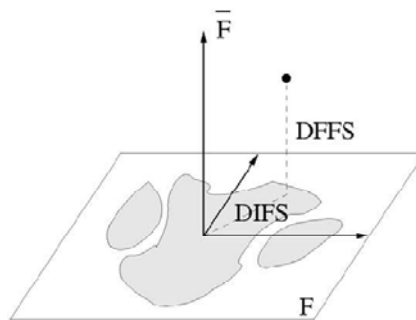
²⁷⁰ modified k-means algorithm

²⁷¹ Mahalanobis distant

²⁷² bootstrap

می شود و تمامی الگوهایی که سیستم فعلی اشتباهاً چهره اعلام کرده است را جمع آوردی می کند. داده های اشتباهاً مثبت سپس به پایگاه داده به عنوان نمونه های جدید غیرچهره اضافه می شوند. این روش بوت استرپ از مشکلات جمع آوری صریح نمونه های الگوهای غیرچهره اجتناب می کند و در کارهای بعدی نیز از آن استفاده شده است.

یک روش یادگیری احتمالاتی بینایی بر مبنای تخمین چگالی در فضای با ابعاد بالا به کمک تجزیه فضاهای یگه در مقاله دیگری آمده است. تحلیل مؤلفه های اساسی برای تعریف زیرفضایی که الگوهای چهره را بهینه بازنمایی می کند، استفاده شده است. این مؤلفه های اصلی، همبستگی خطی عمده بین داده ها را حفظ می کند و قسمت های خردتر آنرا دور میریزد. این روش فضای بردار را بین دو فضای منحصراً متقابل و مکمل تجزیه می کند (شکل ۳-۴۶)؛ چگالی در زیرفضای اصلی (که با مؤلفه های اصلی تعریف می شوند) و مکمل متعامد آن (که در PCA استاندارد دور ریخته می شود). یک گوسی چند متغیره و یک توزیع توأم گوسی برای یادگیری ویژگیهای محلی چهره استفاده می شود. این توزیعهای احتمال سپس بر مبنای تخمین بیشترین احتمال وقوع برای تشخیص شیء استفاده می شوند. روش پیشنهادی برای مکانیابی، کد کردن و شناسایی چهره استفاده می شود. در مقایسه با روش چهره یگه استاندارد، این روش پیشنهادی کارایی بهتری در شناسایی چهره نشان می دهد. در زمینه آشکارسازی چهره، این روش تنها به مکانیابی می پردازد.



شکل ۳-۴۶: تجزیه زیرفضای چهره به دو زیر فضای مؤلفه های اساسی F و فضای مکمل متعامد \bar{F}

در مقاله دیگری یک روش تشخیص بر اساس تلفیق تحلیل‌های فاکتور پیشنهاد شده است. تحلیل فاکتور^{۲۷۳} یک روش آماری برای مدل کردن ساختار کواریانس داده های دارای ابعاد بالا با استفاده از تعداد کمی از متغیرهای راکد است. روش FA از چند منظر مشابه PCA است با این حال PCA بر خلاف FA یک مدل چگالی مناسب برای داده ها درست نمی کند، چون در همه جای زیرفضای مؤلفه های اصلی آن هزینه کد کردن داده ها یکسان است (یعنی چگالی در این جهات نرمال نشده است). به علاوه PCA در برابر نویز غیرمستقل در ویژگیهای داده مقاوم نیست زیرا مؤلفه های همبند واریانس داده های ورودی را بیشینه می کنند که واریانسهای ناخواسته نیز جزو آنهاست.

۳-۲-۵-۳- شبکه های عصبی

شبکه های عصبی در مسئله های شناسایی الگو مانند شناسایی نوری حروف، شناسایی اشیاء، و راهبری خودکار روبات موفق عمل کرده است. از آنجایی که شناسایی چهره به صورت یک مسئله شناسایی الگوی دو کلاسه است، معماری

²⁷³ factor analysis = FA

های گوناگونی از شبکه عصبی برای آن پیشنهاد شده است. مزیت استفاده از این شبکه های عصبی برای آشکارسازی چهره امکانپذیر بودن آموزش این سیستم برای دربرگرفتن پیچیدگی تابع توزیع کلاسی الگوهای چهره است. با اینحال یک عیب آن اینست که معماری شبکه باید بسیار خوب تنظیم شود (تعداد لایه ها، تعداد گره ها، نرخهای یادگیری، ...) تا نتایج درخشانی به بار آورد.

یکی از اولین روشهایی که از شبکه عصبی سلسله مراتبی استفاده کرد بدین شکل بود. سطح اول متشکل از دو زیرشبکه موازی بود که در آن ورودیها مقادیر شدت روشنایی تصویر اصلی و شدت روشنایی تصویر فیلتر شده با فیلتر 3×3 سوبل²⁷⁴ هستند. ورودیهای شبکه سطح دوم از خروجیهای زیر شبکه های مقادیر ویژگیهای استخراج شده (مانند انحراف معیار مقادیر پیکسلهای الگوی ورودی، نسبت تعداد پیکسلهای سفید به کل پیکسلهای دوسطحی شده درون پنجره، و گشتاورهای هندسی) هستند. مقدار خروجی سطح دوم نشان دهنده حضور چهره در ناحیه است. آزمایشات نشان می دهند که این روش در حالتی که تمام چهره های درون تصویر آزمایش یک اندازه باشند می تواند چهره ها را پیدا کند.

مقاله دیگری یکی از اولین شبکه های عصبی برای آشکارسازی چهره را ارائه داده است. این شبکه دارای چهار لایه با 1024 و ورودی، 256 واحد در لایه پنهان اول، هشت واحد در لایه پنهان دوم، و دو واحد خروجی است. این پیکر بندی بعدها نیز مجدداً استفاده شده است.

یک مقاله به پویس تصویر ورودی با یک شبکه عصبی دارای تاخیر زمانی (با میدان گیرندگی 25×20 پیکسل) برای تشخیص اشیا اشاره کرده است. برای کنار آمدن با تغییرات اندازه، تصویر ورودی با تبدیل موجک تجزیه می شود.

روش دیگری برای تشخیص اشیا در تصاویر استفاده از شبکه های عصبی پیچشی²⁷⁵ است. ابتدا نمونه های تصاویر چهره و غیر چهره 20×20 ساخته شده و سپس یک شبکه عصبی برای یافتن مکان حدودی تصاویر در یک مقیاس آموزش داده می شود. شبکه دیگری برای تعیین مکان دقیق چهره ها در یک مقیاس آموزش می بیند. با دادن یک تصویر، شبکه اول نواحی احتمالی کاندید چهره بودن را تعیین می کند. این کاندیدها توسط شبکه دوم تصدیق می شوند.



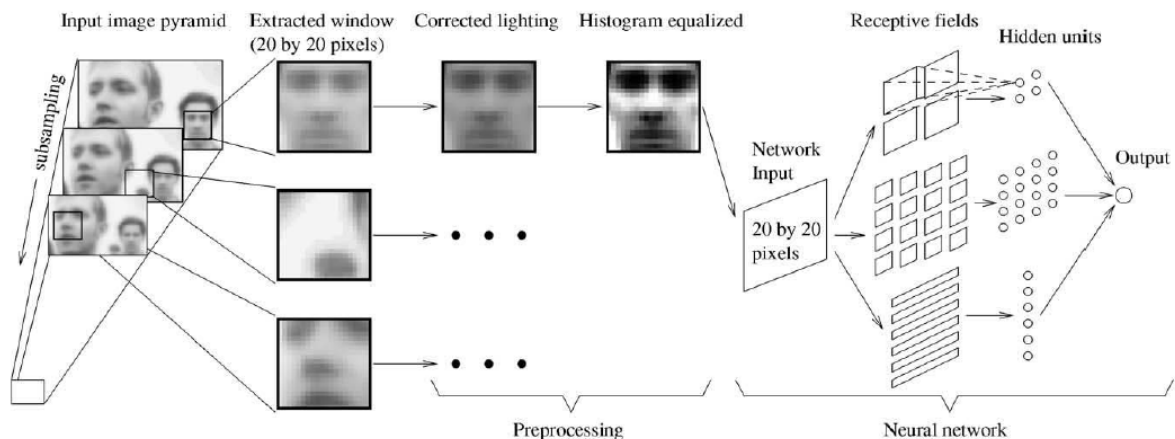
شکل ۳-۴۷: نمونه های هر کلاس چهره برای SOM کوهونن که هر کدام مرکز یک خوشه هستند.

²⁷⁴ Sobel filters

²⁷⁵ convolutional neural network

روش دیگری از تشخیص چهره با شبکه عصبی وجود دارد که در آن تعداد زیادی نمونه از چهره ها و غیرچهره ها توسط الگوریتم SOM کوهونن به تعداد کمتری از نمونه های فشرده تبدیل می شوند. یک پرسپترون چندلایه برای یادگیری این مثالها مبادرت می کند. فاز تشخیص از پویش هر تصویر در دقت های مختلف تشکیل شده است. برای هر مکان و اندازه پنجره پویش، محتوای آن به یک اندازه ثابت نرمالسازی می شوند و میانگین و واریانس برای کاستن از تاثیر شرایط نورپردازی نرمالسازی می شوند. سپس هر پنجره نرمالیزه با یک MLP دسته بندی می شود (شکل ۳-۴۷).

در مقاله ای دیگر که بی شک یکی از موفقترین روشهای آشکارسازی چهره با شبکه های عصبی بوده است و به روش رولی^{۲۷۶} معروف است، از دو مؤلفه استفاده می شود: مؤلفه ای شامل چندین شبکه عصبی برای تشخیص الگوهای چهره و مؤلفه تصمیم گیری که تصمیم نهایی را بر مبنای خروجی شبکه های مؤلفه قبل بدست می دهد. همانطور که در شکل ۳-۴۸ نشان داده شده است، نخستین مؤلفه این روش شبکه عصبی ایست که یک ناحیه 20×20 پیکسلی از تصویر را می گیرد و یک خروجی در بازه ۱- تا ۱ بدست می دهد. با دادن یک الگوی آزمون، خروجی شبکه عصبی نشان می دهد که تصویر یک چهره است (نزدیک به ۱) یا الگویی غیرچهره دارد (نزدیک به ۰). برای آشکارسازی چهره در هر جای تصویر، شبکه عصبی به تمام نقاط تصویر اعمال می شود. برای آشکارسازی چهره هایی بزرگتر از 20×20 پیکسل، تصویر ورودی باید مکرراً زیرنمونه گیری شود و شبکه به هر مقیاس اعمال شود. نزدیک به 1050 نمونه چهره از اندازه ها، راستاها، موقعیتها و نورپردازیها برای آموزش شبکه استفاده شده اند. در هر تصویر ورودی، چشمها، نوک بینی، و گوشه ها و وسط دهان به طور دستی برچسب گذاری شده اند و برای نرمال کردن چهره به همان مقیاس، راستا و موقعیت استفاده شده اند. مؤلفه دوم این روش باید تشخیصهای دارای همپوشانی را ترکیب کرده و بین خروجی چند شبکه عصبی یکی را انتخاب کند. روشهای قضاوت^{۲۷۷} ساده ای مانند عملگرهای منطقی (AND و OR) و روشهای رأی گیری برای افزایش کارایی استفاده شده اند. چندین سیستم دیگر نیز با روشهای قضاوت متفاوت تولید شده اند.



شکل ۳-۴۸: دیاگرام سیستم روش رولی. هر چهره پیش از خوانده شده به شبکه های عصبی پیش پردازش می شود. برای تعیین اینکه بر اساس خروجی این شبکه آیا چهره ای در تصویر موجود است چندین روش قضاوت استفاده می شود.

²⁷⁶ Rowley

²⁷⁷ arbitrate scheme

یکی از محدودیتهای این روش اینست که تنها تصاویر تمام رخ و صاف را آشکارسازی می کند. این مدل با استفاده از یک شبکه مسیریاب توسعه داده شده است. چنین شبکه ای هر پنجره ورودی را پردازش می کند تا راستای چهره احتمالی را تعیین کند و پنجره را به راستای کانونی دوران دهد. این پنجره ورودی به عنوان ورودی مدل بالا استفاده می شود. با وجود اینکه این سیستم برای تصاویر صاف نرخ آشکارسازی کمتری از حالت اولیه دارد. این سیستم روی مجموعه آزمایشی بزرگی دارای ۷۶/۹ درصد آشکارسازی و پاسخهای اشتباهاً مثبت اندکی است.

۳-۲-۵-۴- ماشینهای بردار پشتیبان

ماشینهای بردار پشتیبان^{۲۷۸} می توانند برای آموزش دسته بندی کننده های تابع چند جمله ای، شبکه های عصبی، و تابعهای شعاعی استفاده شوند. در حالی که خیلی از روشها آموزش یک دسته بندی کننده (مثلاً بیزی، شبکه عصبی، و تابعهای شعاعی) بر مبنای کمینه کردن خطای آموزش و ریسک تجربی است، SVM ها بر مبنای قاعده استنتاجی دیگری کار می کنند که کمینه کردن ریسک ساختاری نام دارد و هدف آن اینست که کران بالای خطای تعمیم را حداقل کند. یک دسته بندی کننده SVM یک دسته بندی کننده خطی است که در آن ابرصفحه جدا کننده طوری انتخاب می شود که خطای دسته بندی مورد انتظار را برای الگوهای دیده نشده کمینه کند. این صفحه بهینه با ترکیب وزندار مجموعه کوچکی از بردارهای آموزشی که بردارهای پشتیبان نامیده می شوند تعریف می شود. تخمین این ابرصفحه بهینه برابر حل کردن یک مسئله برنامه نویسی درجه دو با قیود خطی است. بهر حال محاسبات از هر دو لحاظ زمان و حافظه حجیم است.

در مقاله ای روشی کارا برای آموزش SVM در مسائل بزرگ پیشنهاد شده و برای تشخیص چهره اعمال شده است. بر اساس دو مجموعه داده حاوی ۱۰۰۰۰۰۰ الگوی آزمون ۱۹×۱۹ پیکسل، سیستم ارائه شده تقریباً به دقت بهتری از روشهای مبتنی بر توزیع دست یافتند در حالی که زمان آن بسیار کمتر بوده است. همچنین SVMها برای تشخیص چهره ها و عابران پیاده در دامنه موجک نیز استفاده شده اند.

۳-۲-۵-۳- شبکه تنک غربالها^{۲۷۹}

مقاله ای یک روش با استفاده از معماری یادگیری SNoW برای تشخیص چهره پیشنهاد داده است تا چهره هایی با ویژگیها و حالات متفاوت، در موقعیتهای گوناگون، و تحت شرایط متفاوت نورپردازی را تشخیص دهد. این مقاله همچنین به بررسی اثرات یادگیری با مؤلفه های اساسی و نیز با ویژگیهای چندمقیاسه نیز پرداخته است. در واقع SNoW یک شبکه تنک از توابع خطی است که از قوانین بروزسانی غربال بهره می جویند. این روش بخصوص برای یادگیری در زمینه هایی طراحی شده است که تعداد بالقوه ویژگیهای دخیل در تصمیمات بسیار زیاد باشد ولی از پیش این امر مشخص نباشد. برخی از مشخصات این معماری یادگیری عبارتند از: واحدهایی که تنک بهم متصل هستند، تخصیص ویژگیها و روابط بصورت داده گرا، مکانیسم تصمیم گیری، و استفاده از یک قانون بروزسانی کارا. این روش در قیاس با اکثر روشهای دیگر آشکارسازی چهره از لحاظ محاسبات برتری دارد در حالی که دقتی در سطح آنها را فراهم می کند.

²⁷⁸ support vector machines

²⁷⁹ sparse network of winnows = SNoW

۳-۲-۵-۶- دسته بندی کننده ساده بیزی

یکی از مقالات معروف در آشکارسازی چهره، یک دسته بندی کننده ساده بیزی را برای تخمین احتمال توأم ظاهر محلی و موقعیت الگوهای چهره (زیرنواحی چهره) در دقتهای مختلف ارائه می دهد. این روش بر ظاهر محلی تاکید دارد زیرا برخی الگوهای محلی یک شکل، از دیگر الگوها یکتاترند. الگوهای شدت روشنایی در اطراف چشم بسیار متمایزتر از اطراف گونه ها هستند. دو دلیل برای استفاده از دسته بندی کننده خام بیزی (و به عبارت دیگر عدم وجود وابستگی آماری بین زیرنواحی) وجود دارد. نخست اینکه این مسئله تخمین بهتری از تابع توزیع شرطی این زیر ناحیه ها می زند. دوم اینکه یک دسته بندی کننده ساده بیزی فرم عملیاتی از احتمال ثانویه را فراهم می آورد که می تواند آمار توأم ظاهر محلی و موقعیت شیء را ذخیره می کند. در هر مقیاس یک تصویر چهره به چهار زیرناحیه مستطیلی تجزیه می شود. این زیرناحیه ها سپس با کمک PCA به فضایی با ابعاد بازتاب می شوند و آمارگان هر زیرناحیه بازتاب شده از نمونه های بازتاب شده برای ظاهر محلی تخمین زده می شود. با این فرمولاسیون چهره زمانی حاضر است که نسبت احتمال وجود چهره از احتمال ثانویه بیشتر باشد. این روش کارایی نسبتاً خوبی دارد و می تواند چهره های چرخیده و نیمرخها را نیز تشخیص دهد. نویسندگان این مقاله، بعدها آنرا با ارائه موجک ارتقا دادند.

۳-۲-۵-۷- مدل مخفی مارکوف

در مدل مخفی مارکوف²⁸⁰ (HMM) فرض می شود که الگوهای داده ها را می توان با یک فرایند تصادفی به صورت پارامتری مدل سازی کرد، به طوری که پارامترهای این مدل با یک روش دقیق قابل محاسبه باشند. در طراحی HMM برای مسائل بازشناسی الگو برای تعیین ساختار مدل، ابتدا باید تعدادی حالت مخفی در نظر گرفت. سپس می توان HMM را آموزش داد تا احتمال انتقال میان حالات را از روی نمونه های آموزشی که رشته های از مشاهدات هستند، فراگیری کند. هدف از آموزش یک HMM، پیشینه سازی احتمال مشاهده داده های آموزشی از طریق تغییر و اصلاح پارامترهای موجود در مدل HMM با استفاده از الگوریتم های استاندارد تقطیع ویتربی²⁸¹ و الگوریتم بام-ولش²⁸² است. پس از آموزش HMM، احتمال خروجی یک مشاهده، کلاس داده موردنظر را تعیین می کند.

به صورت شهودی می توان الگوی چهره را به چند ناحیه از جمله پیشانی، چشمها، بینی، دهان و چانه تقسیم کرد. یک الگوی چهره را می توان با فرایندی که طی آن این ناحیه ها در یک ترتیب مناسب (از بالا به پایین و از چپ به راست) مشاهده می شوند، بازشناسی کرد. به جای تکیه کردن بر روی ترتیب صحیح، مانند تطبیق الگو یا روش های مبتنی بر ظاهر (که در آنها ویژگی های چهره مانند چشمها و بینی باید با توجه به یک نقطه مرکزی مرتب باشند)، در این روش، هدف، ایجاد تناظر میان ناحیه های مختلف چهره با حالاتی از یک مدل مخفی مارکوف با چگالی پیوسته است. روش های مبتنی بر HMM معمولاً با یک الگوی چهره به صورت رشته ای از بردارهای مشاهده برخورد می کنند. هر یک از این بردارها، همانگونه که در شکل ۳-۴۸ چپ نشان داده شده، مجموعه ای از پیکسل ها می باشد.

²⁸⁰ hidden Markov model

²⁸¹ standard Viterbi segmentation method

²⁸² Baum-Welch algorithm



شکل ۳-۴۹: مدل مخفی مارکف برای شناسایی موقعیت چهره.

برای الگوهای چهره، همانگونه که در شکل ۳-۴۹ راست نمایش داده شده، مرزهای میان مجموعه پیکسل‌ها به صورت انتقال احتمالی میان حالت‌ها نشان داده می‌شود و داده‌های تصویر که درون یک ناحیه قرار می‌گیرند با یک توزیع گوسی چندمتغیره مدل‌سازی می‌گردد. یک رشته مشاهده حاوی تمامی مقادیر شدت از یک بلوک تصویر است. حالت‌های خروجی متناظر با کلاس‌هایی هستند که مشاهدات به آنها تعلق دارند. پس از آموزش HMM، احتمال خروجی یک مشاهده تعیین‌کننده کلاسی است که به آن تعلق دارد.

روش HMM برای بازشناسی چهره و مکان‌یابی آن مورد استفاده قرار گرفته است. یک مقاله نشان داد که حالت‌های HMM متناظر با ناحیه‌های چهره هستند (شکل ۳-۴۸ راست). به عبارت دیگر، یک حالت مسئول مدل کردن بردارهای مشاهده پیشانی انسان، و حالت دیگر مسئول مدل‌سازی بردارهای مشاهده چشم‌های انسان است. برای مکان‌یابی چهره، یک HMM با مجموعه وسیعی از تصاویر چهره آموزش داده می‌شود تا یک مدل کلی از چهره انسان ایجاد شود. اگر درست‌نمایی چهره حاصل از هر الگوی مستطیلی درون تصویر از یک حد آستانه بیشتر باشد، آن بخش به عنوان مکان یک چهره شناسایی می‌شود.

مقاله‌های دیگری HMM‌های یک‌بعدی و دوبعدی را برای استخراج ویژگی‌های چهره و بازشناسی آن مورد استفاده قرار دادند. در HMM‌های آنها از ساختار یک چهره برای اعمال محدودیت‌های مربوط به انتقال میان حالات، بهره برداری می‌شود. به دلیل اینکه محدوده‌های مشهود چهره مانند مو، پیشانی، بینی، چشم‌ها و دهان با ترتیب طبیعی و از بالا به پایین قرار دارند، هر یک از این نواحی در HMM پیوسته تک‌بعدی به یک حالت نسبت داده می‌شود. شکل ۳-۴۸ راست این ۵ حالت مخفی را نشان می‌دهد. برای آموزش، هر تصویر به شکلی یکنواخت، از بالا به پایین به ۵ حالت قطعه‌بندی می‌شود (به عبارت دیگر، هر تصویر به پنج ناحیه که همپوشانی ندارند و هم‌اندازه هستند تقسیم می‌شود). این تقسیم‌بندی یکنواخت در مرحله بعد با استفاده از تقطیع ویتربی جایگزین شده و پارامترهای HMM با روش بام-ولش مجدداً تخمین زده می‌شود. همانطور که در شکل ۳-۴۸ راست نشان داده شده، هر تصویر چهره با عرض W و ارتفاع H به چند ناحیه با ارتفاع L و عرض W که دارای همپوشانی هستند تقسیم گردیده است. تعداد P سطر همپوشانی میان بلوک‌های مجاور تصویر وجود دارد و HMM آموزش داده شده برای تعیین وضعیت خروجی مورد استفاده قرار می‌گیرد.

به طریقی مشابه مقاله ای دیگر از HMM و تبدیل کارهون-لووه (KLT) برای مکان‌یابی و بازشناسی چهره استفاده نموده است. به جای استفاده از مقادیر شدت روشنایی به صورت خام، بردارهای مشاهده حاوی ضرایب KLT هستند که از بردارهای ورودی محاسبه می‌شوند. نتایج آزمایشات آنها حاکی از آن بود که بازشناسی چهره از این طریق نتایج بهتری را

در مقایسه با روش قبل در پی دارد. در دادگان MIT، که حاوی ۴۳۲ تصویر از چهره‌های جداگانه است، این سیستم HMM دوبردی، ۹۰ درصد دقت به همراه داشته است.

در تحقیقی دیگر دو روش احتمالاتی برای تشخیص چهره ارائه شده است. در مقایسه با روشی که از یک مجموعه گوسی چندمتغیره برای مدل کردن توزیع الگوی چهره‌ها استفاده می‌کرد، اولین روش از آمارگان مرتبه بالاتر برای تخمین چگالی احتمال استفاده کرده است. هر دو توزیع ناشناخته چهره‌ها و تصاویر غیرچهره بر اساس آمار مراتب بالاتر با ۶ تابع چگالی خوشه‌بندی می‌شوند و از یک شبکه پرسپترون چندلایه برای دسته‌بندی استفاده می‌شود. در این روش بردار ورودی شامل ۱۲ معیار فاصله میان الگوی تصویر و ۱۲ خوشه مدل است. روش دوم این مقاله از یک HMM برای یادگیری انتقال میان نواحی غیرچهره به چهره و چهره به غیرچهره در یک تصویر بهره می‌برد. این روش مبتنی بر تولید یک رشته مشاهده از تصویر و یادگیری پارامترهای HMM متناظر با این رشته است. رشته مشاهده‌ای که باید یادگیری شود ابتدا با محاسبه فاصله زیرتصویر تا مرکز ۱۲ خوشه چهره و غیرچهره که در روش اول تخمین زده شد، محاسبه می‌گردد. پس از پایان مرحله یادگیری، رشته حالت بهینه برای دسته‌بندی دودویی باز هم پردازش می‌شود. نتایج آزمایشات نشان داده است که هر دو روش نرخ شناسایی بالاتری نسبت به روش پایه داشته‌اند داشتند، اما اعلام‌های اشتباه آنها نیز بالاتر بود.

۳-۲-۵-۸- روش نظریه اطلاعاتی

ویژگی فضایی الگوی چهره از جهات مختلف قابل مدلسازی است. محدودیت‌های محتوایی در این میان از دیگر موارد قدرتمندتر است. محدودیت‌های محتوایی در یک الگوی چهره معمولاً با استفاده از یک همسایگی کوچک از پیکسل‌ها تعیین می‌شوند. نظریه زمینه تصادفی مارکوف (MRF) رویکردی یکنواخت و مناسب برای مدلسازی موجودیت‌های وابسته به محتوا مانند پیکسل‌های تصویر و ویژگی‌های همبسته ارائه می‌کند. این مسئله با استفاده از تعیین تاثیر متقابل میان این موجودیت‌ها از طریق توزیع‌های شرطی MRF به دست می‌آید. بنا بر نظریه همزلی-کلیفورد^{۲۸۳}، یک MRF می‌تواند به صورت متوازن با یک توزیع گیبس تعیین شود و پارامترهای آن معمولاً با تخمین بیشینه استنتاجی (MAP) محاسبه می‌شوند. همچنین، توزیع چهره‌ها و تصاویر غیرچهره را می‌توان با هیستوگرام‌ها نیز تخمین زد. با استفاده از اطلاعات نسبی کولبک^{۲۸۴}، فرایند مارکف که تمایز مبتنی بر اطلاعات را میان دو دسته بیشینه کند، موجود است و می‌تواند برای مکان‌یابی مورد استفاده قرار گیرد.

مقاله ای اطلاعات نسبی کولبک را برای مکان‌یابی چهره استفاده کرده است. برای این کار تابع توزیع $p(x)$ به چهره بودن تصویر نسبت داده شده و تابع توزیع $q(x)$ به الگوی تصویر غیرچهره منسوب می‌شود. یک دادگان آموزشی حاوی ۱۰۰ چهره که از هر یک ۹ زاویه وجود داشت برای تخمین توزیع چهره مورد استفاده قرار گرفت. تابع چگالی احتمال غیرچهره از یک مجموعه شامل ۱۴۳۰۰۰ الگوی غیرچهره با استفاده از هیستوگرام‌ها تخمین زده شد. پیکسل‌هایی که حاوی بیشترین اطلاعات^{۲۸۵} (MIP) بودند از درون مجموعه آموزشی انتخاب شدند تا اطلاعات نسبی کولبک را میان $p(x)$ و $q(x)$ بیشینه کنند (به عبارت دیگر، بیشترین امکان جداسازی دو کلاس را فراهم کنند). بنابر آنچه حاصل شده است، توزیع

²⁸³ Hamersley-Clifford

²⁸⁴ Kullback relative information

²⁸⁵ most informative pixels = MIP

MIP بر روی نواحی دهان و چشم‌ها متمرکز شده و می‌توان از ناحیه بینی صرف‌نظر کرد. در مرحله بعد از MIP برای بدست آوردن ویژگی‌های خطی برای دسته‌بندی و بازنمایی استفاده می‌شود. برای آشکارسازی چهره‌ها، یک پنجره از روی تصویر ورودی عبور می‌نماید و فاصله از فضای چهره (DFFS) محاسبه می‌گردد. اگر این فاصله تا زیرفضای چهره از فاصله تا زیرفضای غیرچهره کمتر باشد، این نتیجه حاصل خواهد شد که در آن پنجره یک تصویر چهره وجود دارد.

اطلاعات نسبی کولبک در مقاله ای دیگر نیز برای بیشینه کردن تمایز مبتنی بر اطلاعات میان نمونه‌های مثبت و منفی چهره‌ها به کار گرفته شده است. تصاویر آموزشی هر کلاس (کلاس چهره و کلاس غیرچهره) به صورت مشاهدات یک فرایند تصادفی مورد تحلیل قرار گرفته و با دو تابع احتمال توصیف می‌شوند. در این پژوهش از یک گروه از فرایندهای پیوسته مارکوف به منظور مدل‌سازی الگوی چهره و پس‌زمینه و تخمین مدل احتمال استفاده شده است. با این تدبیر، فرایند آموزش تبدیل به یک فرایند بهینه‌سازی می‌شود که هدف آن انتخاب یک فرایند مارکوف است که بیشینه تمایز مبتنی بر اطلاعات را میان دو کلاس فراهم کند. با استفاده از مدل احتمال آموزش داده شده درستی محاسبه گردیده و برای شناسایی مکان چهره‌ها به کار گرفته می‌شود.

مقاله ای دیگر روشی را ارائه کرده است که از استراتژی‌های مبتنی بر دید و مبتنی بر مدل استفاده می‌کند. در ابتدا، الگوریتمی که تمرکز آن روی ویژگی‌های بینایی بوده و از اطلاعات سطح بالا استفاده می‌کند، برای کاهش فضای جستجو مورد استفاده قرار می‌گیرد. این امر با انتخاب محدوده‌هایی از تصویر که تصویر هدف ممکن است در آنها ظاهر شود، با استفاده از نقشه‌های ناحیه‌ای که توسط یک الگوریتم شناسایی ناحیه تولید می‌شوند (روش آب پخشان^{۲۸۶}) محقق می‌گردد. در درون این نواحی انتخاب شده، با استفاده از ترکیبی از تطبیق الگو و روش‌های تطبیق ویژگی و با بهره‌گیری از یک زمینه تصادفی مارکوف سلسله‌مراتبی و یک تخمین بیشینه احتمال ثانویه، چهره‌ها، مکان‌یابی و شناسایی می‌شوند.

۳-۲-۵-۹- یادگیری استنتاجی

الگوریتم‌های یادگیری استنتاجی نیز برای شناسایی و مکان‌یابی چهره‌ها به کار گرفته شده‌اند. در یک مقاله از الگوریتم C4.5 به منظور یادگیری درخت تصمیم از نمونه‌های مثبت و منفی الگوهای چهره استفاده شده است. هر نمونه آموزشی، یک پنجره ۸×۸ پیکسلی است که با یک بردار حاوی ۳۰ ویژگی بازنمایی می‌شود. این ویژگی‌ها عبارتند از: آنترابی، میانگین و انحراف معیار شدت روشنایی پیکسلها. الگوریتم C4.5 با استفاده از این نمونه‌ها یک دسته بندی کننده را به شکل یک درخت تصمیم ایجاد می‌کند که گره‌های آن یک مجموعه عملیات آزمون خاص را بر روی یک ویژگی خاص اعمال می‌کنند. درخت تصمیم یادگیری شده سپس برای تصمیم‌گیری بر روی وجود یا عدم وجود چهره در تصویر ورودی مورد استفاده قرار می‌گیرد. آزمایشات نشان می‌دهد که بر روی یک مجموعه از ۲۳۴۰ تصویر چهره تمام‌رخ از دادگان FERET، مکان‌یابی چهره‌ها با دقت ۹۶ درصد انجام گرفته است.

مقاله ای دیگر روشی را ارائه کرده است که در آن از الگوریتم Find-S میچل برای یادگیری مفهوم چهره استفاده می‌شود. نویسندگان این مقاله آنها تصور می‌کرده اند که توزیع الگوهای چهره $p(x|face)$ را می‌توان با مجموعه‌ای از

خوشه‌های گوسی تخمین زد و همچنین فاصله یک نمونه چهره تا یکی از مراکز خوشه باید کمتر از یک نسبتی از بیشینه فاصله نقاط درون آن خوشه تا مرکز آن خوشه باشد. سپس الگوریتم Find-S اعمال می‌گردد تا فاصله حد آستانه را فراگیری کند، به گونه‌ای که تصاویر چهره و غیرچهره از یکدیگر متمایز شوند. این روش ویژگی‌های خاصی دارد. نخست اینکه از نمونه های تصاویر غیرچهره (منفی) استفاده نمی‌کند، در حالی که دو روش قبل هر دو از تصاویر چهره و غیرچهره در طراحی مدل استفاده می‌کردند، دوم اینکه تنها محدوده مرکزی چهره برای آموزش استفاده می‌شود، و سوم اینکه بردارهای ویژگی حاوی تصاویری با ۳۲ سطح شدت روشنایی هستند، در حالی که سایر روشها از بیشترین سطح ممکن برای مقادیر شدت روشنایی به عنوان ورودی استفاده می‌کنند. این روش به دقت ۹۰ درصد شناسایی بر روی اولین مجموعه دادگان CMU دست یافته است.

۳-۲-۵- روشهای مدرن

در این بخش روش‌های شناسایی چهره را در چهار گروه اصلی بررسی و دسته‌بندی شده اند. اگرچه، برخی از روش‌ها می‌توانند در بیش از یک گروه طبقه‌بندی شوند. برای مثال، روش‌های تطبیق الگو معمولاً از یک مدل چهره و چند زیرالگو برای استخراج ویژگی‌های چهره استفاده کرده و سپس این ویژگی‌ها را برای مکان‌یابی یا شناسایی چهره‌ها به کار می‌برند. علاوه بر این، مرز میان روش‌های مبتنی بر دانش و برخی روش‌های تطبیق الگو شفاف نیست، زیرا روش‌های مبتنی بر تطبیق الگو عموماً به صورت ضمنی از دانش انسان برای تعیین الگوهای چهره استفاده می‌کنند. در طرف مقابل، روش‌های آشکارسازی چهره را می‌توان به صورت دیگری نیز طبقه‌بندی کرد. برای مثال، این روش‌ها را می‌توان بر اساس وابستگی آنها به ویژگی‌های محلی چهره دسته‌بندی کرد و یا آن را به یکپارچه^{۲۸۷} مورد بررسی قرار داد. در این بین روشهای مدرنی نیز پدید آمده اند که به اجمال به آنها اشاره می‌شود (Zhang 2010) و در کارهای آتی در دسته بندیها گنجانده خواهند شد.

۳-۲-۶- جمع بندی

در این پژوهش سعی بر این بوده است که مروری بر ایده های آشکارسازی چهره انجام پذیرد و دسته های ساختاری برای روشهای معرفی شده در آنها ارائه شود. این اثر به عنوان تکمیل یک مقاله مروری (Yang 2002) با الگوریتم هایی از جدیدترین روشهای تشخیص چهره ارائه شده است که در طبقه بندی ارائه شده توسط این مقاله گنجانده شده اند. آنچه که ادبیات نیاز زیادی به آن احساس میکند، داده های آموزشی و آزمایشی استاندارد و رویه های ارزیابی کارایی نظام مندتر است. این امر کمک می کند تا کاربران الگوریتمهای آشکارسازی چهره بفهمند که کدام روشها در کدام زمینه رقابتی جدی تری به شمار می آیند. همچنین محققان را به سمت تولید الگوریتمهایی که واقعاً کارا باشند سوق می دهد.

با وجود پیشرفتهای چشمگیری که در سالیان اخیر دیده شده است، همچنان کار زیادی باقی مانده است. برای رقابت در دنیای امروزی الگوریتمهای آشکارسازی چهره، سیستم تهیه شده تحت تنوعهای زیادی باید کارایی خوبی از خود نشان دهد: شرایط نورپردازی، تغییرات پس زمینه، راستا، موقعیت، و انسداد جزئی، حالات چهره، و حضور عینک، موهای چهره، و تنوع مدلهای موی سر، و گریم.

آشکارسازی چهره به تنهایی یک مسئله دشوار و جالب است. با این وجود می‌تواند به عنوان یکی از تلاشهایی که برای حل تشخیص اشیا - که از بزرگترین چالشهای بینایی ماشین است - صورت گرفته است نیز دیده شود. کلاس چهره دارای تنوع زیادی در شکل، رنگ و درونمایه است که این به دلیل تفاوت افراد، صلب نبودن، موی صورت، عینک، و گریم و آرایش است. تصاویر تحت شرایط گوناگون نورپردازی و موقعیت سه بعدی و گاهاً در پس زمینه های درهم ریخته گرفته می‌شوند. بنابراین آشکارسازی چهره با جمیع مشکلاتی که برای کاربرد عام آن، تشخیص اشیا، نیز وجود دارد دست و پنجه نرم می‌کند. بهرحال کلاس چهره دارای قاعده مندی بسیار آشکاری است که توسط بسیاری از الگوریتمهای اکتشافی و روشهای مبتنی بر مدل به خدمت گرفته می‌شود و یا توسط روشهای داده گرا، یاد گرفته می‌شود. ممکن است برخی از قواعد در تعریف کلاس بطور کلی مورد انتظار باشد ولی ممکن است آنقدرها هم آشکار نباشد. در خاتمه باید گفت که هرچند چهره ها دارای تنوع درون کلاسی شگرفی هستند، ولی مسئله آشکارسازی چهره، به سادگی، یک مسئله شناسایی دو کلاسه (چهره در برابر غیر چهره) است.

در پایان مهمترین مدلهای بکار گرفته شده در آشکارسازی چهره مرور می‌شود:

تحلیل مؤلفه اصلی: از تبدیل کارهونن-لووه مشتق شده است. با داشتن یک بازنمایی t بعدی از هر چهره در مجموعه آموزش تصویر، PCA به دنبال یک زیرفضای t بعدی می‌گردد که بردارهای پایه آن متناظر با جهت‌های بیشترین واریانس در فضای اصلی تصویر باشند. این یزفضای جدید دارای ابعاد کمتری است ($t \ll S$). اگر عناصر تصویر به عنوان متغیرهای تصادفی در نظر گرفته شوند، بردارهای پایه PCA به عنوان بردارهای یگه نمودار پراکندگی آنها خواهند بود و تصویر حاصل از آنها را چهره یگه می‌نامند.

تحلیل مؤلفه مستقل: این روش وابستگیهای مرتبه دوم و مراتب بالاتر در تصویر ورودی را کمینه می‌کند و تلاش می‌کند که بردارهای پایه خود را که داده ها روی آنها نگاشت داده می‌شوند، مستقل آماری انتخاب کند. دو معماری برای آشکارسازی چهره از این روش استفاده می‌کند: ۱- تصاویر پایه مستقل آماری ۲- بازنمایی کد جدایی پذیر

تحلیل جداکننده خطی: این روش بردارهایی را در فضا می‌یابد که کلاسهها را به بهترین نحو متمایز کنند. برای تمام کلاسهها، ماتریس پراکندگی درون کلاسی (S_B) و بین کلاسی (S_W) تعریف می‌شوند. هدف این روش بیشینه سازی S_B در عین کمینه سازی S_W است که می‌توان آنرا به صورت بیشینه سازی نرخ $\det|S_B|/\det|S_W|$ هم دید. این نرخ زمانی بیشترین مقدار خود را دارد که بردارهای ستونی افکنش ماتریس برابر بردارهای یگه $S_W^{-1} \times S_B$ باشند.

تعقیب تکاملی:^{۲۸۸} یک روش تطبیقی مبتنی بر چهره های یکه که بدنبال بهترین محورهایی می‌گردد که افکنش روی آنها یک تابع برازش را بیشینه کند و در همان حال دقت دسته بندی و قدرت تعمیم سیستم را اندازه بگیرد. از آنجایی که ابعاد فضای راه حل این مسئله بسیار بزرگ است توسط نوع خاصی از الگوریتم ژنتیک به نام تعقیب تکامل حل می‌شود.

تطابق گراف دسته انعطاف پذیر:^{۲۸۹} تمام چهره های انسانی دارای ساختار توپولوژیک مشترکی هستند. چهره را می‌توان با یک گراف بازنمایی کرد که رأسهای آن روی نقاط مطمئتی قرار دارد (بینی، ... و یالهای آن با بردارهای فاصله بویعدی

²⁸⁸ evolutionary pursuit (EP)

برچسب خورده اند. هر رأس حاوی ۴۰ ضریب پیچیده موجک گابور در مقیاسها و راستاهای متفاوت است (فاز و بزرگی) که به آنها جت گفته می شود. شناسایی بر اساس گراف برچسب دار است. یک گراف برچسب دار مجموعه ای از رأسهاست که با یالها متصلند که رأسها با جت ها برچسب گذاری می شود و یالها با فاصله برچسب گذاری می شود.

روشهای کرنلی: رویه چهره یک زیرفضاست که الزاماً خطی نیست. روشهای کرنلی تعمیمی از روشهای خطی هستند. روشهای مستقیم رویه های غیرخطی برای یادگیری این رویه قابل استفاده است.

تبدیل ردگیری^{۲۹۰}: این تبدیل تعمیمی بر تبدیل رادون^{۲۹۱} است و ابزار نسبتاً جدیدی برای پردازش تصویر بشمار می رود که برای شناسایی اشیا تحت تبدیلاتی مانند دوران، جابجایی و مقیاس بندی بکار می رود. برای تولید تبدیل ردگیری باید تابعی را در راستای خط ردگیری تصویر محاسبه کرد. تبدیلات متفاوت را می توان با استفاده از توابع گوناگون از یک تصویر بدست آورد.

مدلهای فعال ظاهر: یک AAM مدل آماری مجتمعی است که مدل تنوع شکل را با مدل تنوع ظاهر در یک چارچوب شکل نرمالیزه شده ترکیب می کند. اگر شکل و ظاهر خاکستری شیء مورد نظر بتواند تقریباً به هر نمونه ای تعمیم یابد، آنگاه یک AAM حاوی یک مدل آماری است. تطبیق با یک تصویر شامل یافتن مدل‌های پارامتر اسن که تفاوت بین تصویر و مثال مدل ساخته شده که روی تصویر بازتاب می شود را کمینه می کند.

مدلهای سه بعدی تبدیل شونده^{۲۹۲}: چهره انسانی رویه ایست که ذاتاً در یم فضای سه بعدی قرار دارد. بنابراین مدل سه بعدی باید برای بازنمایی چهره و خصوصاً تنوعهای آن مانند وضعیت و نورپردازی مناسب باشد. یک روش بر مبنای مدل سه بعدی تبدیل شونده چهره وجود دارد که شکل و بافت را به صورت پارامترهای مدل، مدلسازی می کند و برای بدست آوردن این پارامترها تنها به یک تصویر از چهره نیازمند است.

چارچوب بی‌زی: یک معیار شباهت احتمالی بر مبنای این باور بی‌زی^{۲۹۳} که تفاوت شدت روشنایی تصویر از خصوصیات یک تنوع خاص در ظاهر یک نفر است. دو کلاس تصاویر چهره تعریف شده است: تنوع خود فرد و تنوع بین افراد. شباهت بین چهره ها توسط قانون بی‌زی اندازه گیری می شود.

ماشینهای بردار پشتیبان: با داشتن مجموعه نقاطی که به دو کلاس مربوط است، این ماشینها سعی می کنند ابرصفحه ای را بیابند که بیشترین کسر ممکن از یک کلاس را در یک طرف خود قرار دهد در حالی که فاصله با کلاس دیگر در طرف دیگر صفحه را بیشینه می کند. ابتدا PCA برای استخراج ویژگیهای تصاویر چهره استفاده می شود و سپس تابع جدا کننده بین هر جفت از تصاویر با SVM یاد گرفته می شود.

²⁸⁹ elastic bunch graph matching (EBGM)

²⁹⁰ Trace transform

²⁹¹ Radon transform

²⁹² 3D morphable model

²⁹³ bayesian belief

مدلهای مخفی مارکوف: مجموعه ای از مدل‌های آماری هستند که برای تشخیص خصوصیات آماری یک سیگنال استفاده می‌شوند. این مدل‌ها از دو فرایند متداخل تشکیل شده اند: ۱- یک زنجیره مارکوفی غیرقابل مشاهده با حالاتی محدود، یک ماتریس انتقال حالت احتمالی و یک توزیع احتمال حالت اولیه و ۲- یک مجموعه از تابعهای توزیع منتسب به هر حالت.

بوستینگ و راه حل‌های دسته جمعی: ایده پشت سر بوستینگ استفاده مداوم از یک یادگیرنده ضعیف روی نسخه وزن‌گزی شده ای از زیرمجموعه نمونه آموزشی است تا بتواند مجموعه ای از دسته بندی کننده های هم‌نوع خود را تعمیم دهد. با وجود اینکه هر دسته بندی کننده ممکن است فقط اندکی از گمانه زنی تصادفی بهتر عمل کند، مجموعه حاصل یک دسته بندی کنند بسیار قوی را تشکیل می‌دهد. ویولا و جونز^{۲۹۴} نخستین سیستم آشکارسازی چهره بلادرنگ را با کمک آدابوست^{۲۹۵} بنا کردند که پیشرفت بسیار بزرگی در این زمینه به شمار می‌آید.

²⁹⁴ Viola & Jones

²⁹⁵ AdaBoost

[مدل پیشنهادی]

۱. معماری مدل پیشنهادی
۲. مدل کردن وابستگیهای کواریانسی
۳. نتایج روی تصاویر طبیعی
۴. مدل کردن آشکارسازی چهره
۵. نتایج روی تصاویر آزمایشی

۴-۱- معماری مدل پیشنهادی

بسیاری از کارهای پیشین در آشکارسازی و شناسایی چهره نسبت با این نکته کدام جنبه های محرک چهره برای شناسایی اهمیت دارد بی توجه بوده اند و فرض کرده اند که معیارهای تعریف شده توسط آنها معتبر و کافی است. یک دیدگاه نظریه اطلاعاتی به مقوله به کدگذاری و دیکد کردن تصاویر چهره، اطلاعاتی در مورد محتوای تصاویر چهره بدست می دهد که به ویژگیهای محلی و عمومی شاخص آن توجه را منعطف می کند. چنین ویژگیهایی ممکن است به تصور ما ویژگیهایی مانند چشمان، بینی، لبها، و موها باشند ولی اجباراً دیدگاه نظریه اطلاعاتی از آنها استفاده نمی کند (Turk & Pentland 91).

به زبان نظریه اطلاعات هدف از آشکار سازی چهره، استخراج اطلاعات مرتبط از یک تصویر چهره است - که باید تا جای ممکن کارا انجام گیرد- و سپس مقایسه کدگذاری یک تصویر با کدگذاری مدلهایی از چهره باشد که قبلاً در پایگاه داده ذخیره سازی شده اند. یک راه ساده برای اینکار استخراج محتویات اطلاعاتی یک تصویر برای اخذ تنوع مجموعه تصاویر چهره و استفاده از آن برای کدگذاری و مقایسه انواع چهره هاست (Turk & Pentland 91).

به زبان مدلسازی بالا به پایین نیز آشکارسازی چهره فرایندی سلسله مراتبی در مغز است که از استخراج ویژگیهای ساده در لایه های پایینتر آغاز شده و در لایه های بالاتر به آشکارسازی و شناسایی محرک پیچیده چهره می پردازد. به طور خلاصه این مراحل که در فصل ۲ مفصلاً به آنها پرداخته شد را مرور می کنیم. هنگامی که تصویری روی شبکه تشکیل می شود، توسط گیرنده های نوری مخروطی تحت هموارسازی هیستوگرام یا تبدیل لگاریتمی قرار می گیرند. این تصاویر هنگامی که به قشر خاکستری مخ می رسند به بخش اولیه بینایی و لایه VI هدایت می شوند. در این لایه داده ها شروع به کدگذاری می شوند. این کدگذاری در لایه های ابتدایی ساده است و با پیشروی در طول مسیر بطنی، که به دلیل ذات شناختی مسئله در اینجا دنبال می شود، کدگذاری پیچیده تر و مجردتر می شود که دلیل این امر پیچیده تر شدن گزینندگی نورونهای موجود در این مسیر است. به اعتقاد پژوهشگران، سلولهای ساده لایه VI به تغییر روشنایی در میدان گیرندگی آن حساس هستند. مدلهایی نظیر فیلترهای گابور برای توجیه عملکرد این سلولها پیشنهاد شده است. سلولهای پیچیده لایه VI نقش جمع آوری اطلاعات فعالیت این سلولهای ساده را بر عهده دارند و آنرا در کد خود منعکس می کنند. نظریات متعددی در مورد نحوه این جمع آوری داده ها در این سلولها وجود دارد. تحلیل مؤلفه اصلی، تحلیل مؤلفه مستقل، کدگذاری پراکنده، مخلوط مقیاس گوسی، مؤلفه های گوسی و مؤلفه های کواریانس برای توجیه عملکرد این لایه پیشنهاد شده اند. بر طبق نظریه اطلاعات کدهایی که از نظر آماری کارا تر هستند، همانهایی هستند که قاعده مندیهایی داده ها را به بهترین نحو اخذ می کنند و مدلهای ارائه شده برای این لایه با استفاده از بهینه سازی معیارهایی که تصور می شود در اخذ قاعده مندیهایی موثر است و با اصول بیولوژیکی مطابقت دارد، مدلهای خود را ارائه داده اند که از جمله این معیارها می توان به همبستگی، پراکندگی، اطلاعات متقابل نورونها، استقلال آماری، و تطابق با کواریانس اشاره کرد. در کل سیستمهای بیولوژیکی زیر فشار تکاملی شدیدی هستند، می توان انگاشت که کدهایی که آنها استفاده می کنند بسیار کاراست. در اکثر مدلهای، به جای اینکه کل صحنه در نظر گرفته شود، به دلیل پیچیدگی محاسباتی به تکه های کوچک تصویر محدود می شود که بدلیل محدودیت میدان گیرندگی نورونهای واقعی، این فرض آنچنان که در نگاه اول به نظر می آید محدود کننده نیست. پس از این لایه، لایه V2 قرار دارد که حساسیت آن به زوایا، بافتها، گوشه ها و کانتورها در نظر گرفته شده است. پس از آن هم لایه V4 قرار دارد

که نقش آن تشخیص شکلهای ساده و کانتورهای بسته است. در تمامی این لایه ها اتصالات جانبی وجود دارد که وظایف مختلفی برای آنها فرض شده است که از جمله آنها می توان به تقویت و تضعیف نورونهای همسایگی، هماهنگی واریانس، و نرمال سازی با تقسیم (بر اساس پدیده فرونشانی اطرافیان) اشاره کرد. از سوی دیگر برای شبکه های عصبی چند لایه الگوریتمهای یادگیری، مسیرهای پسخور و الگوریتمهای رقابتی هم طراحی شده اند که همگی از خصوصیت های کلی مغز (که قشر بینایی نیز از آنها مستثنی نیست) الگوبرداری شده اند. ایده هایی مانند کدگذاری جمعیتی نیز در سطح یک لایه بسیار الهام بخش و سازگار با مشاهدات بیولوژیکی بوده است

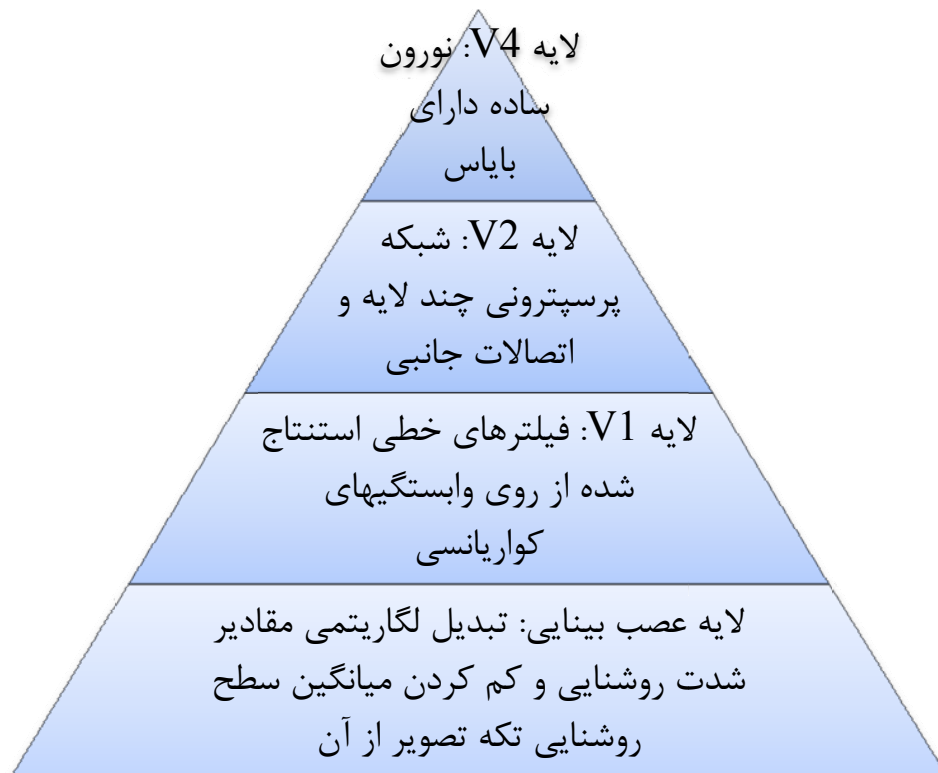
در این پژوهش سعی شده است تا با الهام از مدل های مشاهده شده در مغز، یک چارچوب برای مسئله آشکارسازی چهره معرفی شود. این مدل در قالب یک ساختار شبکه عصبی معرفی می شود و هر کدام از لایه های کلی آن جداگانه بررسی می شود. شایان ذکر است که یادگیری هر کدام از لایه های پایینتر مستقل از لایه های بالاتر صورت می گیرد و تنها در فاز آشکارسازی چهره، داده های تصویری به صورت پیشخور از لایه ای به لایه بعدی منتقل می شود. سازمان این نورونها طوری است که هر چه از لایه های پایین به بالا پیش برویم، به گزینندگی نورونها افزوده می شود، و هر نورون ورودی خود را از تعداد زیادی نورون دریافت می کند. با حفظ نقشه توپوگرافیک نورونها، همانند آن چیزی که در مغز انسان صورت می پذیرد می توان مکان دیده شدن چهره را در طول لایه ها حفظ کرد.

در لایه نخست این مدل که مشابه لایه VI است، هدف بدست آوردن یک بازنمایی جمعیتی است که بتواند ویژگیهایی مشابه لایه VI از خود نشان دهد. همانگونه که در فصل ۳ درباره کاهش افزونگی بحث شد، لایه VI به گونه ای تطبیق یافته است که قاعده مندیهایی آماری موجود در تصاویری که از طبیعت به چشم می رسد را کشف کرده، و یک کدگذاری کارا برای آن ارائه دهد. لذا برای مدلسازی این لایه به ادبیات فهم صحنه های طبیعی مراجعه شود و مدل های ملهم از بیولوژی مد نظر قرار گیرد که در فصل ۳ به تفصیل مورد بررسی قرار گرفت. روش مورد نظر باید بیشترین میزان کاهش افزونگی را داشته باشد، ساختارهای متنوع چهره های انسانی را به طور خودکار بیاموزد، و در مقابل همبستگیهای ناپایستای تصویر (که از غیرصلب بودن چهره، تغییر دیدگاه و نورپردازی که موجب پدیده هایی مانند خود-انسدادی یا خود-سایه ای می شود و ...) مقاوم باشد. برای این مدلسازی از ایده مؤلفه های کواریانس (Karklin 2008) بهره جسته می شود، از فرمولاسیون متغیرهای نهفته که در تحلیل مؤلفه مستقل (Hyvarinen 2000) بخوبی از آن استفاده شده است بهره گرفته می شود، استفاده از درستنمایی لگاریتمی که در این روش استفاده شده است مناسب به نظر می رسد (Hyvarinen 2000, Pham et al 92)، در محاسبه درستنمایی از مدل مؤلفه های واریانس (Karklin 2003a) استفاده می شود و مطابق روشهای استفاده شده در (Hyvarinen 2000, Karklin 2008) از تابع اولیه زبرگوسی لاپلاسین استفاده می شود، با توجه به نیاز به سرعت و نیز تشابه به روش یادگیری هبی (ر.ک. فصل ۲) که در شبکه های عصبی طبیعی و مصنوعی کاملاً شناخته شده است، از روش شیب نوردی برای تقریب مقدار پارامترها و متغیرهای نهفته استفاده می شود و به صورت قانون یادگیری هبی به شبکه دیکته می شود، و سپس آزمایشاتی برای اثبات مناسب بودن کد بدست آمده برای تصاویر طبیعی ارائه می شود.

در لایه دوم این مدل که مشابه لایه V2 است، از بازنمایی بدست آمده در لایه VI استفاده می شود و برای تمایز بازنمایی هایی که احتمالاً از تکه ای از تصویر چهره بدست آمده اند و بازنمایی هایی که چنین نیستند استفاده می شود. برای

این منظور از شبکه پرسپترونی چند لایه استفاده می شود که در ادبیات شبکه های عصبی بسیار شناخته شده است (ایده مشابه (Rowley 98) ولی دارای تفاوت‌های ساختاری). حاصل نورونهای موجود در این لایه یک احتمال خواهد بود که احتمال اینکه تکه تصویر از یک تصویر چهره باشد را مشخص می کند. می توان از اتصالات جانبی برای بهبود این احتمال استفاده کرد. فرض می شود تمام نورونهای این لایه در یک چینش کاشی مانند دو بعدی در کنار هم قرار گرفته اند و هر نورون با تمام همسایگان خود تا شعاعی مشخص در ارتباط است. قدرت این اتصال با یک توزیع گوسی مشخص می شود (ایده از (Rolls 99)) و هر چه فاصله اقلیدسی نورونها از هم بیشتر باشد، وزن اتصال آنها بهم کاهش می یابد. در این اتصالات جانبی با استفاده از قانون یادگیری هب میزان فعالیت نورونهای همسایه همان لایه بروز می شود.

در لایه چهارم که هم ارز لایه V4 تعبیه شده است، نورونها صرفاً از مجموع نورونهای لایه پایینتر استفاده می کنند و با داشتن یک تابع فعالسازی سخت که آنها را با مقدار بایاس مقایسه می کنند در واقع عملیات آستانه گیری را شبیه سازی می کنند. اگر میزان فعالیت نورونی در این لایه برابر یک باشد، نشانگر وجود یک تصویر چهره در میدان گیرندگی آن خواهد بود. نمایی کلی مدل نیز در شکل ۴-۱ آورده شده است.



شکل ۴-۱: نمای کلی مدل پیشنهاد شده

مقدار پارامترهای این مدل در جدول ۴-۱ آورده شده است:

پیش فرض	شرح	پارامتر
۴۰۰	ابعاد ورودی، تعداد پیکسل‌های موجود در میدان گیرندگی نورون ساده V1، پیکسل‌های تکه تصویر	I

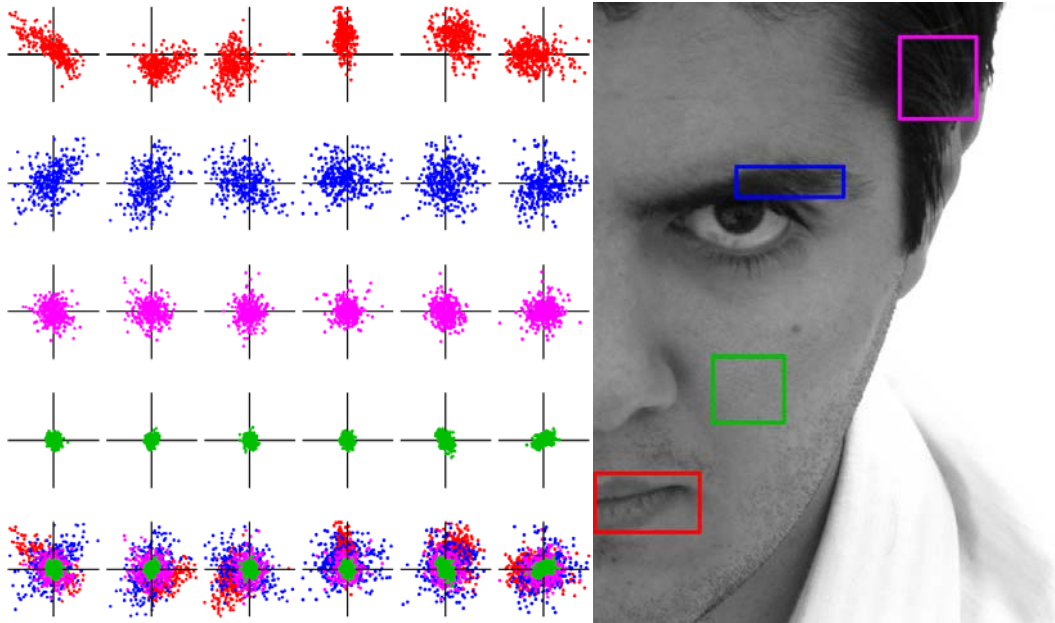
۱۵۰	تعداد نورونهای پیچیده V1، تعداد خروجی های واحد بازنمایی، طول کد حاصل از V1	J
۱۰۰۰ (فراکامل)	تعداد جهت‌هایی که انقباض و انبساط توزیع تکه تصویر در آنها سنجیده می شود	K
۱۰۰	تعداد نمونه های آموزشی در هر بسته ورودی به استنتاج	N
۷۰۵۰۰ تکه چهره ۷۰۵۰۰ غیر چهره	تعداد کل نمونه های آموزشی سیستم	N_{tot}
۵۰	تعداد گامهای مرحله استنتاج درستنمایی بیشینه	$iter_y$
۰.۰۰۵	ضریب تغییر مقدار فعالیت نورون در هر گام آموزش	ϵ_y
۰.۰۰۱	ضریب تغییر مقدار فعالیت جهت کدگذاری توزیع در هر گام آموزش	ϵ_b
۰.۰۱	ضریب تغییر مقدار وزن بردارهای کدگذاری در هر گام آموزش	ϵ_w
۰.۰۱	نرخ تخفیف وزنهاى بردارهای کدگذاری برای جلوگیری از زیاد نشدن آنها	$decay_w$
۱ (توزیع لاپلاسیان)	واریانس تابع احتمال اولیه گوسی که یک توزیع گوسی تعمیم یافته فرض شده است	var_y
۱	واریانس نمودار گوسی وزن تأثیر نورونهای همسایه بر هم در لایه V2	σ_{lat}
۳	شعاع همسایگی	R
۲۰	فاصله پیکسل آغازین میدان گیرندگی دو نورون ساده لایه V1	d_p
۴	آستانه آتش کردن نورون در لایه V4	θ
۳	درشتی میدان گیرندگی نورونهای V4	C

جدول ۴-۱: پارامترهای مدل پیشنهاد شده و مقادیر استفاده شده در پیاده سازی نمونه

۴-۲- مدل کردن وابستگیهای کواریانسی

۴-۲-۱- شرح مسئله

پیشتر به مدل‌های سلسله مراتبی برای وابستگی بین بزرگی ضرایب اشاره شد. این مدلها بر این اساس استوارند که به واریانس توزیع ضرایب اجازه تغییر در فضای داده را بدهند. این تغییرات میزان کشیدگی درستنمایی توأم در راستای محورهای توابع پایه خطی را مشخص می کند. با این حال ساختار همبستگی ضرایب خطی نیز از یک زمینه تا زمینه دیگری تغییر می کند (شکل ۴-۲) و مدل غنی تری برای اخذ این همبستگی های وابسته به زمینه و در عین حال واریانس متغیر مورد نیاز است. در اینجا تعمیمی بر مدل سلسله مراتبی کارکلین (Karklin 2003a) معرفی می شود که در آن ماتریس کواریانس کامل شامل ضرایب خطی درستنمایی توسط متغیرهای تصادفی نهفته شکل داده می شود.



شکل ۴-۲: توزیع محلی تصاویر در تصویر چهره نیز الگوهای همبستگی متفاوتی را نشان می دهد. (راست) یک تصویر چهره با چهار ناحیه متمایز مشخص شده (راست) سطرهای اول تا چهارم خروجی توأم یک جفت از فیلترهای تصویر که روی تکه های 20×20 تصویر اعمال شده است نشان می دهد. این تکه تصویرها از نواحی مشخص شده از تصویر استخراج شده اند و رنگ آنها در نمودار این تناظر را نشان می دهد. لبه ها (سطر اول) و بافتها (سطر دوم تا چهارم) دارای تنوع زیادی هستند. ویژگیهای مختلف تصویر دارای توزیعهای متفاوتی هستند ولی همانطور که در سطر پنجم مشاهده می شود این توزیعها دارای همپوشانی هستند و نمی توان از آنها مستقیماً برای تمایز بین نواحی تصویر استفاده کرد.

این روش اساساً با روش مدل وابسته به واریانس که به آن اشاره شد و روشهای مشابه آن متفاوت است. این تفاوت از آن روست که وابستگیهای بین اندازه ها یا واریانسهای ضرایب خطی تعبیه شده در هیچکدام از مدل‌های مولد، مدل‌های مبتنی بر انرژی^{۲۹۶}، مخلوط مقیاس گوسی، و مدل‌های وابستگی موقتی^{۲۹۷} به طور صریح ساختار وابستگی نایبستا داده ها را مدل نکرده است. می توان نشان داد که توزیع همبسته در بین یک مجموعه از محورهای بسادگی یک توزیع ناهمبسته ولی ناهمبسته در فضایی است که به طور مناسب دوران داده شده است (Karklin 2008). به عنوان مثال متغیرهای تصادفی گوسی با ماتریس کواریانس قطری (ایستا) D هنگامی که توسط یک تبدیل خطی A به فضای جدیدی وارد می شوند دارای کواریانس ADA^T می شوند که الزاماً قطری نیست. بنابراین مدل‌هایی که واریانس ضرایب را تغییر می دهند در همان حال همبستگی بین شدت روشنایی پیکسل‌های تصویر را نیز عوض می کنند. بهرحال هدف این پژوهش اینست که اولاً همبستگیهای مشاهده شده را بدون تکیه به تبدیلات خطی ثابت مدل کند و ثانیاً این همبستگیها را در راستاهایی که توسط یک مجموعه ثابت از توابع پایه تعریف نشده اند ارائه دهد.

ابتدا یک تابع گوسی شرطی از درست‌نمایی ضرایب خطی در نظر گرفته می شود. برای هر نمونه داده، مدل باید کواریانس درست‌نمایی ضرایب C را به عنوان تابعی از متغیرهای نهفته $\vec{\gamma}$ در نظر بگیرد:

²⁹⁶ energy-based models

²⁹⁷ temporal coherence models

$$\begin{aligned} \bar{x} | \bar{y} &\sim N(0, C) \\ C &= f(\bar{y}) \end{aligned} \quad (1-4)$$

از آنجاییکه ماتریس کواریانس مقید نیست، می تواند هر نوع تبدیل خطی را در خود جذب کند $C = AC_0A^T$. بنابراین تلاش می شود که ماتریس کواریانس داده ها (پیکسلهای تصویر)، C ، مستقیماً مدل شود گرچه مدل حاصل می تواند به توزیع ضرایب خطی نیز اعمال شود.

مسئله اصلی اکنون فرموله کردن یک توزیع اولیه سلسله مراتبی است که همان پارامتری کردن ماتریس کواریانس بر حسب متغیرهای نهفته \bar{z} است. همانند مدل واریانس، بازنمایی توزیع شده ای مورد علاقه است که به جای یاد گرفتن یک مجموعه ثابت از ماتریسهای کواریانس مختلف، ابعاد ذاتی تغییر در C را نشان دهد و این ابعاد باید در پارامترهایی که بطور خودکار از داده ها استخراج می شوند، انعکاس داده شوند. از آنجایی که این مدل یک ماتریس کواریانس را تماماً مدل می کند، تعداد پارامترهای آزاد آن از حالت واریانس بسیار بیشتر خواهد بود چنانکه برای یک ماتریس کواریانس در ابعاد $D \times D$ ، اندازه فضای کواریانسهای ممکن $D(D+1)/2$ خواهد بود. با این حال انتظار می رود که بخش اعظم این فضا برای توصیف ساختار تصویر طبیعی نامربوط باشد و تنها قسمت کوچکی از آن حاوی ساختارهای همبستگی ای باشد که در صحنه های طبیعی یا چهره ها به آنها بر می خوریم. یکی از چالشهای این مدل، فرموله کردن مدلی است که آنقدر قوی باشد که به ماتریسهای قطری محدود نباشد و آنقدر قابل توسعه باشد که بتواند با داده هایی با ابعاد بالا مانند تصاویر طبیعی نیز کار کند.

۴-۲-۲- کارهای پیشین

علاوه بر مدل مؤلفه های واریانس (Karklin 2003a) که به آن اشاره شد، مدل‌های سلسله مراتبی دیگری نیز واریانسهای نایبستا یا پارامترهای مقیاس را برای مدل کردن تصاویر طبیعی به خدمت گرفتند (Wainwright et al., 2001; Valpola et al., 2004) اما تنها یک پژوهش به همبستگیهای نایبستای درون داده های تصویری پرداخته است (Karklin 2008). با این وجود مدل کردن کواریانس و تخمین آن به زمینه های پژوهشی متعدد دیگری نظیر پردازش سیگنال، مدل‌های اقتصاد، یادگیری ماشین با روشهای مبتنی بر هسته مربوط است که در آنها به روش خود به پارامتری سازی ماتریس کواریانس می پردازند. همچنین روشهای مختلف تخمین بیزی و کاربردهای فشرده سازی که به احتمال اولیه معقولی برای کواریانس داده احتیاج دارد نیز به این مقوله پرداخته اند که نمونه هایی از آنها عبارتند از (Dempster, 1972; Wax et al., 1984; Leonard and Hsu, 1992; Daniels and Kass, 1999; Pourahmadi, 2004; Asai et al., 2006). از جمله روشهای مدل کردن کواریانس می توان تجزیه طیفی ماتریس کواریانس²⁹⁸ (Boik, 2002) و یا معکوس آن (Dempster, 1972)، فاکتورگیری از معکوس در واریانسهای جزئی و همبستگیهای جزئی معکوس (Wong et al., 2003)، عامل‌های تجزیه چولسکی²⁹⁹ (Smith and Kohn, 2002)، ماتریسهای دوران گیونز³⁰⁰ (Yang and Berger, 1994; Daniels and Kass, 1999)، تبدیلات لگاریتمی ماتریسها³⁰¹ (Leonard and Hsu, 1992; Chiu et al., 1996)، تجزیه

²⁹⁸ spectral decomposition of covariance matrix

²⁹⁹ Cholesky decomposition factors

³⁰⁰ Givens rotation matrices

³⁰¹ matrix-logarithmic transforms

چولسکی (Pourahmadi, 1999, 2000)، و تجزیه همبستگیهای خطای استاندارد (Barnard et al., 2000) را نام برد. بسیاری از این روشها در (Pourahmadi, 2004) و (Daniels and Kass, 1999) مرور شده اند.

یکی از دشواریهای پارامتری سازی ماتریس کواریانس، شرط مثبت-معین بودن ماتریس است. این شرط غالباً به شرطهای ناچوری بر روی عناصر ماتریس منتهی می شود. مسئله مهم دیگری در این زمینه قابلیت تفسیر آماری³⁰² است. خصوصاً برای این پژوهش که در نهایت از متغیرهای نهفته و بازنمایی مدل برای تحلیل راهکار احتمالی نوروها برای کدگذاری تصویر استفاده می شود، این نکته اهمیتی مضاعف می یابد.

یکی از مرتبط ترین کارهای پیشین، تخمین بیزی پارامترهای ماتریس کواریانس (Leonard and Hsu, 1992; Chiu et al., 1996) است. این روشها می کوشیدند که یک تابع احتمال اولیه معقول برای تخمین ماتریس کواریانس بدست آورند. روشی خسیسانه برای تعبیه دانش اولیه برای تخمین کارای این پارامترها لازم است. قرار دادن مستقیم احتمال اولیه روی عناصر ماتریس کواریانس در دسر ساز است زیرا ماتریس حاصل باید ماتریس مثبت-معین باشد. یک روش مستقیم استفاده از خانواده توزیعیهای ویشارت³⁰³ است که معمولاً در مدلسازی بیزی کواریانس به عنوان احتمال اولیه استفاده می شود (Evans, 1965)، ولی این توصیف بسیار محدودی است که تنها یک پارامتر (درجه آزادی) دارد و اجازه تعبیه دانش بیشتری را نمی دهد.

لئونارد و هسو³⁰⁴ یک تبدیل لگاریتمی ماتریس را پیشنهاد داده اند (Leonard and Hsu, 1992) که بسیار مناسب است زیرا هر ماتریس متقارن $A = \log(C)$ هنگامی که به توان برسد یک ماتریس مثبت-معین (یعنی یک ماتریس کواریانس معتبر) بدست می دهد.

$$A = V[\text{diag}(d_1, d_2, \dots)]V^T \quad (2-4)$$

این به توان رساندن ماتریس توسط سری زیر تعریف می شود و دارای خصوصیت مفیدی است:

$$\exp(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \quad (3-4)$$

اگر یک ماتریس متقارن را به صورت تجزیه طیفی بیان کنیم آنگاه اثر به توان رساندن ماتریس محدود به بتوان رساندن عنصر به عنصر مقادیر یکه است:

$$\exp(A) = V[\text{diag}(e^{d_1}, e^{d_2}, \dots)]V^T \quad (4-4)$$

که در آن V بردارهای یکه را جمع می کند و d_i مقادیر یکه A می باشند.

³⁰² statistical interpretability

³⁰³ Wishart family of distributions

³⁰⁴ Leonard & Hsu

تخمین پارامترها در فضای کواریانس لگاریتمی، بردارهای یگه فضای اصلی را حفظ می کند. اگر ماتریس به توان صفر برسد یک ماتریس همبستگی بدست می آید که بدین معنی است که احتمالهای اولیه به مرکز صفر در فضای لگاریتمی کواریانس، یک ماتریس کواریانس همبستگی را ترجیح می دهند. این نوع از احتمالهای اولیه توسط لئونارد و هسو در سال ۱۹۹۲ پیشنهاد شده است: عناصر بالای قطر اصلی ماتریس لگاریتمی کواریانس خطی می شوند، و یک احتمال اولیه گوسی چندمتغیره روی عناصر قرار می گیرد. چيو و همکاران^{۳۰۵} این روش را با مدل کردن عناصر توسط ترکیب خطی ماتریسهای طراحی توسعه داد (Chiu et al., 1996):

$$A = \sum_j y_j A^j \quad (۵-۴)$$

اگر تعداد ماتریسهای طراحی اندک باشد، می تواند تعداد پارامترها (و متغیرهای نهفته اگر y_i ها برای نقاط داده ای مختلف متفاوت باشند) را محدود کند ولی یک توصیف قابل تفسیر از ساختار داده را مهیا نمی کند. در ادامه مدلی بر مبنای این کار معرفی خواهد شد ولی برای اخذ الگوی همبستگی ناپیستا با استفاده از پارامترهایی که کاملاً قابل تفسیر باشند توسعه داده خواهد شد.

۴-۲-۳- تعاریف

درستنمایی داده ها: ابتدا به مدل کردن داده های پیکسلی با یک تابع توزیع احتمالی گوسی چند متغیره و یک مجموعه از توابع پایه خطی در ماتریس لگاریتمی کواریانس پرداخته می شود.

$$\begin{aligned} \bar{x} | \bar{y} &\sim N(0, C) \\ C &= \sum_j y_j A^j \end{aligned} \quad (۶-۴)$$

هر A^j یک ماتریس مربعی با اندازه ای برابر با ماتریس کواریانس است. همانند مدلهای گذشته فرض می شود که میانگین توزیع این درستنمایی $p(x|y)$ صفر است. این ساده سازی برای راحت کردن محاسبات بسیار مطلوب است و می توان در کارهای آینده به آن پرداخت ولی این روش محاسباتی (با استفاده از شیب نوردی صعودی و تابع احتمال ثانویه متغیر نهفته) نسبت به مینیموم های محلی که از دو فاکتور میانگین و واریانس توأمان ایجاد شود واکنش ضعیفی دارد. سوال دیگری که مطرح است اینست که آیا این مدل به قدر کافی انعطاف پذیر است که بتوان میانگین یک توزیع ناپیستا را به ازای هر نمونه آموزشی قرار داد؟ (این نمونه های بجای کاهش واریانس و متمرکز کردن آن در حوالی آن نقطه، کواریانس محلی را مدل می کنند).

هر ماتریس متقارن را می توان به صورت حاصلجمع ماتریسهایی با رتبه یک نمایش داد. مثلاً می توان اینکار را با انجام یک تجزیه مقدار مفرد^{۳۰۶} و سپس محاسبه یک جمع وزندار حاصل ضرب خارجی بردارهای یگه انجام داد که بردارهای یگه دارای نرم یک بوده و ضرایب آنها چیزی جز مقادیر یگه متناظر نیست.

$$A^j = \sum_k w_{jk} \vec{b}_{jk} \vec{b}_{jk}^T \quad (۷-۴)$$

انتخاب ضرب خارجی ماتریسها منجر به تفسیر پذیری قابل توجهی برای پارامترها می شود. حالتی را در نظر بگیرید که کواریانس لگاریتمی تنها از یک ماتریس $\log(C) = A^j$ تشکیل شده باشد و بردارهای مؤلفه آن \vec{b}_{jk} متعامد باشند. از آنرو که به توان رساندن ماتریس تنها روی مقادیر یگه یک ماتریس تأثیر می گذارد، تغییری در راستای مشخص شده توسط بردارها نمی دهد. بنابراین این بردارها نمایانگر جهتهایی هستند که هم در فضای اصلی و هم در فضای بتوان رسانده شده انقباض و انبساط صورت میگیرد و مقیاس در هر راستا توسط $\{e^{w_{jk}}\}$ داده می شود. با این وجود اگر مجموعه بردارها متعامد نباشند یا اگر ماتریسهای چند مؤلفه ای $\log(C) = \sum_j A^j = \sum_{jk} w_{jk} \vec{b}_{jk} \vec{b}_{jk}^T$ اضافه شدند که مؤلفه های آنها متقابلاً متعامد نبود، برهم نهی و تعامل بین بردارهای \vec{b}_k نیز باید در نظر گرفته شود. تعامل اینها با تغییر شکل بردارهای یگه دخیل در جمع نهایی است. بردارهای تقریباً متعامد جملات بینابینی^{۳۰۷} بزرگی را تولید نمی کند و ساختار طیفی ماتریس حاصل را چندان تحت تأثیر قرار نمی دهد. البته بردارهایی که کمتر متعامد هستند ساختار طیفی را طوری تغییر می دهد که جملات بینابینی را منعکس کند.

اگر از رابطه ۷-۴ مستقیماً استفاده شود و یک مجموعه متفاوت از \vec{b}_{jk} ها برای هر A^j بکار گرفته شود، عملاً فضای پارامترها کوچکتر نشده است (و همان $O(D^2)$ پارامتر را داریم). در اینجا فرض می شود که اگر تعداد محدودی از جهتها را به انحاء متفاوت با یکدیگر ترکیب کنیم می تواند خیلی از الگوهای همبستگی متداولی که در صحنه های طبیعی دیده می شود را توصیف کند. به عبارت دیگر، توزیع تصاویر با تغییر توزیع کانونی (توزیع گوسی دایروی شکل) در طول تعداد محدودی از جهتها قابل توصیف است که این تعداد می تواند بسیار بیشتر از ابعاد داده ها باشد (یعنی تعداد جهتها که K است در این بازه است $D \ll K \ll D^2$). تمام مؤلفه های کواریانس از این گروه جهتهای ممکن برای دستکاری توزیع به طور متحدالشکل بهره می برند ولی هر کدام آزادند که ترکیب خطی وزندار متفاوتی از این بردارها را استفاده کنند.

$$A^j = \sum_k w_{jk} \vec{b}_k \vec{b}_k^T \quad (۸-۴)$$

از آنجاییکه ضرایبی که A^j را ترکیب می کنند مستقل فرض شده اند، وزن w_{jk} بطور کارآمدی تغییرات همبستگی در شکل توزیع در جهتهای \vec{b}_k را اخذ می کند. به طور مثال اگر کش آمدن توزیع در یک راستا معمولاً با فشرده شدن توزیع در راستای دیگری هم وقوع است، هر دو بردار \vec{b}_k به یک z مربوط می شوند و وزنه های آنها مختلف علامه خواهد بود.

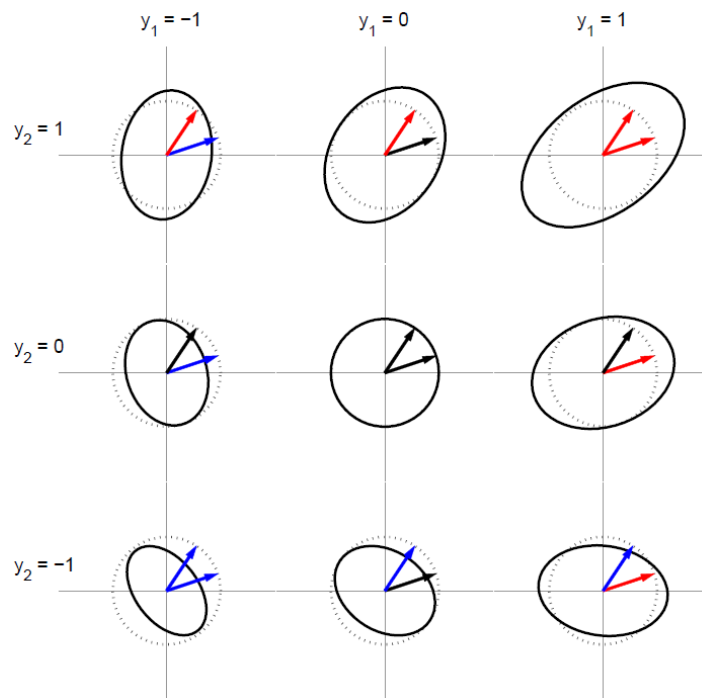
³⁰⁶ singular value decomposition

³⁰⁷ cross-terms

بنابراین ماتریس لگاریتمی کواریانس به صورت زیر تعریف می شود:

$$\log(C) = \sum_j y_j A^j = \sum_{jk} y_j w_{jk} \vec{b}_{jk} \vec{b}_{jk}^T \quad (9-4)$$

برای روشن شدن این تغییر در مدل باید افزود که وزنهای w_{jk} مورد نیاز است زیرا مطلوب است که متغیرهای y_j الگوهای همبستگی مستقل را کدگذاری کند. بردارهای \vec{b}_k توزیع را در طول یک راستای خاص دستکاری می کند و این دستکاریها (در مجموع) ممکن است همبسته باشد. بردارهای \vec{b}_k همواره طول واحد خواهند داشت $\|\vec{b}_k\| = 1, \forall k$ بنابراین وزنهای مسئول بزرگی انحراف مدل شده در کواریانس نیز هستند. یعنی هرچه بزرگی وزنهای بیشتر باشد، انحراف از توزیع کانونی بیشتر خواهد بود. علامت وزنهای جهت تأثیرگذاری را منعکس می کند و به یک y_j اجازه می دهد که افزایش تغییر در یک راستا w_{jk} مثبت) و کاهش تغییر در راستای دیگر w_{jk} منفی) را یا هم انجام دهد. از این پس در این گزارش به متغیرهای درون بردار \vec{y} ضرایب کواریانس گفته می شود. اثر ضرایب کواریانس بر روی توزیع توأم در شکل ۴-۳ به نمایش در آمده است. هنگامی که متغیر نهفته برابر صفر تنظیم شده باشد، ماتریس کواریانس برابر ماتریس همانی I است که با توزیع کانونی، گوسی دایره ای شکل متناظر است.



شکل ۴-۳: تصویری نمادین از اثر ضرایب کواریانس بر روی ساختار کواریانس داده ها. هر بیضی یک کانتر دو بعدی از توزیعی واحد است که این توزیع همان گوسی دو بعدی $p(x|y, \theta)$ می باشد و θ در بردارنده پارامترهای مدل است. این مدل ساده سازی شده از دو مؤلفه کواریانس تشکیل شده است که هر کدام با وزن $w_{jk} = I$ به تنهایی از رو بردار b_1 یا b_2 مرتبط است. یعنی ماتریس وزنهای برابر ماتریس همانی است $W = I_2$. ضرایب کواریانس مثبت (قرمز) در راستای متناظر خود در فضای تصویر منبسط می کنند و ضرایب منفی (آبی) منقبض می کنند. هنگامی که $y = 0$ ، توزیع همان توزیع گوسی پیش فرض خواهد بود (بخش مرکزی تصویر)

پیش از آموزش این مدل روی داده های تصویر، ابتدا برای پیش پردازش و ایتنینگ روی آنها اعمال می شود تا ساختار همبستگی عمومی آنها حذف شود. بنابراین کواریانس تعریف شده توسط مدل تنها انحرافات از آمارگان کلی مجموعه داده ها را منعکس می کند. این انتخاب از انتخاب مدل سلسله مراتبی واریانس (Karklin 2003a) که در حالت پیش فرض وقتی تمام ضرایب مراتب بالاتر برابر صفر است، به کواریانس AA^T دست می یابد کمی طبیعی تر است. این ماتریس الزاماً یک ماتریس همبندی نیست مگر اینکه تابعهای پایه خطی متعامد باشند و اگر A ماتریس فراکامل باشد اصلاً نمی تواند یک ماتریس همبندی باشد (البته وقتی که تعداد ابعاد زیاد است، حتی یک مجموعه بردار فراکامل نیز به تعامد نزدیک می شود). از سوی دیگر، مدل کواریانس اجازه می دهد که از یک مجموعه فراکامل از ویژگیهای خطی \vec{b}_k استفاده شود و بازهم کواریانس پیش فرض را همان ماتریس همبندی قرار می دهد و توزیع کانونی را گوسی چند متغیره دایروی در نظر می گیرد. مزیت مهم دیگر این مدل اینست که حتی وقتی که تعداد ویژگیهای خطی فراکامل است، هیچ محاسبه توزیع حاشیه ای روی ضرایب خطی انجام نمی دهد (آنها در این مدل صریحاً حساب نمی شوند). با اینحال این بدان معنا نیست که هر کدگذاری خطی و تناظر به نوروهای نسبتاً خطی مانند سلولهای ساده در محاسبات مدل مستتر باشند. درباره این موضوع بیشتر بحث خواهد شد.

درستنمایی لگاریتمی به صورت زیر تعریف می شود:

$$L = -\frac{1}{2} \log |\det C| - \frac{1}{2} \vec{x}^T C^{-1} \vec{x}$$

$$L = -\frac{1}{2} \text{Tr} \left(\sum_j y_j A^j \right) - \frac{1}{2} \vec{x}^T \exp \left(- \sum_j y_j A^j \right) \vec{x} \quad (10-4)$$

$$L = -\frac{1}{2} \sum_{jk} y_j w_{jk} - \frac{1}{2} \vec{x}^T \exp \left(- \sum_{jk} y_j w_{jk} \vec{b}_k \vec{b}_k^T \right) \vec{x}$$

و از آنجاییکه بردارهای \vec{b}_k دارای نُرم واحد هستند، در جمله دترمینان لگاریتمی جایی ندارند.

احتمال اولیه متغیر نهفته: در این مدل متغیرهای نهفته همان ضرایب \vec{y} هستند در حالی که \vec{b}_k و وزنها w_{jk} پارامترهایی هستند که باید تخمین زده شوند. همانند قبل می توان توزیع مستقل لاپلاسی را به عنوان توزیع اولیه \vec{y} در نظر گرفت:

$$\log p(\vec{y}) = \sum_j \log p(y_j) \propto - \sum_j |y_j| \quad (11-4)$$

یک چنین احتمال اولیه متقارنی به طور ضمنی فرض می کند که برای یک الگوی همبستگی داده شده، چنانکه با A^j تعریف می شود، ضرایب مثبت و منفی دارای احتمال یکسان هستند؛ و آشکارتر از آن فرض می کند که هر دوی این مجموعه ضریب های مثبت و منفی یک مجموعه بهینه از مؤلفه های کواریانس را می سازد. چنین فرضی درباره تقارن در مدل های خطی نیز وجود داشته است، ولی در این مدل ساختار تصویر که توسط دو قطب ضرایب نشان داده می شوند بسیار متفاوتند. شکل ۳-۴ این مشکل را به تصویر کشیده است: یک مؤلفه کواریانسی توزیعهای بخش راست، چپ و میانی را بر اساس علامت ضریب خود بازنمایی می کند، و مؤلفه دیگر آنرا به صورت بالا، پایین و وسط تبدیل می کند. بنابراین مدل

تقارنی را فرض می کند که ممکن است در داده ها وجود نداشته باشد. در واقع مدل هیچ راهی برای اخذ مثلاً توزیعی که در بخش میانی بالای شکل دیده می شود ندارد و بازنمایی یادگرفته شده ممکن است برای داده های آموزش بهینه نباشد.

این مشکل ممکن است در کارهای آینده با روشهای پیشرفته برای تخمین پارامترهای مدل در آینده رفع شود. ولی در این پژوهش نتایج با همین احتمال اولیه لاپلاسی (متقارن) روی ضرایب کواریانس بدست آورده شده است.

۴-۲-۴- استنتاج متغیر نهفته

گرادیان دقیق: گرادیان دقیق توسط مشتقهای جهتدار ماتریس به توان رسیده محاسبه می شود. مشتق جهتدار نسبت به ماتریس Q اینگونه تعریف می شود

$$\nabla_Q \exp(A) = V \left[(V^{-1} Q V) \odot \Phi(A) \right] V^{-1} \quad (۱۲-۴)$$

که در آن $A = V D V^{-1}$ همان تجزیه طیفی است، عملگر \odot همان ضرب عنصر به عنصر (هدامارد^{۳۰۸}) دو ماتریس است و $\Phi(A)$ متشکل از

$$[\Phi(A)]_{ij} = \begin{cases} \frac{e^{d_i} - e^{d_j}}{d_i - d_j} & i \neq j \\ e^{d_i} & i = j \end{cases} \quad (۱۳-۴)$$

است که در آن d_i همان λ_i مین عنصر ماتریس قطری D است (ر.ک. به رابطه ۱۴۴ از مقاله (Najfeld and Havel, 1995)) هر متغیر y_j ماتریس لگاریتمی کواریانس را به طور خطی توسط A^j تحت تأثیر قرار می دهد، لذا از مشتق جهتدار $\nabla_{A^j} \exp(A)$ برای بدست آوردن تابع لگاریتمی درستنمایی استفاده می شود:

$$\frac{dL}{dy_j} = \frac{1}{2} \left(\bar{x}^T \left(V \left[(V^{-1} A^j V) \odot \Phi(A) \right] V^{-1} \right) \bar{x} - \text{Tr}(A^j) \right) \quad (۱۴-۴)$$

توجه شود که این عملیات مستلزم تجزیه فرکانسی $A = \sum_j y_j A^j$ به ازای هر بروزرسانی \bar{y} است و اینکار برای ماتریسهای بزرگ بسیار کند است.

تقریب زنی با استفاده از بسط سریها: به منظور تسریع محاسبات علی الخصوص برای اجتناب از محاسبه مجدد تجزیه طیفی A در هر بار تکرار، می توان از بسط سریها برای تقریب ماتریس توانی استفاده کرد (Karklin 2008). از رابطه ۳-۴ برمی آید که دومین جمله درستنمایی لگاریتمی می تواند به صورت زیر بیان شود:

$$-\frac{1}{2}\bar{x}^T C^{-1}\bar{x} = -\frac{1}{2}\bar{x}^T \exp(-A)\bar{x} = -\frac{1}{2}\left(\bar{x}^T \bar{x} - \bar{x}^T A\bar{x} + \frac{1}{2}\bar{x}^T AA\bar{x} - \frac{1}{6}\bar{x}^T AAA\bar{x} + \dots\right) \quad (۱۵-۴)$$

اگر B را ماتریسی حاوی بردارهای \vec{b}_k و \vec{w}_j را بردار ستونی وزنه‌های $(w_{j1}, w_{j2}, \dots, w_{jK})^T$ در نظر بگیریم، می‌توان گرادینان تقریب درست‌نمایی لگاریتمی (تعریف شده توسط سری) را به صورت زیر داشت:

$$\frac{d\hat{L}}{dy_j} = \frac{1}{2}\vec{w}_j^T \left(-1 + (B^T \bar{x})^2 - (B^T \bar{x}) \odot (B^T A\bar{x}) + \frac{1}{3}(B^T \bar{x}) \odot (B^T AA\bar{x}) - \dots \right) \quad (۱۶-۴)$$

که $\vec{1}$ برداری به طول K است که تمام مؤلفه‌های آن یک هستند. این رابطه برای پیاده‌سازی در MATLAB بسیار راحت است و می‌توان حتی بدون حلقه نیز آنرا پیاده نمود. البته تقریبهای سریع دیگری نیز موجود است (Najfeld and Havel, 1995) که می‌توان در کارهای آینده به آنها پرداخت.

این شکل محاسبه استنتاج مورد نیاز در این مدل را بسیار واضح می‌کند و رابطه آن با محاسبات پیشخوری و مدل‌های کلاسیک نورونهای غیرخطی آشکار است. می‌دانیم که مدل سعی می‌کند که ماتریس کواریانس را به سمتی پیش ببرد که نقطه داده شده را پوشش دهد. استنتاج در مدل تقریباً با نگاشتن محرک بر روی ویژگیهای خطی $B\bar{x}$ ، به توان دو رساندن حاصل، و مقایسه آن با یک، یعنی عملاً رابطه $(B\bar{x})^2 - 1$ هم ارز است. سپس حاصل مقدار y_j را بر حسب وزن متناظر \vec{w}_j بالا یا پایین می‌برد. جملات مرتبه بالاتری که شامل جمع کامل درون A می‌شوند ($BA\bar{x}$ و غیره) کواریانس که باعث پدید آمدن وضعیت کنونی $\vec{1}$ شده اند را نشان می‌دهند و می‌کوشند تا تأثیر جمله مرتبه اول را خنثی کنند.

این محاسبات شبیه مدل‌های کلاسیک «انرژی» سلولهای پیچیده است که در آن محرک روی دو ویژگی نگاشت می‌شود و مقادیر حاصل به توان دو رسیده و جمع می‌شوند تا خروجی نورون را بدست دهند. در اینجا تعداد و تأثیر نسبی این نگاشتهای مختلف توسط وزنه‌های w_{jk} تعریف می‌شوند و به طور خودکار از روی داده‌ها یاد گرفته می‌شوند.

راه حل تقریب به فرم بسته: برای پیاده‌سازی روش به شکل شبکه عصبی مطلوب است که راه حل را با یک شیب نوردی صعودی تکراری با گامهای پیشخور تقریب زد. این کار بخصوص وقتی مفید است که بخواهیم استنتاج بسیار سریع انجام شود (مثلاً در کاربردهای برخط) و ارتباط مستقیم تری بین محاسبات مدل و پردازش نورونی پیشخوری سریع اطلاعات بینایی برقرار باشد. با وجود اینکه راه حل تحلیلی برای تخمین MAP از $\hat{\vec{1}}$ را نمی‌توان بدست آورد اما چند روش دیگر برای اینکار موجود است.

کواریانس لگاریتمی تک نمونه^{۳۰۹}، یک روش استفاده از تخمین تک-نمونه ای کواریانس

³⁰⁹ single sample log-covariance

$$\hat{C} = \bar{x}\bar{x}^T \quad (17-4)$$

و محاسبه مجموعه ای از ضرایب است که این تخمین را بخوبی تقریب می زند، به طوری که $\log(\hat{C}) \approx \hat{y}$. بهر حال کواریانس تک نمونه یک ماتریس مفرد است و تنها یک بردار یگه دارد با مقدار یگه غیرصفر دارد (در طول بردار \bar{x}). تمام ابعاد دیگر دارای حجمی نیستند و یک کدگذاری ایده آل باید تمام این ابعاد را در حین تطبیق با مقیاس ورودی فشرده کند. در طی استنتاج MAP، احتمال اولیه به مرکز صفر بوسیله جریمه کردن مقادیر بزرگ \bar{y} از واگرایی جلوگیری می کند.

اگر به ماتریس کواریانس تک-نمونه، یک ماتریس کواریانس یکرخت با واریانس کوچک (ϵ) بیفزاییم^{۳۱۰}،

$$\hat{C} = \bar{x}\bar{x}^T + \epsilon I \quad (18-4)$$

مقدارهای یگه این ماتریس برابر $(\|\bar{x}\|^2 + \epsilon, \epsilon, \dots, \epsilon)$ خواهند بود و بردارهای یگه آن شامل $\bar{x} / \|\bar{x}\|$ و یک مجموعه از بردارهای متعامد خواهند بود. لگاریتم ماتریس دارای بردارهای یگه مشابهی خواهند بود و مقدارهای یگه آنها عبارتست از $(\log(\|\bar{x}\|^2 + \epsilon), \log \epsilon, \dots, \log \epsilon)$. از آنجایی که مدل یک مجموعه از بردارهای پایه خطی در فضای ماتریسهای لگاریتم کواریانسی را به خدمت می گیرد، می توان مجموعه ای از ضرایب را محاسبه کرد که این ماتریس لگاریتم کواریانسی غیر مفرد را تقریب بزند:

$$\sum_j \hat{y}_j A^j \approx \log(\hat{C}) \quad (19-4)$$

ابتدا توابع پایه برداری می شود ($\vec{a}_j = \text{vec}(A^j)$)، سپس ماتریس لگاریتمی کواریانس برداری می شود و در نهایت بردار \bar{y} که بهترین تقریب از آنرا بدست می دهد تخمین زده می شود:

$$\begin{aligned} [a_1; a_2; \dots; a_j] \bar{y} &= \text{vec}(\log \hat{C}) \\ \hat{y} &= [a_1; a_2; \dots; a_j]^\dagger \text{vec}(\log \hat{C}) \end{aligned} \quad (20-4)$$

این کار به کمینه کردن خطای مربعی در فضای لگاریتمی کواریانس بستگی دارد (وزن دهی مساوی به تمام عناصر ماتریس کواریانس لگاریتمی) که ممکن است با داشتن ساختار فضا کاملاً درست نباشد چون تغییر در برخی از عناصر کواریانس لگاریتمی، درستنمایی لگاریتمی را بیش از بقیه تحت تأثیر قرار می دهد.

تقریب توان تابع، روشی دیگر است که بردارهای \vec{b}_k را تقریباً متعامد فرض کرده (در فضایی با ابعاد زیاد، آنها تقریباً متعامد فرض می شوند)، جملات بینابینی را نادیده گرفته، و با جملات درستنمایی لگاریتمی جاهایی کار داشت که توان ماتریس به صورت عنصر به عنصر ارزیابی می شود که اینکارها بجای تکیه بر محاسبه توان ماتریس انجام می شود. ماتریس قطری Z با عناصر $z_{kk} = \sum_j y_j w_{jk}$ و ماتریس B بردارهای ستونی \vec{b}_k را جمع می کند. آنگاه:

$$L = \log p(\bar{x} | \bar{y}) = -\frac{1}{2} \text{Tr}(Z) - \frac{1}{2} \bar{x}^T \exp(-BZB^T) \bar{x} \quad (21-4)$$

اگر ماتریس B متعامد یکه³¹¹ باشد (که در واقع مؤلفه های برداری آن نرمال شده هستند ولی دقیقاً متعامد نیستند) آنگاه می توان رابطه را اینگونه بازنویسی کرد:

$$L = -\frac{1}{2} \text{Tr}(Z) - \frac{1}{2} \bar{x}^T B [\text{diag}(e^{-z_{11}}, e^{-z_{22}}, \dots)] B^T \bar{x} \quad (22-4)$$

$$L = -\frac{1}{2} \sum_j y_j \sum_k w_{jk} - \frac{1}{2} \sum_k (e^{-z_{kk}} [B^T \bar{x}]_k [B^T \bar{x}]_k)$$

مشتق نسبت به y_j برابرست با

$$\frac{\partial L}{\partial y_j} \propto -\sum_k w_{jk} + \sum_k \frac{\partial z_{kk}}{\partial y_j} e^{-z_{kk}} [B^T \bar{x}]_k^2 \quad (23-4)$$

$$\frac{\partial L}{\partial y_j} \propto -\sum_k w_{jk} + \sum_k w_{jk} e^{-z_{kk}} [B^T \bar{x}]_k^2$$

با مساوی قرار دادن این رابطه با صفر رابطه زیر برای مقدار بهینه y_j بدست می آید:

$$\sum_k w_{jk} = \sum_k w_{jk} e^{-z_{kk}} [B^T \bar{x}]_k^2 \quad (24-4)$$

یکی از جوابها وقتی بدست می آید که $e^{-z_{kk}} [B^T \bar{x}]_k^2 = 1$. این منجر به تخمین زیر می شود:

$$\begin{aligned} e^{z_{kk}} &= [B^T \bar{x}]_k^2 \\ z_{kk} &= \log([B^T \bar{x}]_k^2) \\ [W\bar{y}]_k &= \log([B^T \bar{x}]_k^2) \\ \hat{\bar{y}} &= W^\dagger \log([B^T \bar{x}]_k^2) \end{aligned} \quad (25-4)$$

که در آن W ماتریس تمام وزنهای w_{jk} است و W^\dagger شبه معکوس این ماتریس است. این تقریب مدل مؤلفه های کواریانس را به فرم مدل مؤلفه های واریانس (Karklin 2003a) فرموله می کند که در آن واریانسهای لگاریتمی در راستاهای \bar{b}_k توسط یک تبدیل خطی، در اینجا W ، تعریف می شوند.

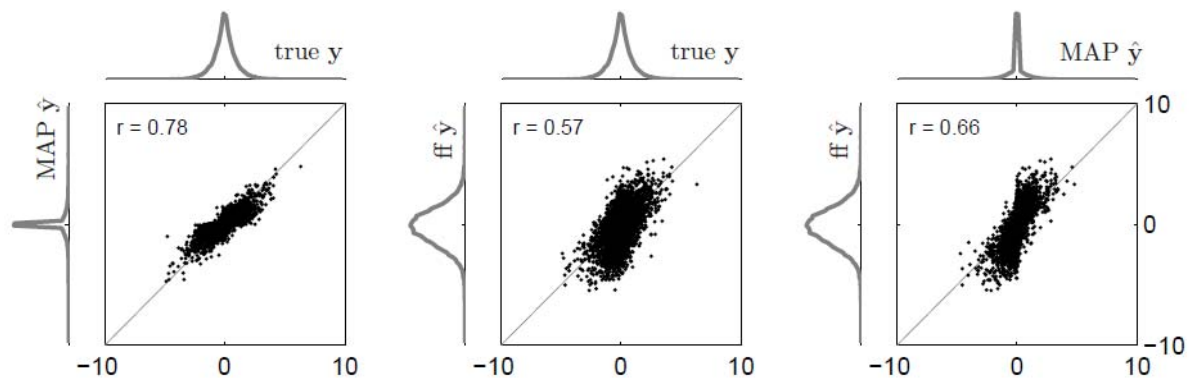
هیچ یک از این دو کدگذاری تقریبی پیشخور احتمال اولیه پراکنده روی \bar{y} را پشتیبانی نمی کنند. برای محاسبه تقریبی پیشخوری که پراکنده هم باشد، می توان روشهایی که برای مدلهای خطی با قیدهای پراکندگی استفاده می شوند را

³¹¹ orthonormal

به خدمت گرفت؛ روشهایی مانند روشهای تعقیب پایه^{۳۱۲} (Chen et al., 2001) یا تقریب زندهای کوچک کننده^{۳۱۳} (Hyvärinen, 1999). این روشها با در نظر گرفتن یک جریمه در تابع هدف برای راه حلهای ناپراکنده الگوریتم را متمایل به یافتن پاسخهای پراکنده تر می نمایند. نمونه ای از این جریمه ها در زیر آمده است:

$$L = \left\| \log([B^T \bar{x}]^2) - W\bar{y} \right\|^2 + \phi(\bar{y}) \quad (۲۶-۴)$$

تقریب زندهای بالا (بدون فید پراکنده‌گی) روی داده های ساختگی که در آن تابع مولد ضرایب معلوم است (و از توزیعهای مستقل لاپلاسین استخراج شده است) و نیز روی تصاویر طبیعی با پارامترهای مدل که با مجموعه بزرگی از تصاویر آموزش دیده است آزموده شده است (karklin 2007). تقریبهای مبتنی بر نگاهت مستقل و محاسبه واریانس لگاریتمی (رابطه ۴-۲۶)، مقادیر MAP را بسیار بهتر از تخمین کواریانس لگاریتمی تک نمونه ای (رابطه ۴-۲۰) تقریب می زند (اثبات نظری این نتایج به کارهای آینده موکول می شود. احتمالاً تقریبهای کم خطا در فضای عناصر ماتریس کواریانس لگاریتمی در ساختار مسئله منعکس نمی شود). همبستگیهای بین تخمینها و مقادیر محاسبه شده MAP در هر تکرار در شکلهای ۴-۴ و ۴-۵ آورده شده است.



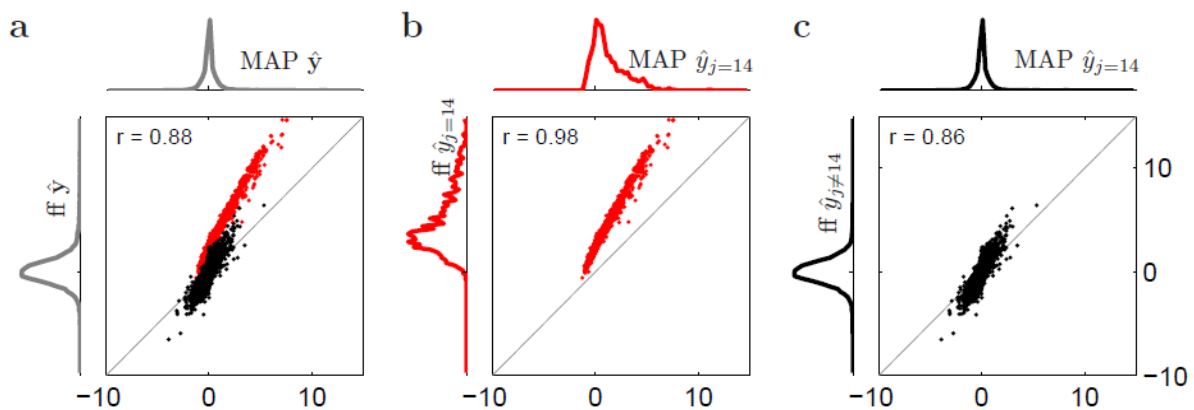
شکل ۴-۴: مقادیر متغیرهای نهفته واقعی در مقایسه با تخمینهای MAP (\hat{y}_{MAP}) و تخمینهای پیشخور آنها (\hat{y}_{ff}) در یک مجموعه داده ساختگی (که مستقیماً از مدل نمونه گیری شده است) ابعاد داده ها ۹۹، ابعاد y برابر ۲۰ و پارامترهای مدل از یک توزیع نرمال استاندارد استخراج شده اند. در شکل مقایسه تخمین MAP با مقادیر واقعی (چپ)، تقریب پیشخور با مقادیر واقعی (وسط)، و تقریب پیشخور با تخمین MAP (راست) دیده می شود. ضرایب همبستگی نیز در شکل آورده شده است. مقادیر MAP با مقادیر واقعی بسیار همبسته اند و چنانکه انتظار می رفت بسیار پراکنده اند (توزیع MAP نباید با تابع احتمال اولیه مطابق باشد زیرا مقدار بیشینه احتمال ثانویه را اخذ می کند. تخمین پیشخور با مقادیر MAP همبسته است (راست) ولی از تخمین MAP و مقادیر واقعی بسیار کمتر است.

برای داده های ساختگی حتی می توان تخمینهای MAP و مقادیر مولد \bar{y} را نیز مقایسه کرد (شکل ۴-۴ چپ). تخمین MAP بسیار خوب با مقادیر واقعی متناظر است (همبستگی آنها ۰.۷۸ است) در حالی که تناظر بین تخمین MAP و تقریب پیشخور آن کمی معتدلتر است و برابر ۰.۶۶ می باشد (شکل ۴-۴ راست). همانگونه که انتظار می رفت، توزیع مقادیر

³¹² basis pursuit methods

³¹³ shrinkage estimator

MAP از یک تابع احتمال اولیه لاپلاسی پراکنده تر است. تخمین پیشخور تقریباً گوسی است و تمرکز زیاد مقادیر MAP حول صفر را ذخیره نمی کند. الگوی مشابهی برای مدلی که روی تصاویر طبیعی آموزشی داده شده مشاهده گردید ولی ساختار پارامترهای یاد گرفته شده به مقید کردن تقریبهای تخمین پیشخور کمک می کند و در عین حال با یاس را معین می کند. به طور متوسط، تخمین پیشخور بسیار با مقادیر MAP همبسته است (شکل ۴-۵ چپ). بهرحال مقادیر MAP بدست آمده برای ضرایبی که برای کنتراست عمومی استفاده می شوند توزیع متفاوتی دارند و تخمین پیشخور این ضرایب هر چند هم که با مقدار MAP همبستگی بالایی داشته باشد، دارای مقیاس مشابهی نیستند (شکل ۴-۵ وسط). تخمین پیشخور برای مابقی ضرایب خوب است (همبستگی برابر ۰.۸۶) بنابراین نبود احتمال اولیه بدین معناست که آنها پراکنده نیستند (شکل ۴-۵ راست).



شکل ۴-۵: تقریب پیشخور تخمین MAP در مقایسه با تقریبهای MAP که از شیب نوردی صعودی روی تکه تصویرهای 10×10 بدست آمده است. از آنجایی که مؤلفه کواریانس متناظر با تغییرات عمومی (y_{14} , واحد واریانس DC) توزیع متفاوتی از مقادیر MAP نشان می دهد، با قرمز نشان داده شده است (و در شکل وسط نیز جداگانه نشان داده شده است). وقتی که مدل با داده های تصاویر طبیعی تطبیق داده می شود، تقریب پیشخور از زمانی که پارامترهای مدل تصادفی هستند بهتر (و اریب تر) است (شکل راست) زیرا بازنمایی مدل برای بازنمایی الگوهای کواریانس پراکنده شده است.

این شبیه سازیها نشان می دهد که تقریب پیشخور پیشنهاد شده تقریباً خوب عمل می کند و بیانگر اینست که این روشها وقتی استنتاج سریعی لازم است سودمند هستند. کارایی آنها می تواند با لحاظ کردن فیدهای پراکندگی و محسوب کردن وابستگی بین مؤلفه های نامتعامل کواریانس بهبود می یابد. توزیع نامتقارن ضرایب DC کواریانس نشان می دهد که احتمال اولیه لاپلاسی برای واحد کنتراست DC مناسب نیست. در مدل های خطی مانند ICA، اثر مشابهی برای ضرایب روشنایی DC^{314} نیز مشاهده شده است که آنها هم پراکنده نیستند و این مؤلفه معمولاً در آموزش مستثنی واقع می شود. بنابراین احتمالاً می توان تطبیق این مدل را با یک تابع اولیه انعطاف پذیر برای این ضرایب، و یا با اصلاح مدل برای مستثنی کردن این راستا از فضای ماتریسهای کواریانس بهبود داد.

³¹⁴ DC luminance coefficients

نگاشت استنتاج تقریبی به محاسبات نورونی: یک تقریب پیشخوری از استنتاج برای نگاشت محاسبات درون مدل به پردازش سیستم تصویری نیز سودمند است. عملیات بینایی که در این پژوهش به مدل کردن آن پرداخته شده است - پردازش سریع تصویری که در فوویا می افتد - ماهیتی سریع دارد و برای سیگنالها زمانی اندکی برای انتشار یا رفت و برگشت در نواحی بینایی قشر خاکستری مخ باقی می گذارد. مکانیسم های نورونی باید اطلاعات میدان بینایی را به صورت یک پردازش پیشخور تجمیع کنند که این اطلاعات ممکن است توسط اتصالات بازدارنده محلی دروازه بندی شوند. بر اساس رابطه ۴-۲۵ این محاسبات عبارت است از نگاشت بر روی یک مجموعه از ویژگیهای خطی $B^T \bar{x}$ که تقریباً با فعالسازی سلولهای ساده در $V1$ متناظر است (و ویژگیهای \bar{b}_k نماینده میدان گیرندگی سلولهای ساده هستند). خروجی این لایه اول نورونها سپس به توان دو می رسد و مقدار آنها با یک مقایسه می شود: هر چه مقدار بزرگتر باشد مقدار عبارت $\log([B^T \bar{x}]^2)$ نیز بزرگتر خواهد بود و برعکس هر چه کوچکتر باشد، منفی تر خواهد بود. این اطلاعات توسط نورونهای بعدی، \bar{y} ، به تناسب قدرت اتصال که با ماتریس W^T نشان داده می شود وزن گذاری شده و سپس جمع آوری می شود. این وزنها شبیه چه هستند؟ در این پیاده سازی، ماتریس W در تمام فضای ویژگیهای خطی گسترده نشده است (تعداد ابعاد کواریانس معمولاً بسیار کمتر از فضای توسعه یافته \bar{b}_k هاست) و عملگر شبه معکوس ماتریسی را می سازد که ورتهای آن به طور کیفی مشابه ماتریس اصلی است.

هر پراکنده سازی که جزئی از مدل باشد ولی در مکانیسم استنتاج آورده نشده باشد، از بازدارندگی جانبی بین نورونهای مرحله دوم نشأت می گیرد.

۴-۲-۵- تخمین پارامترها

روش تخمین بیشینه درستنمایی: پارامترهای مدل با بیشینه کردن درستنمایی داده ها بیشینه می شود. با در نظر گرفتن $\theta = \{w_{jk}, \bar{b}_k\}$ ، درستنمایی بیشینه (ML) اینگونه حساب می شود (Karklin 2003a):

$$\hat{\theta} = \arg \max \langle p(\bar{x}_n | \theta) \rangle_n \quad (۲۷-۴)$$

(که برای سهولت دنبال کردن کار، n نوشته نشده است). اکنون می توان از چارچوب بیزی استفاده کرد و احتمال اولیه این پارامترها را مشخص نمود ولی هدف در اینجا یافتن تخمینهای بیشینه است. ابتدا باید روی متغیرهای نهفته \bar{y} توزیع حاشیه را بدست آورد:

$$p(\bar{x} | \theta) = \int p(\bar{x} | \theta, \bar{y}) p(\bar{y}) d\bar{y} \quad (۲۸-۴)$$

بهرحال به خاطر شکل بسیار غیرخطی درستنمایی و نیز احتمال اولیه غیرگوسی $p(\bar{y})$ ، حل کردن این انتگرال غیرممکن است و باید همانطور که قبلاً هم اشاره شد، از مقدار آن در بیشترین حالت برای تقریب آن استفاده کرد

$$p(\bar{x} | \theta) \approx p(\bar{x} | \theta, \hat{y}) p(\hat{y}) \quad (29-4)$$

$$\hat{y} = \arg \max p(\bar{y} | \bar{x}, \theta)$$

همانند بالا، می توان بجای استفاده از تقریب از روشهای نمونه گیری مانند روش زنجیره مارکوفی مونت کارلو³¹⁵ استفاده کرد و با آنها توزیع ثانویه هم متغیرهای نهفته و هم پارامترهای مدل را تقریب زد. در عمل تخمین MAP برای بازیابی پارامترها کاراست لذا این روش جایگزین را به کارهای آینده می سپریم.

برای بدست آوردن تخمین های درستنمایی بیشینه از شیب نوردی صعودی استفاده شده است. همانند تخمین متغیرهای نهفته، می توان از تعریف رسمی گرادیان استفاده کرد که به تجزیه طیفی A در هر تکرار نیازمند است، و یا می توان از تقریب سریعتری که روی بسط سریها استوار است بهره برد. گرادیانهای تقریبی که در شبیه سازیها استفاده شده است و بر بسط سریها استوار است به شرح زیر است (اثبات در (Karklin 2007)):

$$\frac{\partial \hat{L}}{\partial b_k} = \left(\sum_j w_{jk} y_j \right) \left(I + \bar{x} \bar{x}^T - \frac{1}{2} (\bar{x} \bar{x}^T A + A \bar{x} \bar{x}^T) + \frac{1}{6} (\bar{x} \bar{x}^T A A + A A \bar{x} \bar{x}^T + A \bar{x} \bar{x}^T A) + \dots \right) b_k \quad (30-4)$$

$$\frac{\partial \hat{L}}{\partial w_{jk}} = \frac{1}{2} \left(-1 + (b_k^T \bar{x})^2 - (b_k^T \bar{x})(b_k^T A \bar{x}) + \frac{1}{3} (b_k^T \bar{x})(b_k^T A A \bar{x}) + \frac{1}{6} (b_k^T A \bar{x})^2 - \dots \right) y_j$$

برای مجموعه تصاویر، گرادیانها برای هر تکه تصویر ارزیابی می شوند و میانگین آنها گرفته می شود. اینکار درستنمایی لگاریتمی مورد انتظار روی تمام مجموعه داده ها را بیشینه می کند. پیش از هر گام شیب نوردی، بردارهای \vec{b}_k نرمال سازی می شوند تا از برقراری $\|\vec{b}_k\| = 1, \forall k$ اطمینان حاصل گردد. نرم بردار وزنهای هر متغیر نهفته، $\|w_j\|$ ، به طور تدریجی تنظیم می شود تا واریانس مطلوب γ را حفظ کند. اگر به وزنها اجازه تغییر آزادانه داده شود، تخمین MAP برای ساخت توزیع حاشیه ای متغیر نهفته به یک وضعیت منحنی منجر می شود که در آن پارامترها بزرگ و بزرگتر می شوند در حالی که واریانس متغیر نهفته کاهش می یابد.

۴-۲-۶- رابطه با مدل مؤلفه های کواریانس

با وجود اینکه این مدل به بازنمایی ماتریسهای کواریانس می پردازد، رابطه نزدیکی با مدل مؤلفه های کواریانس (Karklin 2003a) دارد و در حقیقت ایده اولیه تولید این مدل از تعمیم مدل قبل بدست آمده است. تغییر واریانس در یک بُعد اگر فضا دوران داده شده باشد، به مثابه معرفی جملات کواریانس بین هر دو بعد است. این مدلها در فضای وایتینگ شده اعمال می شوند (بنابراین توزیع کانونی دارای آمارگان $1/f$ است)، لذا واریانس متغیر در این مدل روی مقدار پیکسلها تأثیری نمی گذارد و روی جهتایی کار می کند که ترکیبات پیکسلها دیکته می کنند.

اگر مدل واریانس دارای توزیع گوسی شرطی باشد یعنی

³¹⁵ Markov chain Monte Carlo

$$p(s_i | \vec{v}) = N(0, \exp([B\vec{v}]_i)) \quad (۳۱-۴)$$

آنگاه توزیع در فضای داده ها پس از نگاشت روی توابع پایه A اینگونه خواهد بود:

$$p(\vec{x} | \vec{v}) = N(0, A \exp([B\vec{v}]A^T)) \quad (۳۲-۴)$$

که ماتریس $[e^{B\vec{v}}]$ یک ماتریس قطری است که عناصر روی قطر آن $e^{[B\vec{v}]_i}$ هستند.

بنابراین مدل مؤلفه های واریانسی، کواریانس را در فضای داده ها به شکل $C = A[e^{B\vec{v}}]A^T$ نشان می دهد که ارتباط تنگاتنگی با فرمول به توان رساندن ماتریس که در مدل کواریانس استفاده شده است دارد که $C = \exp\left(\sum_{jk} y_j w_{jk} \vec{b}_k \vec{b}_k^T\right)$. اگر مجموعه بردارهای \vec{b}_k کامل و متعامد باشد، مقادیر یکه ماتریس لگاریتمی کواریانس $A = \sum_{jk} y_j w_{jk} \vec{b}_k \vec{b}_k^T$ برابر خود \vec{b}_k می باشد و فرمول به توان رساندن ماتریس تا حد به توان رساندن عنصر به عنصر مقادیر یکه ساده می شود که $\sum_j y_j w_{jk} = W\vec{y}$. بنابراین ماتریس کواریانس که به شکل $C = B[e^{W\vec{y}}]B^T$ است دقیقاً فرمی مشابه آن چیزی است که از مدل کواریانسی بدست می آید.

این هم ارزی اگر بردارهای \vec{b}_k فراکامل یا فروکامل و نامتعامد باشند دیگر برقرار نخواهد بود. در مدل واریانس اثر یکایک توابع پایه در فضای داده جمع می شود ولی در مدل کواریانسی، اثر هر بردار \vec{b}_k تغییر بردارهای یکه ساختار A است که در نتیجه آن انقباض یا انبساط توزیع چگالی در یک راستا با به توان رسیدن مقادیر یکه کل ماتریس مشخص می شود. نتیجه اینست که مدل واریانسی، و مجموعه فراکامل توابع پایه A یک ماتریس کواریانس کانونی ($\vec{v} = 0$) برابر AA^T خواهد داشت در حالیکه در مدل کواریانسی توزیع پیش فرض همان گوسی یکرخت با $C=I$ است.

این نکته ممکن است هنگامی که ابعاد داده بالاست و \vec{b}_k ها تقریباً متعامد است چندان محسوس نباشد ولی با افزایش فراکامل بودن ویژگیهای خطی این اثر محسوستر می شود. هنگامی که یک عضو فروکامل ویژگیهای خطی استفاده می شود، مدل واریانسی تمام فضای داده را پوشش نمی دهد (ابعادی وجود خواهند داشت که واریانس در طول آنها صفر خواهد بود). از سوی دیگر مدل کواریانسی دارای پیش فرض ۱ برای تمام جهات بدون پوشش در فضاست هرچند این مدل نیز نمی تواند هیچ تغییری در این راستاها را منعکس کند.

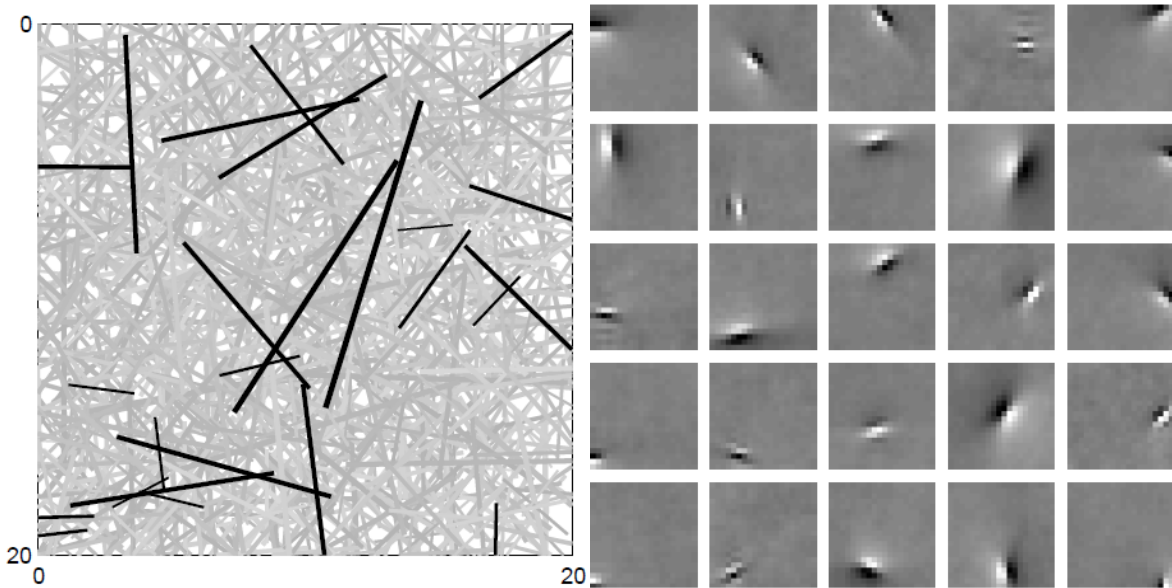
یک مزیت دیگر مدل کواریانسی اینست که به طور ضمنی توزیع حاشیه ای ضرایب خطی \vec{d} را محاسبه می کند. در مدل واریانسی امکان محاسبه توزیع حاشیه ای احتمال ثانویه $p(\vec{v} | \vec{x})$ وجود دارد. البته بجای اینکار از تخمین MAP ضرایب خطی و واریانس ضرایب مؤلفه ها استفاده می شود. یک ضعف احتمالی برای مدل مؤلفه های کواریانسی این فرض است که توزیع شرطی داده ها گوسی است در حالی که مدل واریانسی عمومی تر است و می تواند شامل توزیعهای متنوعی برای ضرایب خطی باشد و نیز می توان لگاریتم پارامتر مقیاس را مدل کرد و مقادیر مطلوب را تخمین زد.

۴-۳- نتایج روی تصاویر طبیعی

۴-۳-۱- نتایج روی تصاویر طبیعی

در این پیاده سازی، مدل با یک مجموعه بزرگ از تکه تصویرهای 20×20 که به طور تصادفی از مجموعه ۱۱۰ عکس خاکستری فام از صحنه های بیرون از منزل (Van Hateren and van der Schaaf, 1998) انتخاب شده است، آموزش داده شده است. تعداد ضرایب کواریانس $(\vec{y} = (y_1, \dots, y_r))$ در مدل برابر ۱۵۰ می باشد و تعداد ویژگیهای تصویر که برای توصیف توزیع استفاده می شود (\vec{b}_k) برابر ۱۰۰۰ است. با آزمایشهای زیاد این نتیجه حاصل شد که برای تکه ها 20×20 تصویر، مؤلفه های کواریانسی بیشتر، مجموعه ای واضح تر و تفسیرپذیرتر از این را بدست نمی دهد و حتی اندازه مؤلفه های کواریانسی زیادی هم رو به کاهش می رود. همچنین با تعداد مختلفی از \vec{b}_k آزمایش صورت گرفته است و این نتیجه حاصل شده است که یک مجموعه کامل (۴۰۰ تایی) بدترین درستی را برای داده تولید می کند در حالی که افزایش میزان فراکاملی تا بیش از $2/5$ برابر (بیش از ۱۰۰۰) بردارهای تکراری تولید می کند. در این پیاده سازی مجموعه بردارهای \vec{b}_k و وزنهای w_{jk} که نورونها را به آنها مرتبط می کنند همزمان بهینه سازی می شوند تا توزیع تصویر را بخوبی بازنمایی کنند.

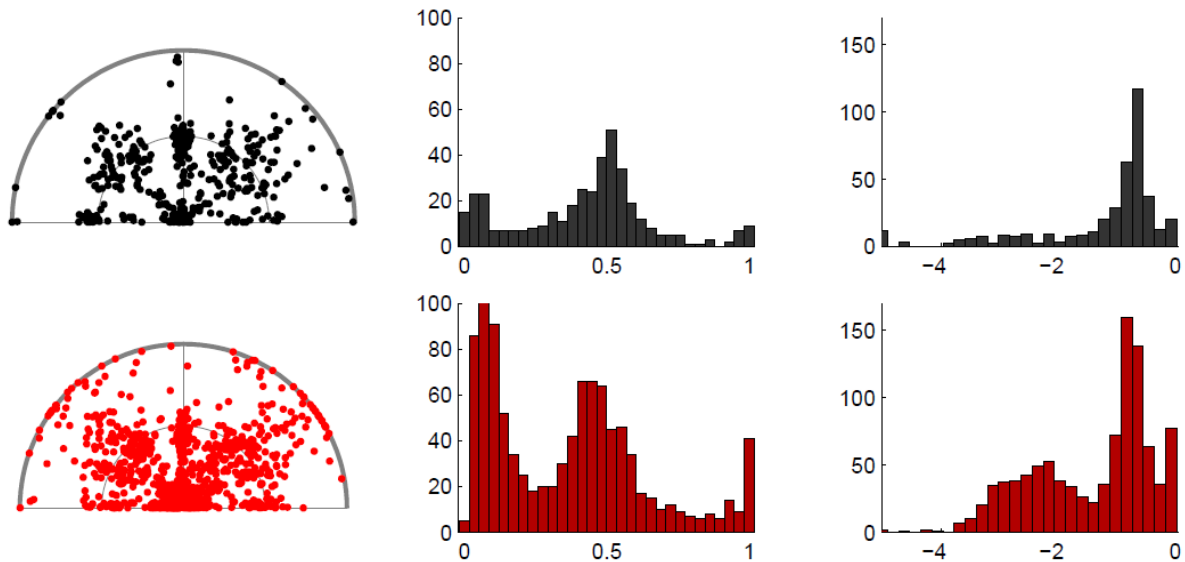
همانگونه که گفته شد مجموعه ۱۱۰ عکس خاکستری فام از صحنه های بیرون از منزل (van Hateren and van der Schaaf, 1998) به عنوان داده های آموزش استفاده می شود، که این انتخاب به طریقه سنتی و بر اساس روال مقاله های معتبری مانند (Hyvarinen 2000, Schwartz 2001, Karklin 2003a, Doi 2005, Karklin 2008) صورت گرفته است. شدت روشنایی پیکسلها ابتدا تبدیل لگاریتمی می شود (مشابه تبدیلی که در سلولهای مخروطی شبکه صورت می گیرد (van Hateren, 1997)). سپس تصاویر از یک فیلتر پایین گذر عبور داده می شود و به پایین نمونه برداری می شود تا آثار فرکانسی نمونه گیری گوشه ها را از بین ببرد. تکه تصویرهای 20×20 از همه مجموعه داده ها استخراج می شود. مقدار میانگین روشنایی از هر بسته کسر می شود (که آموزش مدل را سرعت می بخشد ولی روی پاسخ نهایی تأثیر چندانی ندارد). برای نمایش ویژگیهای تصویر، نتایج به فضای تصویر اصلی بازگردانده می شوند.



شکل ۴-۶: هنگامی که مدل با تصاویر طبیعی تطبیق می یابد، بردارهای b_k که توسط مدل برای توصیف توزیع تصویر استفاده می شوند، ویژگی‌هایی راستدار و محلی هستند. (راست) ۲۵ بردار نمونه (از ۱۰۰۰) بردار کلی به شکل تصویر در آمده اند. هر بردار الگوی همبستگی را در توزیع تصویر تغییر می دهد: پیکسل‌هایی که رنگ مشابهی دارند، مثلاً دو پیکسل سفید همبسته ترند و پیکسل‌هایی که رنگ‌هایی متضاد دارند ضد همبستگی بیشتری دارند. (چپ) مجموعه کامل ویژگی‌های تصویر که در پهنای یک تکه تصویر 20×20 پیکسل گسترده شده اند. هر خط منعکس کننده راستا، موقعیت مکانی درون تکه تصویر، و مقیاس (که با طول خط نشان داده شده است) برای یک ویژگی استخراج شده است. ویژگی‌هایی که در تصویر راست به نمایش در آمده اند در تصویر چپ با سیاه مشخص شده اند.

بردارهای \vec{b}_k جهت انبساط و انقباض کلی شکل توزیع تصویر را مشخص می کند. شکل ۴-۶ راست یک زیرمجموعه نمونه را پس از آموزش نشان می دهد. اگر به چشم یک تکه تصویر به آنها نگاه شود هر کدام یک ویژگی راستدار و لبه مانند هستند. مجموعه کامل ۱۰۰۰ تایی آنها در تمام گستره مکانی تکه تصویر پخش می شوند (شکل ۴-۶ چپ) و تمام راستاها و فرکانسهای مکانی تصاویر طبیعی را پوشش می دهند (در شکل نیامده است). این ویژگی‌های تصویری راستدار و باندگذر با تصویرهای که باعث بیشترین تحریک سلولهای ساده درون بخش اولیه قشر بینایی مغز می شود مطابقت دارد (Jones and Palmer, 1987; van Hateren and van der Schaaf, 1998; Ringach, 2002). پیش از این نیز بازنمایی های مشابهی با استفاده از مدل‌های آماری خطی بوسیله تطبیق با تصاویر طبیعی بدست آمده بود که کارایی کد بدست آمده را بیشینه می کرد (Olshausen and Field, 1996; Bell and Sejnowski, 1997). در مدلی که در اینجا پیشنهاد شده است با وجود اینکه این ویژگی‌ها صریحاً برای بازسازی تصویر اصلی استفاده نشده است ولی در عوض برای اصلاح توزیع کدگذاری شده به کار گرفته شده است (پیکانها در شکل ۴-۳). بنابراین با وجود اینکه تفاسیر قدیمی از کدهای حسگری اولیه حکایت از این داشتند که این کدها برای بازسازی منصفانه محرک سازگار می شوند، این نتایج بدست آمده تفسیر جدیدی را به آنها می افزاید: این کدها تغییرات در توزیعهای تصویر را انتقال می دهند و به ناحیه های تصویری بعدی قشر مخ امکان تشکیل بازنمایی های مجردتر را می دهند.

بررسی مجموعه ویژگیهای خطی همچنین نشان داده است که همانند مدل سلسله مراتبی واریانس (Karklin, 2003a)، خصوصیات جمعیتی این ویژگیها با جمعیتی که از مدلهای خطی مانند ICA بدست می آید متفاوت است. مجموعه کامل بردارهای \vec{b}_k دارای مؤلفه های فرکانس پایین بیشتری نسبت به توابع پایه ICA یا بردارهایی است که وقتی بدست می آید که تعداد آنها با ابعاد داده برابرست (شکل ۴-۷).



شکل ۴-۷: مقایسه بردارهای خطی b_k در یک تنظیم کامل (۴۰۰، سیاه) و فراکامل (۱۰۰۰، قرمز). ستون چپ نمودار پراکندگی فرکانس قله و راستای فیلترهای گابور که بر ویژگیهای خطی منطبق شده اند. مقیاس واحدهای شعاعی چرخه بر پیکسل است و خط ضخیم حد نایکوئیست^{۳۱۶} را نشان می دهد. ستون وسط هیستوگرام فرکانس مکانی است. ستون سمت راست هم هیستوگرام فرکانس لگاریتمی است. توجه شود که این نمودارها خصوصیات ویژگیهای b_k را نشان می دهد که شباهت زیادی به توابع پایه در ICA دارد. از آنجایی که مدل محرک را به صورت خطی کد نمی کند، برای آن فیلترهایی معادل فیلترهای مدلهای خطی نمی توان متصور شد.

این مدل یک مزیت احتمالی دیگر را نیز برای بازنمایی فراکامل فراهم می کند. واحدهای مدل توزیع تصویر را کدگذاری می کنند و برای توصیف این توزیعها به ویژگیهای خطی تکیه می کنند (شکل ۴-۳). انواع مختلف کواریانس به مجموعه متفاوتی از جهت‌های زیرین احتیاج دارند تا توزیع آنها دستکاری شود. برای مثال یک مؤلفه کواریانس که تنها محتوای فرکانسی تصویر را تغییر می دهد باید مجموعه های ویژگیهای خطی \vec{b}_k مرتب شده بر اساس فرکانسی را به طور تفاضلی تغییر دهد و تعداد کافی از ویژگیهای با فرکانس بالا و پایین باید برای آن وجود داشته باشد تا بتواند ساختار کواریانس را توصیف کند.

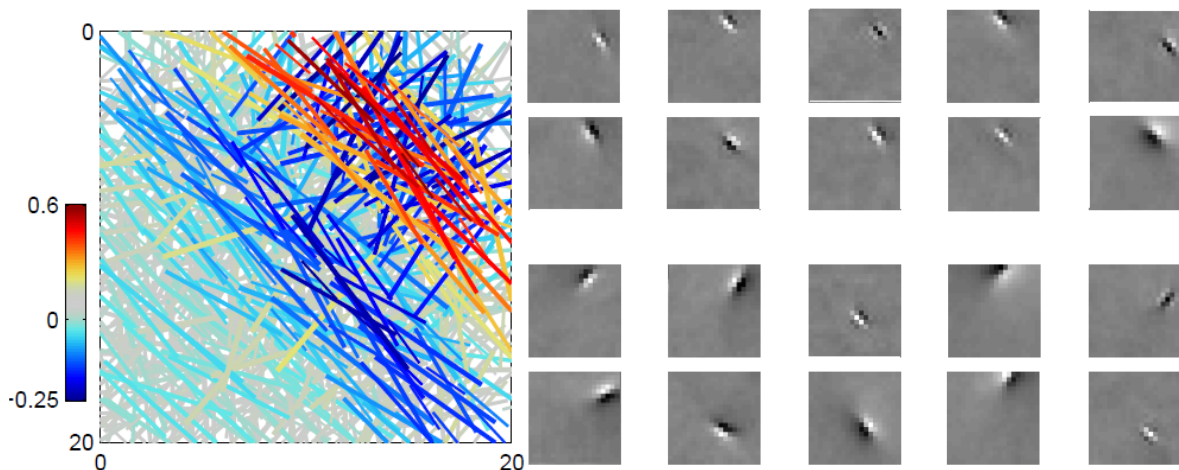
برای اینکه یک واحد دیگر بتواند تغییرات در محتوای راستهای تصویر را مستقلاً کدگذاری کند، باید به مجموعه ای از ویژگیهای خطی که تمام راستهای گوناگون را پوشش می دهند دسترسی داشته باشد. یک پایه کامل، که توسط ICA یا یک مدل سلسله مراتبی با تعداد محدودی ویژگی ساخته می شود، حاوی مجموعه کاملی از ویژگیها نیست. از نگاهی دیگر می

³¹⁶ Nyquist limit

توان دید که ویژگیهای راستادار و محلی برای کدگذاری کارای تصاویر طبیعی بهینه نیستند ولی می توانند بهترین و فشرده ترین توصیفی از توزیعهای تصویر و ساختارهای تصویری مراتب بالاتر را ارائه کنند.

۴-۳-۲- تحلیل تک تک واحدهای کواریانسی

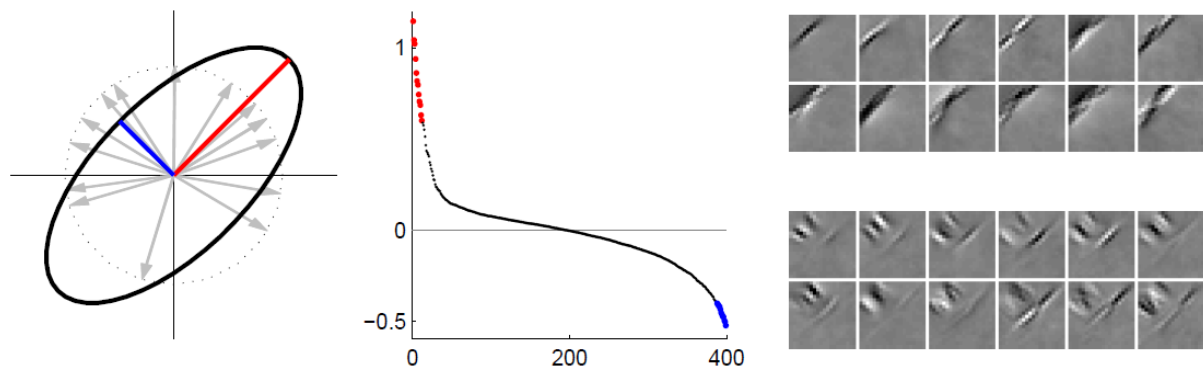
دومین مجموعه پارامترها، وزنهایی است که نورونهای مدل را به مجموعه بردارهای \vec{b}_k مربوط می کند و نقش هر نورون را در شکل دهی به توزیع کدگذاری شده تصویر معین می کند. یک مجموعه از وزنهایی آموزش دیده برای یک نورون مدل معمولی در شکل ۴-۸ راست آمده است. این نورون قویترین اثر خود را برای ویژگیهایی که در سمت راست بالای تکه تصویر قرار دارند نشان می دهد، فعالیت نورون برای آنهایی که زاویه 45° دارند (راستای «مرجح» نورون) افزایش می یابد و برای آنهایی که در راستای عمود قرار دارند و یا همراستا هستند ولی موقعیت آنها متفاوت است فعالیت کمتری دارد. بجای اینکه نورون تنها به یکسری ویژگیهای تصویری فرو نشاننده یا برانگیخته کننده پاسخ دهد، اینبار برای توصیف الگوی تنوع زیرین یک توزیع خاص تصویر، به تعداد زیادی از این ویژگیها واکنش نشان می دهد (مجموعه ای که در شکل ۴-۸ راست نمایش داده شده است). در حالی که اهمیت کاربردی این زیرواحدها در تغییر ساختار آماری توزیع کدگذاری شده است، آنها همچنان ویژگیهایی که برای بیشترین و کمترین تحریک این نورون لازم است نشان می دهند.



شکل ۴-۸: (چپ) وزنهایی یک نورون معمولی برای بردارهای b_k که در شکل ۴-۷ ترسیم شده بودند. (به ترتیب) رنگها نمایانگر علامت و بزرگی وزنهاست (به نمودار رنگی توجه شود). وزنهایی مثبت (رنگهای گرمتر) نشانگر افزایش تغییر در ویژگی متناظر هستند. وزنهایی منفی (رنگهای سردتر) نمایانگر کاهش فعالیت هستند. ویژگیهایی که توسط این نورون تغییری نمی کنند با رنگ خاکستری مشخص شده اند. (راست) بردارهای b_k متناظر با ده وزن بیشینه (سطرهای بالا) و ده وزن کمینه (سطرهای پایین) تصویر سمت راست. اینها به عنوان زیرواحدهای برانگیزاننده یا فرو نشاننده برای این نورون عمل می کنند. هنگامی که این نورون فعال شود، حضور ساختار تصویری دارای راستای خاصی در سمت بالا چپ، نبود ساختار تصویری در راستای عمود بر آن، و نبود ساختار تصویری با راستای موازی ولی با مکانی جابجا شده را مشخص می کند.

توجه شود که یک نورون مدل دارای یک محرک انگیزشی بهترین نیست بلکه این نورون برای تمام تصاویر درون توزیعی که نورون کدگذاری می کند فعال می شود. برعکس محرکهایی که در قسمتی از فضای تصویر قرار دارند که این نورون برای آنها احتمال کمی قائل است از فعال شدن آن جلوگیری می کنند.

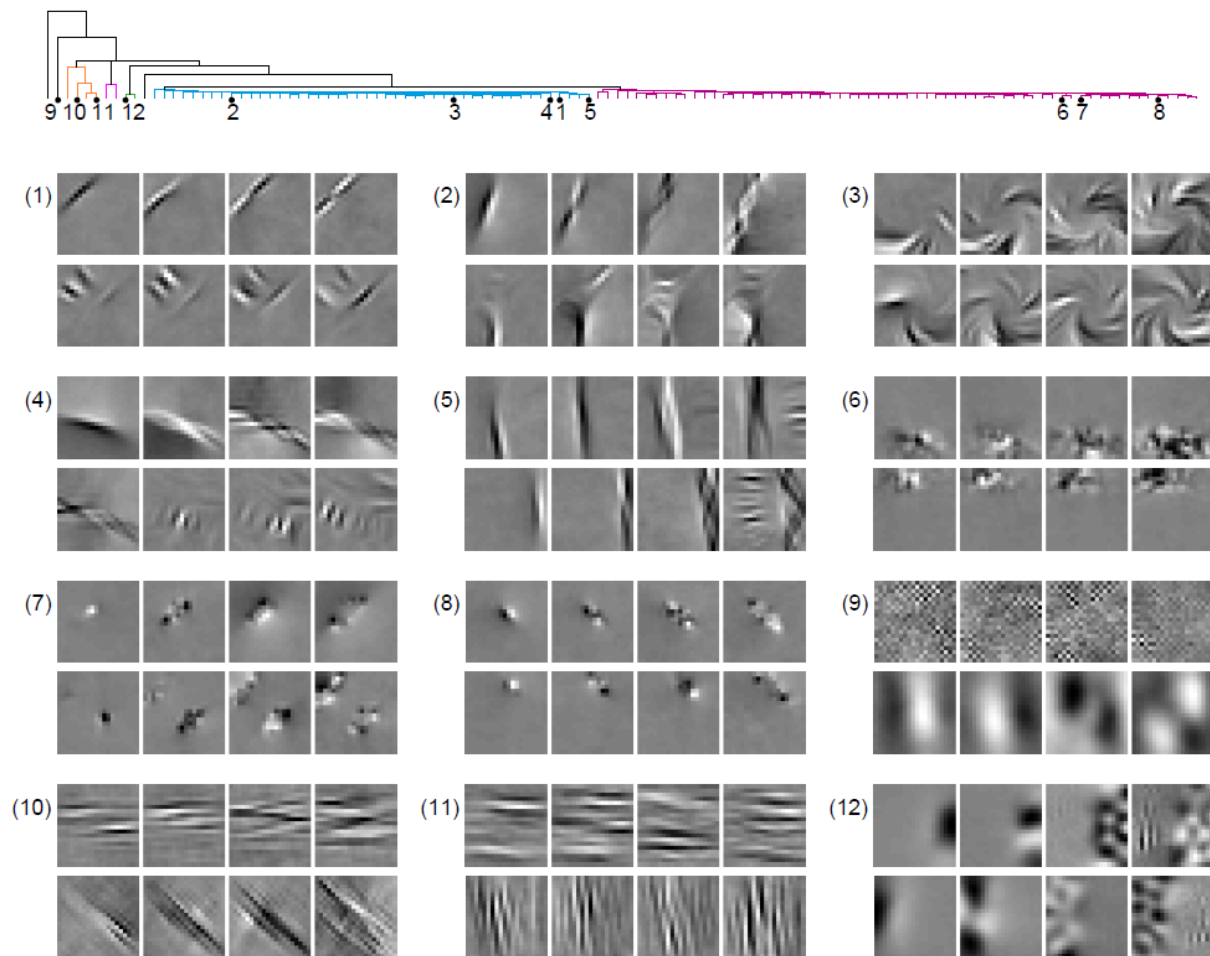
این مدل اجازه محاسبه تحلیلی راستایی که بیشترین انگیزشی و بیشترین فرونشانی برای نورون دارد را می دهد که این انگیزش و فرونشانی با توجه به قدر مطلق مقادیر z_j تعریف می شود. با ۱ قرار دادن فعالیت نورون z_j و صفر قرار دادن بقیه، ماتریس کواریانس که این نورون کدگذاری می کند پیدا می شود، $C = \exp(A^j)$. تحلیل طیفی این ماتریس نشان می دهد که ابعاد بزرگترین انقباض یا بیشترین انبساط عبارتند از بردارهای یگه متناظر با راستای بزرگترین بردارهای یگه فضای تصویر که در آن واریانس بیشترین افزایش را نسبت به توزیع کلی (وقتی که $\bar{z} = 0$ است) دارد. مقادیر یگه نشانگر اینست که توزیع فشرده شده است. این تحلیل برای یک نورون مدل در شکل ۴-۹ و برای بخشی از جمعیت در شکل ۴-۱۰ پایین آمده است.



شکل ۴-۹: (چپ) شمایی از اثر یک نورون بر توزیع کدگذاری شده. نورون از ویژگیهای زیربنایی تصویر (پیکانهای خاکستری) استفاده کرده است تا توزیع کانونی (دایره نقطه چین) را به توزیع دیگری (بیضی سیاه) مبدل سازد. تأثیر نهایی نورون بر روی توزیع از تجزیه طیفی نتایج حاصل از ماتریس کواریانس لگاریتمی بدست می آید (ر.ک. به متن) جهت‌هایی بیشترین انقباض و بیشترین انبساط به ترتیب با کوچکترین و بزرگترین مقدار یگه متناظرند (خطوط آبی و قرمز) (Karklin 2008). (وسط) مجموعه کامل ۴۰۰ مقدار یگه مقیاس همه جهت‌ها در فضای تصویر را توصیف می کند. (راست) بردارهای یگه ای که مثبت ترین و منفی ترین مقدارهای یگه را دارند به شکل تصویری به ترتیب در بالا و پایین تصویر قرار دارند. این مقدارهای یگه کرانی در تصویر میانی مشخص شده اند.

۴-۳-۳- کدگذاری جمعیتی

برای هر تصویر ورودی، فعالیت توأم جمعیتی از نورونهای مدل، توزیع تصویر استنتاج را توصیف می کند. برای فهم این کدگذاری جمعیتی باید از تحلیل خوشه ها برای شناسایی نورونهایی با عملکرد یکسان استفاده کرد (شکل ۴-۱۰ بالا) و سپس به بازنمایی نورون نماینده آنها نگاه کرد (شکل ۴-۱۰ پایین).



شکل ۴-۱۰: تحلیلی از جمعیت واحدهایی که مدل یاد گرفته است. (بالا) ۱۲۰ تا از فعالترین نورونها (از مجموع ۱۵۰ تا) که بر اساس جنبه ای از تصویر که کدگذاری می کنند به طور سلسله مراتبی خوشه بندی شده اند. این خوشه بندی دو رسته بزرگ و چند دسته خاص از نورونها را مشخص می کند (پایین) برای تهیه یک توصیف آگاهانه از هر نورون، ویژگیهایی از تصاویر که بیشتر از همه آنها را برانگیخته یا فرونشانی می کنند تشخیص داده شده اند. در این تصویر دوازده نورون نمونه از جمعیت یاد گرفته شده انتخاب شده اند. برای هر یک از نورونها چهار ویژگی تصویری برانگیخته کننده در سطر بالا و چهار ویژگی تصویری فرونشاندنده در سطر پایین آورده شده اند. اعداد مشخص شده جای نورون در دندوگرام شکل بالا را نشان می دهد.

برای خوشه بندی جمعیت نورونها، بررسی شده است که کدام پارامتر از ویژگیهای تصویری \vec{b}_k بهتر می تواند مقدار وزنه‌های هر نورون، W_{jk} ، را خوشه بندی کند. به عنوان مثال نورون شکل ۴-۸ به ساختار راستدار و محلی حساس است و لذا وزنه‌های آن در برابر ویژگیهای زیرین تصاویر (رنگها در شکل ۴-۸ چپ) در بهترین حالت توسط مکان و راستای ویژگیهای \vec{b}_k توصیف می شوند (۹۳٪ واریانس در W_{jk} ها با رگرسیون روی این دو پارامتر تعریف می شوند). برای هر نورون، یک بردار تعریف شده است که میزان مشارکت پارامتر ویژگیها (مکان، راستا، فرکانس، و تمام ترکیبات آنها) را در توصیف وزنه‌های نورونها نشان می دهد. عناصر این بردار هشت بُعدی با $null$ (به عنوان جمله بایاس در رگرسیون)، loc ، fr ، or ، $loc-fr$ ، $loc-or$ ، $fr-or$ و $loc-or-fr$ نشان داده شده است و مقدار آنها بسته به آنکه هر جمله چقدر واریانس را در $\{W_{jk}\}$ توصیف کرده است

بین ۰ تا ۱ آورده شده است. از آنجایی که رابطه بین این پارامترها و وزن‌ها غالباً غیرخطی و احتمالاً چندقله ایست، از رگرسیون چندبعدی غیرپارامتری استفاده شده است. هر مجموعه از مقادیر پارامترها در چند ظرف^{۳۱۷} با فواصل مساوی (معمولاً ۵ ظرف) تقسیم می‌شوند و از مقدار میانگین هر ظرف (آب‌ظرف برای حالت چندبعدی) برای تخمین ضرایب رگرسیون استفاده می‌شود.

پس از آنکه ضرایب رگرسیون بدست آمد، آنرا از مقادیر بیشینه والدهایشان کم می‌کنیم تا حاصل تنها بهره‌نشان (دادن واریانس) بدست آمده از افزودن جمله جدید را منعکس کند. به عنوان مثال اگر رگرسیون روی loc و or حدود ۶۰٪ واریانس را بیان می‌کند و افزودن fz تنها آنرا تا ۶۳٪ بالا می‌برد، مقدار این عنصر در بردار برابر ۰.۰۳ قرار می‌گیرد. در پایان از روش خوشه بندی سلسله مراتبی استاندارد^{۳۱۸} (الگوریتم پیوند^{۳۱۹} که به ماتریس فواصل بلوک شهری^{۳۲۰} بین بردارهای ورودی اعمال می‌شود) استفاده می‌شود که دندوگرام شکل ۴-۱۰ بالا را می‌سازد.

روشهای دیگری برای خوشه بندی موجود است از قبیل روشهایی که در تشخیص گروههای متمایز عملکردی نوروها استفاده شده است (Hegd'e and Van Essen, 2003; Gallant et al., 1996). ولی استفاده از این روش وقتی عملکرد نورو از پیش معلوم نیست و محرک غنی و پیچیده ای مانند تصاویر طبیعی موجود است، دشوار است. یک روش معمول ساختار میدان گیرندگی نورو را با توجه به راستای غالب و اندازه گیری وسعت تنظیم راستا، گستردگی ویژگیهای برانگیخته کننده و فرونشاندن، و موقعیت مکانی مورد ارزیابی قرار می‌دهد. برای مدل فعلی این روش تنها برای برخی از نوروها خوب کار می‌کند. بررسی شهودی این نکته را نشان می‌دهد که تنظیم راستا و گستردگی فرونشانی برای توصیف این جمعیت ناهمگن کافی نیستند و سایر گزینه های تحلیل نیز به انتظاری که از رفتار نورو می‌رود وابستگی زیادی دارند. با این وجود تحلیلی که از پژوهشهای فیزیولوژیکی برآمده باشد ممکن است به درد اینکار بخورد، مخصوصاً وقتی که تناظر نوروهای قشر مخ و بازنمایی مدل مورد بررسی قرار گیرد.

خوشه بندی جمعیت دو دسته بزرگ و چند دسته خاص از نوروها را مشخص می‌کند و جمعیت نوروها تعدادی از ویژگیهای مشاهده شده در سلولهای بینایی قشر مخ را نیز به نمایش می‌گزارند. یک مجموعه بزرگ از نوروهای حساس به راستا و محلی تشخیص داده شده اند (به عنوان مثال نورو شکل ۴-۸ چپ). بسیاری از آنها پدیده بازداری از فعالیت به ازای راستاهای متقاطع و نواحی اطراف که در شکل ۴-۸ نیز به آن اشاره شد نشان می‌دهند که مربوط به سلولهای V1 و V2 می‌باشد که نسبت به راستا گزینندگی دارند. این نوروها تصاویر متنوعی را کدگذاری می‌کنند، که برخی از آنها انحناء و ویژگیهای پیچیده تری از آنها را شامل می‌شود (الگوهای ۵ الی ۹).

مجموعه بزرگ دیگری از نوروها برای مشخص کردن کنتراست (انرژی) در مدل استفاده می‌شوند (الگوهای ۱۰ الی ۱۲). هر یک از آنها تنها مکان حدودی انرژی کنتراست در تصویر را مشخص می‌کند (و با مجموعه وسیعی از توزیعهای تصویر متناظر است) ولی فعالیت توأم آنها به عنوان قیودی کار می‌کند که برای تعلق به توزیع تصویر باید آنها را ارضا کند. با وجود

³¹⁷ bin

³¹⁸ standard hierarchical clustering

³¹⁹ linkage algorithm

³²⁰ city block distance

اینکه نورونهای قشری در هیچ چارچوبی برای تشخیص این کد بررسی نشده اند، ولی زیرمینه های محلی کنتراست با این مشاهدات که برخی از نورونهای قشر خاکستری به الگوهای مرتبه دوم (انرژی) در تصویر حساسند (Zhou and Baker, 1994; Mareschal and Baker, 1998b) سازگار است.

در بین نورونهای مدل برخی محتوای فرکانس مکانی تصویر را تحلیل می کند (مثلاً الگوی ۱). هنگامی که نورونهای الگوی ۱ فعال است، استنتاج می کند که تصویر ورودی از مجموعه ای از تصاویر با فرکانس مکانی خاص آمده است. توجه شود که هر فعالیت نورون در مدل می تواند مثبت یا منفی باشد. فعالیت مثبت در اینجا به معنی فرکانس بالا ساختار تصویر است در حالی که فعالیت منفی به معنای فرکانس مکانی پایین است. این نورون چیزی درباره موقعیت مکانی ساختار در تصویر یا راستای غالب آن نمی گوید و تصاویری که آنرا فعال می کنند بسیار متفاوتند و تنها قیود فرکانس مکانی را ارضا می کنند.

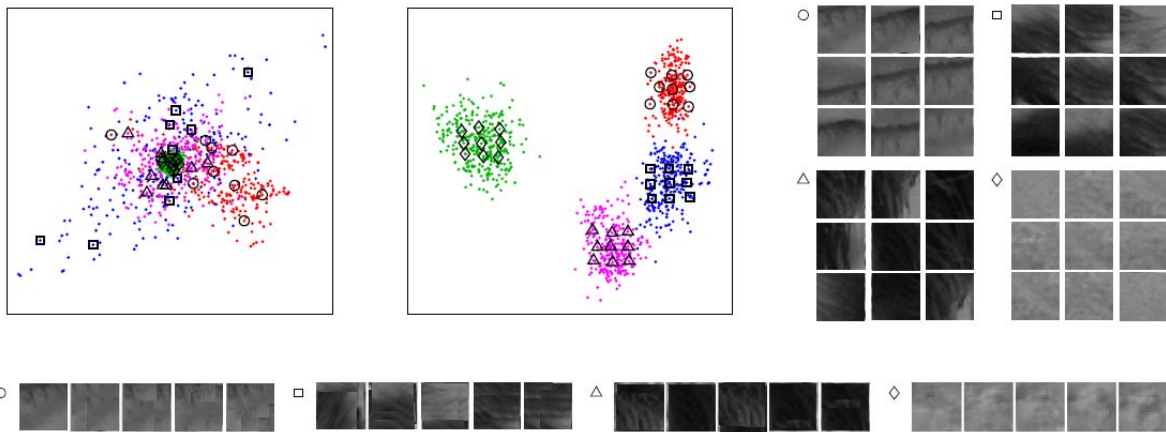
سایر نورونهای جمعیت ساختار عمومی راستا را منتقل می کنند (الگوی ۲ و ۳) ولی به محتوای فرکانس مکانی آن حساس نیستند. این خصوصیات کدگذاری در نورونهای V4 دیده شده اند که برخی از آنها کاملاً برای راستا تنظیم شده اند در حالی که برخی دیگر اطلاعات فرکانسی را کدگذاری می کنند (David 2006). سایر نورونهای مدل نشانگر کنتراست در فرکانس مکانی در موقعیتهای تصویر هستند (الگوی ۴) که مرزی برای بافتهایی مشخص می کنند که با ویژگیهای آماری آنها همخوانی دارد. مطالعات کدگذاری مرز بافتها در قشر بینایی مخ به محرکهای ساختگی ساده ای محدود شده است (Lamme, 1995; Lee et al., 1998; Nothdurft et al., 2000; Rossi et al., 2001; Song and Baker, 2007). نتایج بدست آمده در اینجا راههایی برای استفاده از بافتهای پیچیده تر پیشنهاد می کند و یک چارچوب آماری برای تحلیل پاسخ نورونها بدست می دهد.

۴-۳-۴- تعمیم و تمایز نواحی تصویر

در این قسمت نگاهی به روش استفاده مدل از جمعیت نورونها برای بازنمایی تصاویر انداخته می شود. اگر مدل می تواند روی تنوع زیادی که در تصاویر طبیعی وجود دارد تعمیم یابد، تکه های تصویر که در فضای تصویر اصلی بسیار پراکنده بودند و امکان جداسازی خطی آنها با یک کد خطی ساده موجود نبود، باید در فضای بازنمایی مدل دارای خوشه بندی تنگاتنگی باشند. این مسئله می تواند با نگاشت بازنمایی ۱۵۰-بُعدی مدل از مجموعه تصاویر روی دو بُعد (چنانکه بری فضای تصویر در شکل ۴-۲ انجام شد) روشن شود. شکل ۴-۱۱ بالا-وسط دوتا از خطی-جدایی پذیرترین ابعاد را در فضای آنرا برای چهار نوع تصویر نشان می دهد (مثالها در شکل ۴-۱۱ بالا-راست نشان داده شده اند). این محاسبات توسط روشهای تحلیل جدایی پذیری خطی (LDA) استاندارد انجام شده است که در آن فرض شده است که توزیعها گوسی چندمتغیره با کواریانسهای برابر برای هر خوشه است، S_w کواریانس درون خوشه و S_b کواریانس بین خوشه هاست و هدف محاسبه بردارهای یگه $(S_w^{-1} S_b)$ است.

تحلیل مشابهی در فضای پیکسل انجام شد و نتوانست خوشه ها را از یکدیگر جدا کند با وجود اینکه مجموعه نقاط متناظر با ابرو به نوعی جدایی پذیر است (بطور متوسط این تصاویر دارای لبه های عمودی تاریک-روشن قابل درکی است). نگاشت مجموعه آنرو دو مؤلفه اصلی اول نیز خوشه های نسبتاً جدایی پذیری بدست می دهد که مبین اینست که تفاوت

بین این خوشه‌ها در فضای بازنمایی مدل برجسته است. همانطور که مدل فرض کرده بود، با کدگذاری توزیع تصویر بجای ویژگیهای دقیقی در هر تصویر، مدل می‌تواند تصاویری که از لحاظ مفهومی به هم شبیهند را با بازنمایی شبیهی کدگذاری کند و نوعهای متمایز تصویر را از هم جدا کند.



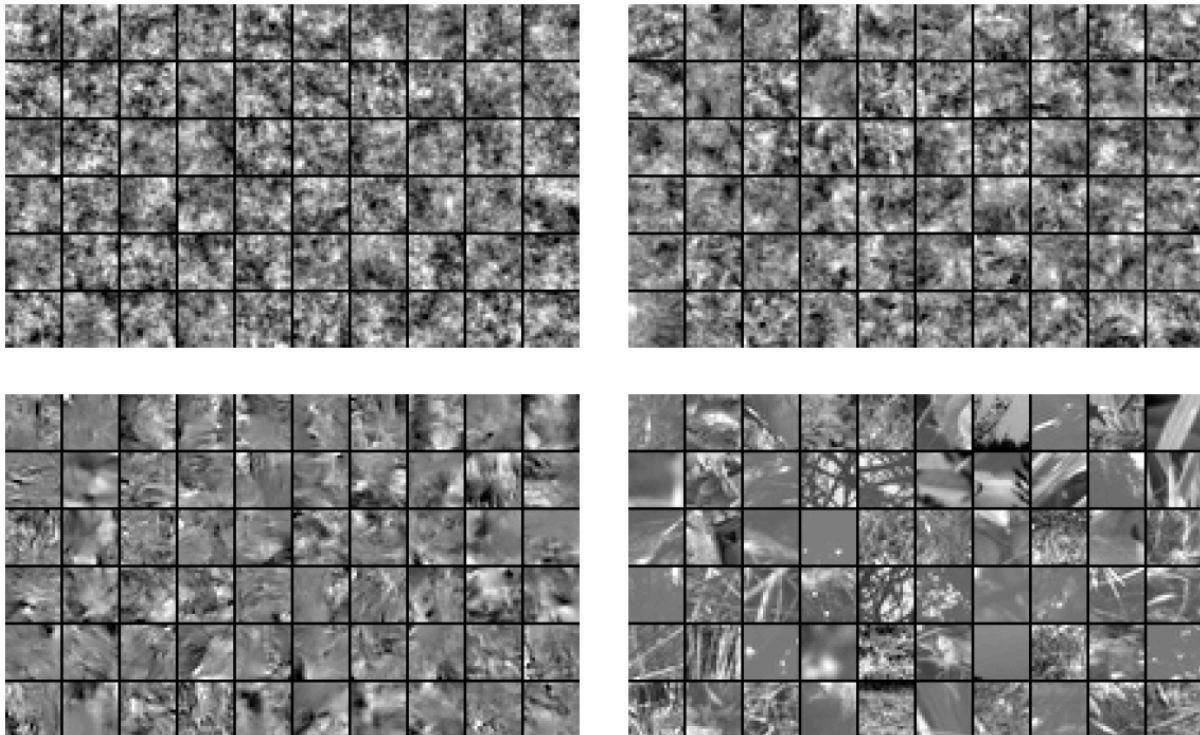
شکل ۴-۱۱: بازنمایی مدل توانسته است روی تنوع نواحی چهره تعمیم یابد و بین انواع مختلفی از انواع تصویری تمایز قائل شود. (بالا چپ) بازنمایی خطی تصاویری که از چهار ناحیه شکل ۴-۲ استخراج شده اند همپوشانی زیادی دارند (رنگها مشابه تصویر ۴-۲) (بالا وسط) یک افکنش دو بعدی از بازنمایی مدل گروه جدا پذیری از خوشه‌ها را نشان می‌دهد. (بالا راست) هر گروه 3×3 از نمادهای شکل بالا-وسط که در شکل بالا-چپ نیز حضور دارند. علیرغم تنوع در ظاهر لبه‌ها و بافتها، بازنمایی مدل در هر مدل تعمیم یافته است ولی همچنان آنها را تمایز می‌دهد. (پایین) تکه‌های تصویری که از بازنمایی مدل از هر یک از چهار نوع نمونه گیری شده اند. داده‌ها از یک گوسی چند متغیره $N(0, C(y_r))$ استخراج شده است که y_r ضرایب کواریانس برای تصویر بالا-چپ هر کدام از گروههای قسمت بالا-راست همین شکل است.

همچنین می‌توان با نمونه برداری از مدل با استفاده از ضرایب کواریانس ثابت، نوع ساختاری را که مدل کدگذاری می‌کند آزمود. میانگین تخمینهای MAP برای هر ناحیه تصویر توزیع «مرکزی» متناظر با هر نوع تصویر را بدست می‌دهد. پس باید مقدار \bar{y} را برابر با این مقدار قرار داد و تکه‌های تصویر را نمونه برداری کرد (شکل ۴-۱۱ پایین). تصاویری که از مدل استخراج شده اند دارای ساختار کواریانسی میانگین هر ناحیه از تصویر هستند - تکه‌های لب دارای تفاوت کنتراست در لبه هاست، تصاویر موها دارای ساختارهای غالب عمودی هستند، تصاویر ابرو مورب و افقی است، و تکه‌های تصویری گونه دارای کنتراست کمی هستند - ولی بسیاری از سرنخها و قاعده مندیهای آماری که نواحی را تعریف می‌کنند به وضوح توسط مدل اخذ نشده اند.

۴-۳-۵- ساخت تصاویر طبیعی

یکی از راههای مقایسه مدل با روشهای دیگر، آزمودن داده‌هایی تولید شده با نمونه گیری تصادفی از مدل است. همانطور که در فصل ۳ اشاره شد، روشهای خطی تصاویری مشابه نمونه‌های طبیعی تولید نمی‌کردند. شکل ۴-۱۲ تصاویری مه بطور

تصادفی از توزیع نمونه برداری شده اند را در کنار مجموعه ای از تصاویر طبیعی نشان می دهد (که توزیع آنها توسط مدل پیشنهادی یاد گرفته شده است). همانند قبل جمله DC کنار گذاشته می شود و توسط مدلها تولید نمی شود.



شکل ۴-۱۲: تکه های تصویری نمونه گرفته شده از مدل های مولد گوناگون. این تکه های 20×20 از مدل های PCA (چپ بالا)، ICA بدون نویز (راست بالا) و سلسله مراتبی کواریانس نایستا (مدل پیشنهادی) (پایین چپ) استخراج شده است. تصاویر آموزشی طبیعی نیز برای مقایسه آورده شده است (پایین راست)

مدل های سلسله مراتبی برخی از ساختارهای ناهمگن تصاویر طبیعی را اخذ کرده است و نواحی دارای کنتراستها، فرکانس مکانی، راستا و برخی از بافتهای جالب توجه را می سازد که شبیه چمت و پوست درخت هستند. با اینحال این مدل نتوانسته است که ویژگیهای دیگر تصاویر طبیعی مانند کانتورهای طولانی، همراستایی فاز در لبه ها، و ساختارهای با مقیاس بزرگ را اخذ کند. توجه شود که تکه های تصویر نسبتاً کوچک هستند و این همه ساختار تنها در همین ابعاد دیده شده است. بسط دادن این کار به تکه های بزرگتر تصویر و تحلیل نوع ساختارهایی که از داده ها یاد می گیرند جالب خواهد بود زیرا بسیاری از الگوهایی که برای تکه های تصویر 20×20 بدست آمده است در تکه های کوچکتر یافت نشده است. بهرحال این اندازه تکه های تصویر به بیشترین ابعادی رسیده است که با روشهای آموزش کنونی می توان محاسبات آنرا انجام داد.

۴-۴- مدل کردن آشکارسازی چهره

۴-۴-۱- شرح مسئله

روش کدگذاری مؤلفه های کواریانسی، که یکی از مدل‌هایی است که ویژگی‌های بخش بینایی مغز انسان را تا حد خوبی مدل می کند، در همین فصل ارائه شد. از خصوصیات بارز این روش می توان به ارائه یک بازنمایی مناسب برای هر تکه تصویر اشاره کرد که کد بدست آمده از آن، تکه های تصویری شبیه به هم را با کد مشابهی بازنمایی می کند. از سوی دیگر می توان به این نکته اشاره نمود که کد حاصل به ویژگی‌های کواریانسی تصاویر مربوط است و بازنمایی آن از روی آمارگان مرتبه دوم و بالاتر بدست آمده است و تنوع شدت روشنایی پیکسل (آمارگان مرتبه اول) در آن خللی وارد نمی کند. همینطور عدم وابستگی نوع کدگذاری به نمونه های خاص آزمایشی و آموزش آن با دادگان کلی امکان تطبیق بیش از حد^{۳۲۱} به مجموعه آموزشی را از بین می برد.

با استفاده از این خصوصیات کدگذاری مطرح شده می توان از آن در یک سیستم آشکارسازی چهره استفاده کرد. چنین سیستمی با تکیه بر بازنمایی تکه های تصویری، سعی بر این خواهد داشت که ابتدا میزان شباهت هر تکه تصویری را به بخشی از چهره مشخص کند، سپس با استفاده از این معیار شباهت به چهره، کاندیدهایی برای چهره بودن را تعیین کند و سپس در صدد تایید یا رد این کاندیدها بر بیاید.

۴-۴-۲- معیار شباهت به تکه ای از تصویر چهره

با توجه به توانایی بازنمایی نسبتاً جدایی پذیر مدل مؤلفه های کواریانسی، این ایده به ذهن خطور می کند که ویژگی‌های تصویر چهره در افراد مختلف نیز از این قاعده مستثنی نخواهد بود و احتمالاً در خوشه هایی مشابه قرار خواهند گرفت. از این نکته می توان برای بدست آوردن معیاری برای شباهت تکه تصویر به بخشی از چهره فرد استفاده کرد.

برای مهیا کردن چنین معیاری از یک شبکه عصبی پرسپترونی استفاده می شود. این شبکه ورودی خود را از نورونهای بازنمایی کننده مدل موله های کواریانسی می گیرد که در آن میزان فعالیت نورونها، \bar{a} ، بازنمایی نهایی تصویر را بر اساس مؤلفه های آموزش دیده w_{jk} و \bar{b}_k بدست می دهد. این مدل در لایه آخر خود دارای یک نورون است که باید میزان شباهت به چهره را در بازه (۰,۱) بدست بدهد. از آنجایی که بازنمایی بدست آمده در آزمایشات، کاملاً خطی جدا پذیر نیست نمی توان از جدا کننده خطی برای آن استفاده کرد لذا شبکه پرسپترونی به فرم چند لایه در نظر گرفته می شود و تابعهای فعالسازی نورونهای لایه ها نیز غیرخطی فرض می شود.

در این پیاده سازی شبکه پرسپترونی دارای سه لایه، با تابع فعالسازی سیگموئیدی برای لایه میانی، و تابع سیگموئیدی لاجستیک برای لایه پایانی در نظر گرفته می شود. مجموعه آموزشی بدین ترتیب مهیا می شود که مجموعه ای

³²¹ overfitting

از تصاویر برش داده شده چهره انسان (قابی مستطیل شکل از بالای موهای سر تا پایین چانه و از حد فاصل دو گوش و کمی بیرونتر از آنها) به عنوان مجموعه آموزشی مثبت، و از برخی عکسهای لبه ها، گوشه ها، بافتها و سطوح هموار و نیز سازه های مصنوعی به عنوان مجموعه آموزشی منفی استفاده می شود. لازم به ذکر است که مجموعه آموزشی مثبت مشتمل بر تصاویر چهره در اندازه ها، شرایط نورپردازی، راستا، حالات چهره، مدل‌های مختلف مو، رنگ پوست، و در برخی موارد حضور عینک (طبی) و موهای صورت (ریش و سبیل) است. در مجموعه آموزشی منفی نیز الگوهایی وجود دارد که ویژگیهای ظاهری چهره را تداعی می کنند. برای آموزش شبکه ابتدا از هر تصویر آموزشی تکه تصویرهایی 20×20 پیکسل استخراج می شود و به مدل مؤلفه های کواریانسی داده می شود تا آنرا بازنمایی کند. این مدل بازنمایی را به صورت خروجی 150 نورون به لایه اول شبکه عصبی وارد می کند. شبکه عصبی با استفاده از قانون پس انتشار خطا و 20 نورون لایه میانی خود، و با توجه به اینکه تکه تصویر از مجموعه آموزشی مثبت یا منفی بوده است، وزنه های شبکه را آموزش می دهد. تکه های تصویر غیرچهره ای که میزان فعالیت نورون لایه آخر برای آنها از آستانه 0.5 بالاتر باشد، الگوهای «شبه چهره» نامیده می شود. این الگوها در طول آموزش به طریقه بوت استرپ مکرراً به شبکه آموزش داده می شوند تا میزان فعالیت نورونها برای آنها اصلاح شود. پس از فاز آموزش وزنه های شبکه ثابت می شود و خروجی شبکه احتمال اینکه یک تکه تصویر از تصویر چهره یک انسان اخذ شده باشد را در خروجی خود نشان می دهد.

برای آزمودن برتری کدگذاری حاصل در مرحله قبل ابتدا باید به کاهش ابعاد کدگذاری حاصله از 400 بعد به 150 بعد توجه کرد که ورودی شبکه عصبی را کمتر کرده، تخمین صفحه جدا کننده در آنرا تسهیل می کند. در ضمن برای این منظور آزمایشی بدین ترتیب صورت می دهیم که سه شبکه عصبی کاملاً مشابه هم در نظر گرفته می شود که تنها در تعداد ورودی با هم متفاوتند. شبکه نخست مطابق آنچه که برای مدل توصیف شد آماده می شود. شبکه دوم با 400 ورودی، از تمام پیکسل‌های تکه تصویر که شدت روشنایی آنها بین صفر و یک نرمالسازی شده است ورودی می گیرد. برای شبکه سوم ابتدا روی کل مجموعه آموزش وایت‌نینگ صورت می گیرد و میانگین سطح روشنایی از تمام آنها کم شده و سپس مقادیر نرمالسازی شده آنها در بازه صفر و یک به شبکه داده می شود. با روش تصدیق متقاطع 10 بار تا کردن^{۳۲۲}، میانگین درصد تشخیص برای شبکه ها به ترتیب 89.2% ، 56.8% و 79.5% بوده است که نشان دهنده برتری روش پیشنهادی، به دلیل کاهش ابعاد (از دید شبکه عصبی) و اخذ وابستگیها و بالا بردن جدایی پذیری خطی (از دید لایه کدگذاری) بوده است. در این آزمایش تشخیص دادن چهره در شبکه معادل بالاتر بودن میزان فعالیت نورون لایه آخر از آستانه ای برابر 0.5 می باشد و بالعکس.

برای آزمودن میزان کارایی این معیار آزمایش زیر با الهام از (Turk & Pentland 91) ترتیب داده شده است: آزمون بازیابی تکه تصویرهای مشابه. ابتدا تکه ای 20×20 پیکسلی از تصویر یک چهره انتخاب می شود. سپس در یک تصویر که همان فرد حضور ندارد، تمام تکه های 20×20 پیکسل تصویر بازنمایی شده و ذخیره می شود. فاصله اقلیدسی بازنمایی تصویر پروب با بازنمایی تمام تصاویر موجود در گالری محاسبه شده و نزدیکترین گزینه ها مشخص می شود. با توجه به اینکه شبکه عصبی موجود به کل تصویر دسترسی ندارد و تنها بر اساس ظاهر هر تکه تصویر و بازنمایی آن در مورد آن تصمیم گیری می کند، وجود احتمالاتی بالا برای تصویر غیرچهره و احتمالاتی پایین برای چهره اجتناب ناپذیر است. شکل ۴-۱۳ میزان کارایی این روش را با توجه به آزمون بازیابی نشان می دهد.



شکل ۴-۱۳: آزمون بازیابی تکه تصویرهای مشابه (راست) تصویر بزرگنمایی شده پروب (چپ) ۵ تکه تصویر دارای کمترین فاصله اقلیدسی به بازنمایی پروب

۴-۳-۴- مدل بازنمایی های همخوان

در ادامه روند اطلاعات در مدل، اکنون داده های کدگذاری شده در لایه دوم هستند. برای روشن شدن مدل باید توجه داشت که میدان گیرندگی سلولهای عقده ای شبکه دارای همپوشانی است و بخشی از داده های بینایی یک نورون در نورون مجاور آن نیز یافت می شود. ولی باید توجه داشت که هر نورون به طریقه خود اطلاعات بینایی را کدگذاری می کند و ممکن است این کدگذاری وابسته به ویژگیهایی از تصویر باشد که در میدان گیرندگی نورون مجاور قرار نگیرد (مانند یک لبه عمودی که نورون همسایه افقی ممکت است به دلیل اندک جابجایی میدان گیرندگی خود آنرا در قاب خود نبیند). از سوی دیگر نورونهای لایه اول به فرکانس مکانی، راستا و موقعیت یک ویژگی حساسند و چنانکه مشاهده شد با اندکی دوران یا جابجایی یا به علت درشتی الگوی مورد نظر، الگو از حالت ترجیح سلولهای ساده خارج شده و حتی فرونشانی آن نیز بشود. به همین دلیل نورونهای همسایه ممکن است دارای کدزاریهایی کاملاً مشابه یا کاملاً متفاوتی باشند. از سوی دیگر برخی از الگوهای یک تصویر غیرچهره ممکن است الگویی مانند یک تکه از تصویر چهره را برای یک نورون داشته باشد (شباهت در مؤلفه فرکانسی، مکانی، راستا، یا ترکیبی از آنها) و احتمال بروز خطا در یک نورون متصور است که همانگونه که ذکر شد استفاده از دانش موجود در همسایگی هم خالی از دردسر نیست.

برای تصحیح این مشکل، از مکانیسم اتصالات جانبی در یک لایه استفاده می شود. در این مکانیسم نورونها در لایه به صورت کاشی کاری فرض می شوند و هر نورون با تمام نورونهای دیگر اتصالات مستقیم برقرار می کند. این اتصالات در یک شعاع همسایگی محدود هستند (وزن اتصالات در بیرون از این شعاع صفر فرض می شود). وزن اتصالات درون شعاع همسایگی از یک توزیع گوسی تبعیت می کند که با افزایش فاصله اقلیدسی دو نورون در نقشه توپوگرافیک وزن آنها، w_{ij}^{lat} هم کاهش می یابد:

$$w_{ij}^{lat} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\|i-j\|^2}{2\sigma^2}\right) \quad (۳۳-۴)$$

پارامترهایی که در این لایه روی نتیجه تأثیرگذار خواهند بود، واریانس توزیع گوسی همسایگی (σ^2) و فاصله میدان گیرندگی نورونهای لایه V1 از یکدیگر است. اگر واریانس توزیع همسایگی زیاد باشد، با چهره تشخیص دادن یک تکه تصویر، همسایگان بیشتری تحت تأثیر قرار خواهند گرفت ولی این تأثیر کمتر خواهد بود. بالعکس اگر واریانس توزیع همسایگی کم باشد، با چهره تشخیص دادن یک تکه تصویر، همسایگان نزدیک تحت تأثیر قرار خواهند گرفت ولی این تأثیر زیاد خواهد بود. میزان این تأثیر بر اساس قانون یادگیری هب عبارتست از:

$$y_i = y_i^{bup} + w_{ij}^{lat} y_j^{bup}, \forall j \in N_i, j \neq i \quad (۳۴-۴)$$

که در آن y_i^{bup} میزان احتمال چهره بودن بازنمایی ارائه شده از نورونهای درون میدان گیرندگی نورون i در لایه V1، و N_i همسایگی نورون i است. برای در نظر نگرفتن همسایگی می توان شعاع یا واریانس آنرا صفر در نظر گرفت و عملاً تأثیر اتصالات جانبی را حذف کرد. با توجه به استفاده از تابع سیگموئید لاجستیک برای y_i^{bup} مقدار فعالیت نورونها همواره مثبت است.

اگر فاصله میدان گیرندگی نورونها کم باشد (در کمترین حالت برابر ۱) آنگاه همپوشانی نورونها از یک تکه تصویر زیاد خواهد بود و هر نورون نسبت به نورون مجاور تغییر اندکی را تجربه می کند و اگر فاصله آنها زیاد باشد (معمولاً در بیشترین حالت برابر عرض تکه تصویر استخراج شده) آنگاه برای انتقال اطلاعات یک نورون در مورد یافتن چهره به نورونهای مجاور به اتصالات جانبی نیاز است.

۴-۴-۴ آشکارسازی چهره

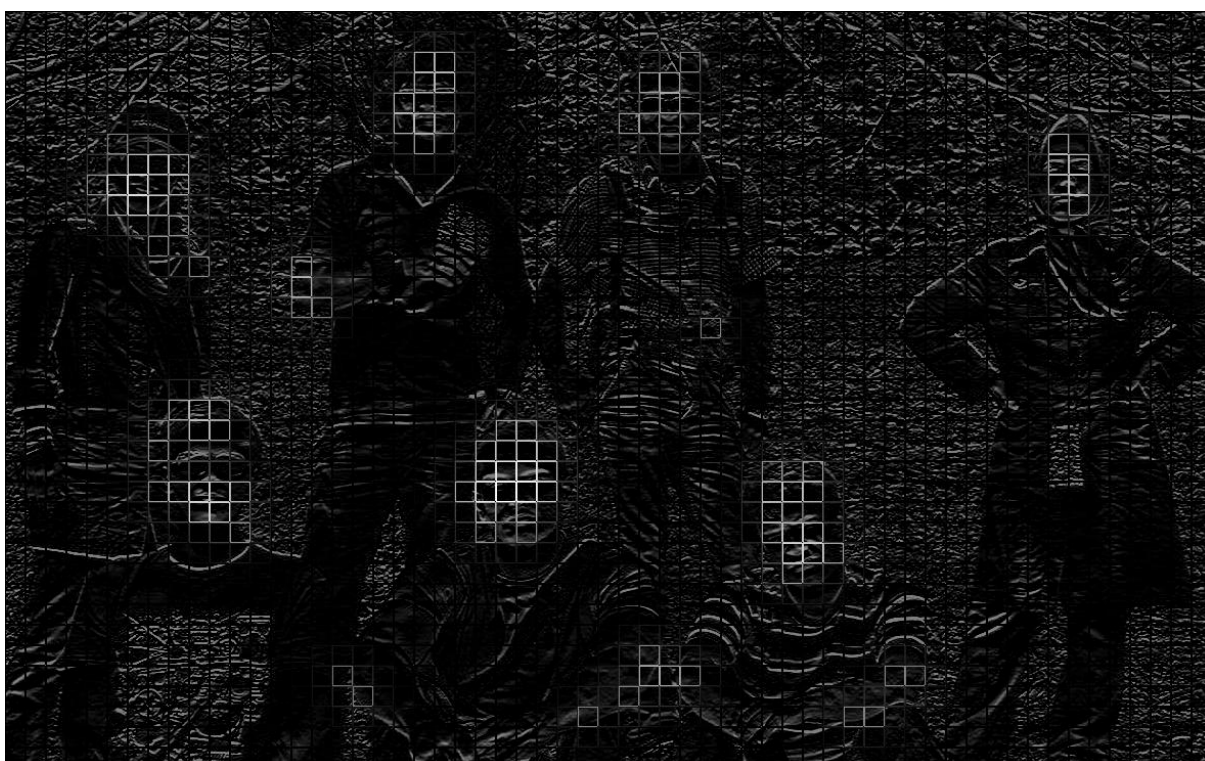
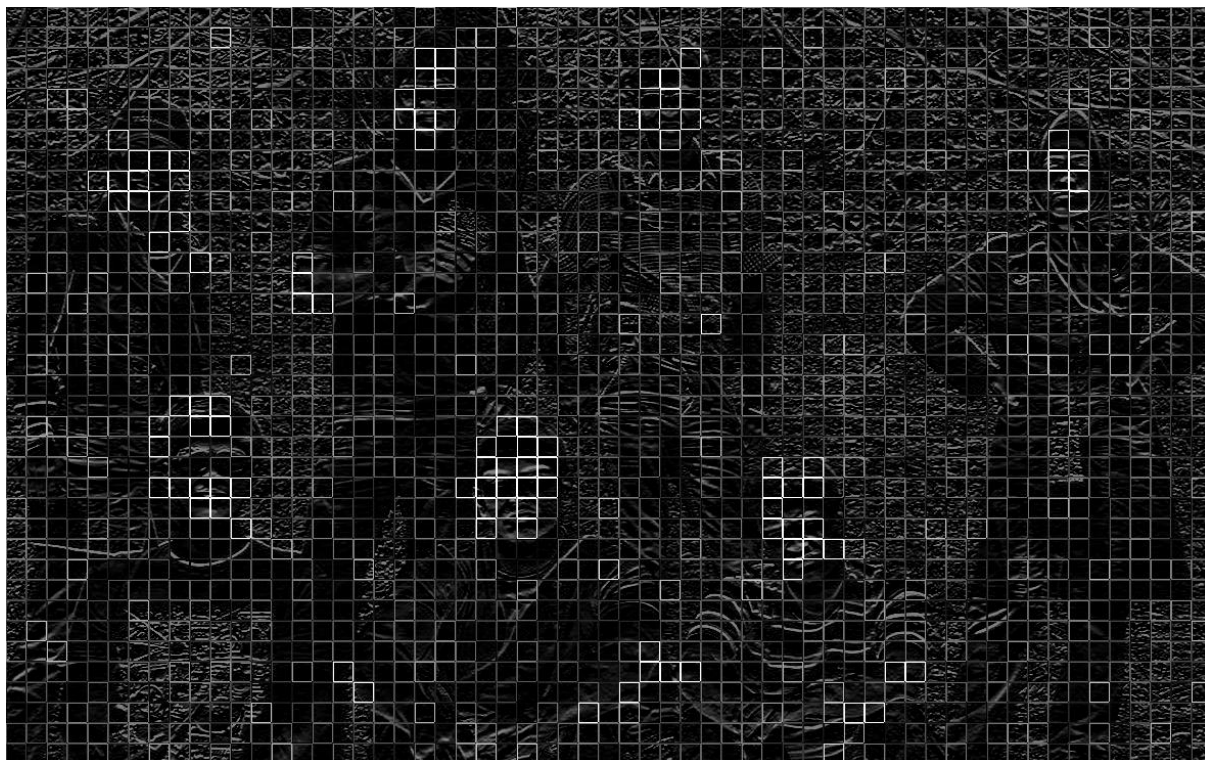
پس از پردازش داده ها در لایه V2، اکنون داده های کدگذاری شده به صورت فعالیت نورونهای این لایه هستند. با توجه به اینکه مقادیر بیشتر در این لایه، نشانگر احتمال بیشتر حضور چهره در آن می باشند، کفایت در میدان گیرندگی نورون لایه بالاتر که معادل V4 است، حوزه این مقادیر زیاد مشخص شود.

برای اینکار سه راهکار می توان در نظر گرفت: در راهکار نخست می توان از شبکه عصبی برای بیشینه گیری روی فضای نورونهای لایه پایین استفاده کرد که اینکار خوبترین کاندید برای چهره بودن را نشان می دهد. عیب این روش اینست

که تنها یک کاندید را معرفی می کند و اگر اتصالات جانبی دارای واریانس مناسبی نباشند این کاندید هم ممکن است نادرست باشد. راهکار دوم استفاده از خوشه بندی بدون ناظر برای تعیین خوشه های تصاویر چهره است که این روش دارای دو ایراد بزرگ است: بهم ریختگی نظم توپوگرافیک شبکه در لایه V4 و نیاز به تعداد خوشه ها. در ضمن اگر نواحی غیر چهره پیوسته نباشد برای قرار دادن آنها در یک خوشه مشکل بوجود می آید. راهکار سوم که در این پژوهش از آن استفاده شده است، استفاده از یک نرون ساده (مطابق مدل مک کالاک و پیتز^{۳۲۳}) با تابع فعالسازی پله، بایاس مشخص، و وزنهایی برابر یک است. چنین نرونی در میدان گیرندگی خود که توسط پارامتر درشتی، محدوده آن مشخص می شود، میزان احتمال تمام نرونها را جمع کرده با یک آستانه مقایسه می کند و اگر احتمال وجود چهره در آن ناحیه از آستانه بیشتر باشد، با داشتن خروجی یک آنرا اعلام می کنند. پارامتر درشتی برابر تعداد تکه های تصویر حاضر در یک بعد مربعی است که نرون آنرا میدان گیرندگی خود می داند و یک شدن آن برابر احتمال حضور چهره در همان مربع است. نرونها این لایه دارای همپوشانی است به طوری که تفاوت میدان گیرندگی یک نرون با نرون سمت راست آن تنها در سمت چپترین تکه های تصویری نرون سمت چپ و سمت راستترین تکه های تصویری نرون سمت راست می باشد.

مراحل آشکار سازی چهره در شکل ۴-۱۴ نمایش داده شده است. در مرحله نخست بر اساس تنظیم پارامتر همپوشانی میدان گیرندگی، از تصویر اصلی در فواصل مشخص تکه تصویرهای 20×20 پیکسلی استخراج می شود. کنتراست این تصاویر توسط ابتدا توسط تبدیل لگاریتمی تنظیم می شود. سپس میانگین سطح روشنایی از تکه تصویر کاسته می شود. سپس تصویر پیش پردازش شده به صورت 400 ورودی شدت روشنایی به لایه نخست وارد می شود. در این لایه با توجه به جهت های \vec{b}_k و وزنه های w_{jk} که از قبل آموزش دیده و ثابت شده اند، برای تهیه بازنمایی هر تکه تصویر از یک روال تکراری برای استنتاج MAP استفاده می شود که بازنمایی تکه تصویر در پایان این تکرارها در خروجی نرونها لایه اول دیده می شود. در لایه دوم این بازنمایی که توسط 150 نرون نشان داده می شود، با استفاده از شبکه پرسپترون چند لایه که وزنه های آن ثابت شده است، این بازنمایی به یک عدد تبدیل می شود که احتمال اینکه هر تکه تصویر بخشی از تصویر یک چهره باشد را نشان می دهد. این احتمال در خروجی یک نرون به ازای هر تکه تصویر یافت می شود. سپس در این لایه اتصالات جانبی بر فعالیت نرونها اثر می گزارند و مقادیر آنها را بروز می کنند. در لایه بعد بر اساس پارامتر درشتی محدوده آشکارسازی چهره، وجود چهره در هر محدوده با جمع احتمال چهره بودن تکه های تصویر درون آن و آستانه گیری از آنها مشخص می گردد. خروجی یک در این لایه نشانگر حضور چهره در محدوده چند تکه تصویر زیرمجموعه نرون فعال است.







شکل ۴-۱۴: مراحل آشکارسازی چهره توسط مدل پیشنهاد شده (مدل بازنمایی های همخوان). تصویر اول ورودی سیستم را نشان می دهد که در آن ۷ چهره از روبرو مشخص شده است. تصویر دوم پس از پیش پردازش توسط مدل سلولهای مخروطی بدست آمده است. میدان گیرندگی آنها (در این حالت بدون همپوشانی) را نشان می دهد. تصویر سوم احتمال چهره بودن هر تکه تصویر را با استفاده از لایه اول و دوم مدل بدست می دهد. قابهای دور تکه های تصویر احتمال را به تصویر کشیده است که قابهای روشنتر احتمال بیشتر را نشان می دهند (قاب کاملاً سفید=احتمال یک). در تصویر چهارم تأثیر اتصالات جانبی روی احتمالات مشخص می شود. در این تصویر شعاع همسایگی ۳ در نظر گرفته شده و واریانس توزیع گوسی همسایگی یک فرض شده است. احتمالها پس از اعمال اتصالات جانبی، با تقسیم بر بزرگترین مقدار نرمال سازی شده و نمایش داده شده اند. در تصویر پنجم تأثیر لایه سوم مدل با آستانه ای برابر ۴ نشان داده شده است. در این لایه درشتی میدان گیرندگی برابر ۳ در نظر گرفته شده است که بدین معناست که هر نورون در این لایه از ۹ نورون در لایه پایینتر ورودی می گیرد. در این تصویر نتیجه مدل مشخص شده است.

۴-۵- نتایج روی تصاویر آزمایشی

برای آزمون این روش سلسله مراتبی چند آزمایش ترتیب داده می شود. در این آزمایشات ابتدا تأثیر تغییرات پارامترها (واریانس اتصالات جانبی، همپوشانی میدانهای گیرندگی اولیه، آستانه تایید کاندید چهره، و درشتی میدان گیرندگی نورونهای سطح بالا) روی عملکرد مدل برای تشخیص چهره در تصاویر ساده چهره/غیر چهره آزموده می شود. سپس این روش بر روی تصاویر دارای نورپردازیهای متفاوت، چند چهره، شلوغ، و دارای چهره های غیر کامل در اثر انسداد، و حالت چهره آزمایش می شود. در پایان تأثیر رنگ پوست روی این مدل بررسی گردید. در انجام این آزمایشها سرعت بسیار کمی بدست آمد. البته

بدلیل استفاده از شبکه عصبی می توان تمام محاسبات درون یک لایه را بطور موازی انجام داد که سرعت را بطور چشمگیری افزایش می دهد، پیاده سازی موازی این سیستم به کارهای بعده موکول می شود.

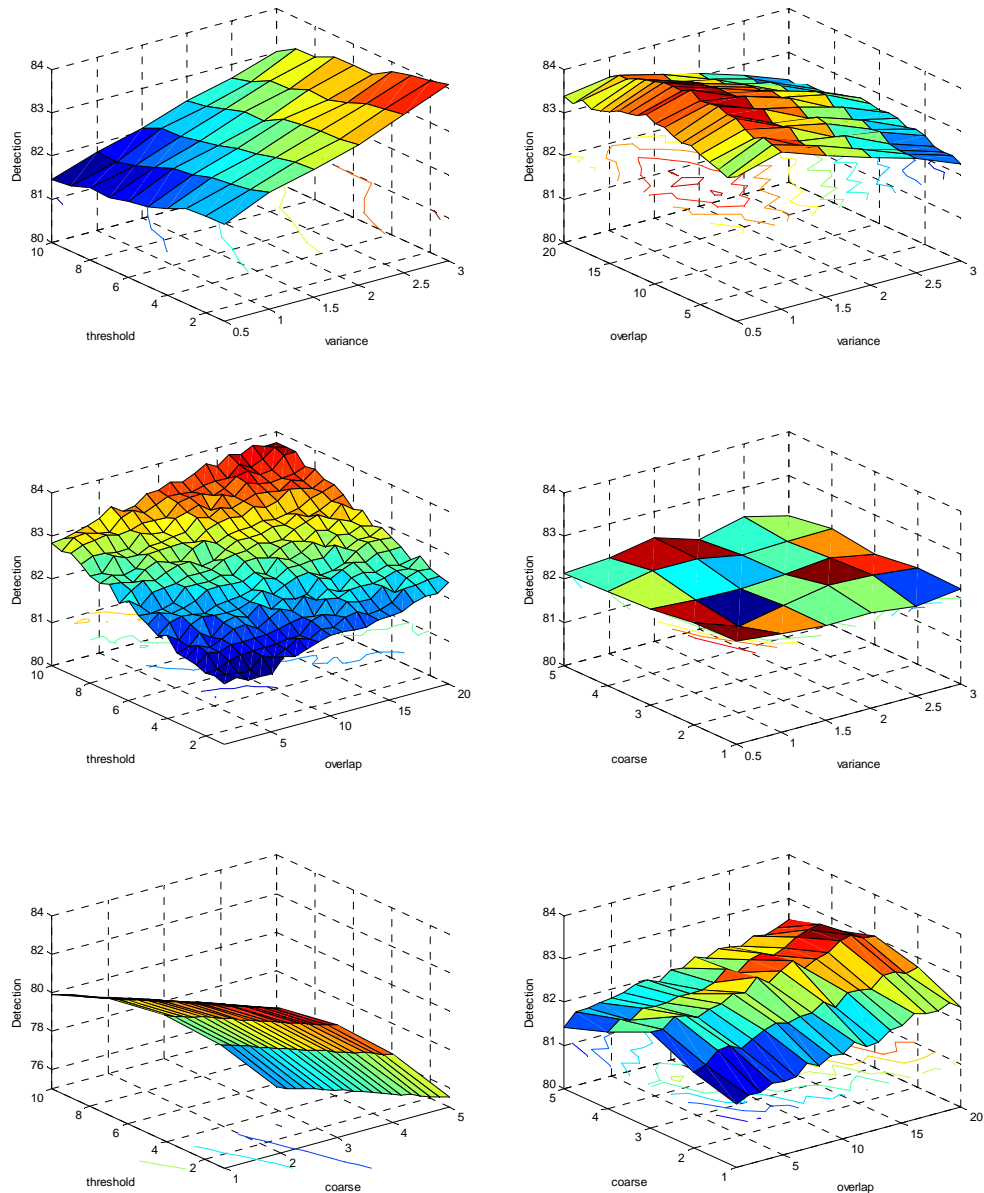
۴-۵-۱- تأثیر پارامترها روی نرخ آشکارسازی چهره ها

در این آزمایش از مجموعه داده های BioID Face Database استفاده می شود که شامل ۲۳ فرد مختلف است که از آنها ۱۵۲۱ تصویر خاکستری فام از تمام رخ موجود است. این تصاویر در شرایط روزمره گرفته شده است و دارای نورپردازی های متنوع و پس زمینه هایی پیچیده است. در کنار این مجموعه موقعیت چشمها هم به طور دستی قرار داده شده است تا صحت آشکارسازی چهره تایید شود (شکل ۴-۱۵). علاوه بر این تصاویر ۱۴۷۹ تصویر غیر چهره نیز استفاده شده اند که بعضی از آنها حاصل بهم ریختن فاز، تار کردن تصویر چهره در حد نا مفهوم شدن، و الگوهای شبیه به چهره (از نظر شهود انسانی) می باشند. برای آموزش سیستم از تکه های تصویر مجموعه تصاویر FERET استفاده شده است.



شکل ۴-۱۵: نمونه ای از مجموعه داده های BioID Face Database

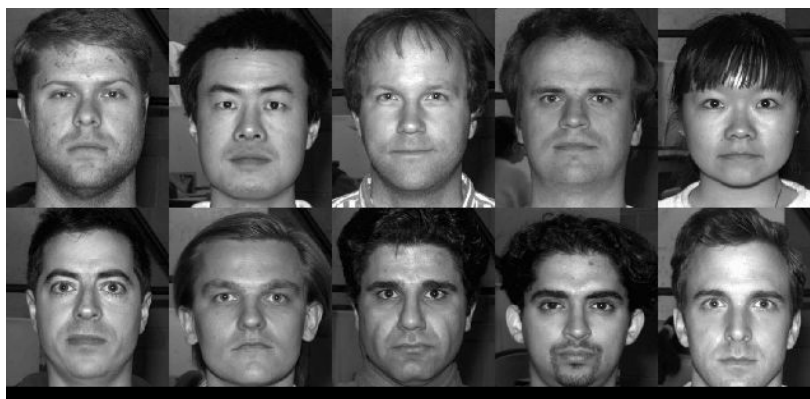
در شکل ۴-۱۶ تأثیر تغییرات پارامترها (واریانس اتصالات جانبی، همپوشانی میدانهای گیرندگی اولیه، آستانه تایید کاندید چهره، و درستی میدان گیرندگی نورونهای سطح بالا) روی عملکرد مدل برای تشخیص چهره در تصاویر ساده چهره/غیر چهره نشان داده شده است.



شکل ۴-۱۶: تأثیر تغییرات پارامترها روی عملکرد مدل برای تشخیص چهره در تصاویر ساده چهره/غیر چهره (بالا راست) نمودار واریانس-همپوشانی: افزایش واریانس به همسایگان بیشتری اجازه دخالت در تصمیم گیری را می دهد و افزایش همپوشانی این همسایگان را نزدیکتر می کند، با ثابت بودن آستانه جمع شدن این دو اثر باعث بروز اعلامهای اشتباهاً مثبت می شود. (بالا چپ) نمودار واریانس-آستانه: هر چقدر واریانس پایینتر باشد تصمیم گیری نورون برای همسایگان خود موثرتر خواهد بود و می توان آستانه را بالا برد (وسط راست) نمودار واریانس-درشتی: هر چه واریانس بیشتر باشد تاثیر همسایگان بر تصمیم گیری درباره یک نورون بیشتر است و می توان درشتی را کاهش داد ولی در مجموع این دو عامل زیاد قابل پیش بینی نیستند (وسط چپ) نمودار همپوشانی-آستانه: هنگامی که همپوشانی زیاد می شود به طور خودکار برای نواحی دارای احتمال چهره بودن بالا، مجموع بالاتر خواهد بود و آستانه باید بالاتر رود (پایین راست) نمودار همپوشانی-درشتی: این دو عامل با ثابت ماندن عوامل دیگر زیاد ارتباطی با یکدیگر ندارند (پایین چپ) نمودار آستانه-درشتی: هر چه درشتی بیشتر باشد، نورونهای زیادتری در مجموع شرکت می کنند و به آستانه بالاتری نیاز است.

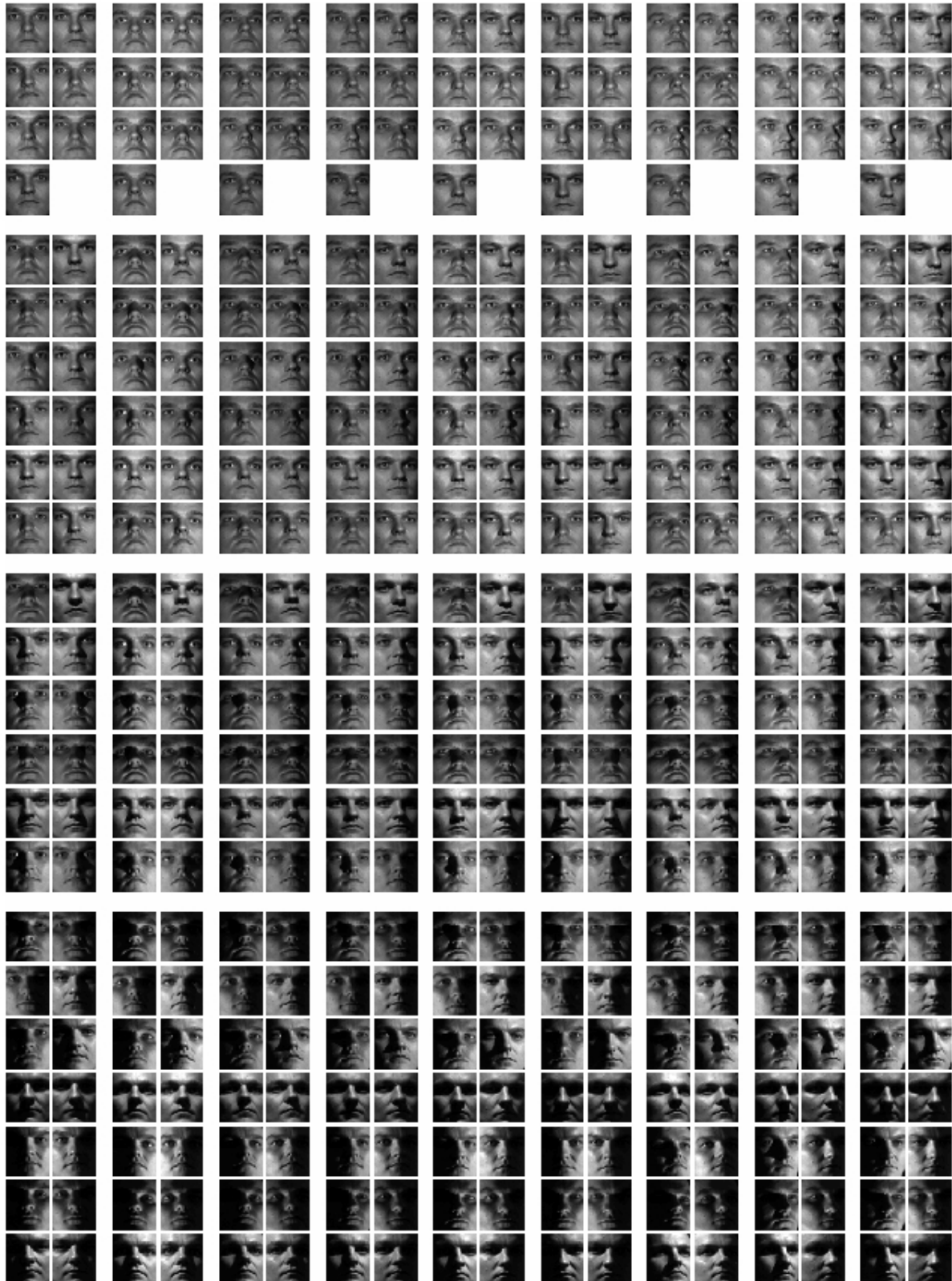
۴-۵-۲- آشکارسازی چهره در شرایط ویژه نورپردازی

در آزمایش روی تصاویر چهره دارای نورپردازیهای متفاوت، نرخ آشکارسازی افت چندانی نداشته است. علت امر آنست که تبدیل لگاریتمی که تصویر را پیش پردازش می کند، عملاً بازه دینامیکی شدت روشنایی تصویر را هموار می کند و در نتیجه کنتراست بین ساختارهای چهره باقی می ماند. در حالتی که سایه پردازی روی چهره الگوهای پیچیده ای را تولید می کند، میزان چهره بودن تکه تصویرهای اطراف نیز نمی تواند باعث شود که آن ناحیه یک چهره قلمداد شود. در این آزمایش تنظیمات پارامترها به گونه ایست که تصویر چهره ها در نورپردازی استاندارد آشکارسازی شده است.



شکل ۴-۱۷: افراد حاضر در مجموعه داده های Yale Face Database B

در آزمایش از مجموعه داده های Yale Face Database B استفاده می شود که شامل تصاویری از ۱۰ فرد مختلف است (شکل ۴-۱۷) و ۵۷۶۰ تصویر مختلف دارد که ۹ وضعیت قرار گیری سر در ۶۴ وضعیت نورپردازی از هر فرد موجود است ($10 \times 9 \times 64 = 4760$ ، شکل ۴-۱۸). موقعیت چهره در هر تصویر نیز در فایل جداگانه ای موجود است که می توان صحت چهره پیدا شده را بطور خودکار توسط آن سنجید. در این آزمایش با تنظیم پارامترها برابر بهترین نتایج آزمایش ۴-۵-۱ اکثر تصویرها آشکارسازی شدند و نرخ آن ۹۱.۳٪ بوده است. دلیل این امر اینست که حتی اگر فرم بخشی از چهره در اثر بروز حالتی در آن تغییر کند، بازهم بدلیل ثابت ماندن ساختار سایر قسمتهای چهره تغییر چندانی در آشکارسازی رخ نمی دهد.



شکل ۴-۱۸: نمونه ای از تصاویر یک فرد در مجموعه داده های Yale Face Database B

۴-۵-۳- آشکارسازی چهره های دارای حالت خاص

در آزمایشات زیر شرایطی مشابه آزمایش بخش ۴-۵-۱ حکمفرماست و بهترین تنظیم پارامترها در آن آزمایش (که منجر به بیشترین دقت آشکارسازی در آن آزمایش شد) در اینجا برای سیستم استفاده می شود.

در آزمایش عملکرد مدل روی تصاویر چهره دارای حالت خاص از مجموعه داده های Japanese Female Facial Expression (JAFPE) Database استفاده می شود که شامل ۷ حالت چهره مختلف است و ۲۱۳ تصویر مختلف دارد (شکل ۴-۱۹). در این آزمایش اکثر تصویرها آشکارسازی شدند و نرخ آن ۹۸.۹٪ بوده است. دلیل این امر اینست که حتی اگر فرم بخشی از چهره در اثر بروز حالتی در آن تغییر کند، بازهم بدلیل ثابت ماندن ساختار سایر قسمت‌های چهره تغییر چندانی در آشکارسازی رخ نمی دهد. پارامتر آستانه مدل در این آزمایش به گونه ای تنظیم شد که برای چهره هر فرد در یکی از تصاویر بدون حالت آن، مجموع تحریک نورون های لایه چهارم آن روی مرکز چهره برابر حد آستانه باشد.



شکل ۴-۱۹: نمونه ای از مجموعه داده های JAFPE

۴-۵-۴- آشکارسازی چند چهره و چهره های ناکامل

در حالتی که چند چهره در تصویر حاضر باشد، سیستم تک تک آنها را آشکار می کند. مشکل زمانی پیش می آید که آستانه مورد نیاز برای برخی از چهره ها با هم فرق می کند که تغییر آستانه باعث اعلام‌های اشتباهاً مثبت (حد آستانه پایین) و یا اشتباهاً منفی (حد آستانه بالا) می شود. همین مشکل آستانه در تصاویر شلوغ نیز به چشم می خورد و در تصاویر شلوغ تنظیم همپوشانی تکه های تصاویر به کمترین میزان خود مطلوب است زیرا اگر الگوی تکه ای تصادفاً مشابه چهره باشد حضور آن در تکه های مجاور با همپوشانی زیاد نیز احتمالاً به احتمال چهره بودن بالایی می انجامد و در اثر تقویت اثر آنها در لایه V2 اعلام چهره اشتباهاً مثبت خواهد بود.

برای چهره های دارای انسداد، اگر انسداد توسط چهره دیگری باشد هردوی آنها تشخیص داده خواهند شد. اما انسداد توسط اشیای غیر چهره بسته به میزان پوشاندن چهره ممکن است منجر به عدم آشکارسازی چهره شود.

برای این آزمایشها از مجموعه Bao Face Database استفاده می شود که حاوی ۲۲۱ تصویر بسیار پیچیده از دو نفر یا بیشتر است. پس از پیدا کردن تنظیمات بهینه برای پارامترها در این مجموعه، نشان داده شد که ۸۸.۱٪ چهره ها آشکار می شود و نتایج بالا از آن استخراج شده اند.

[جمع بندی]

۱. جمع بندی
۲. مسئله ها و راه حلها
۳. مقایسهها سایر روشها
۴. کارهای آتی

۵-۱- جمع بندی

در این پایان نامه چندین موضوع مستقلاً مورد بررسی قرار گرفت و سپس از ترکیب آنها و تلفیق ایده های برخاسته از هر یک از آنها روشی برای آشکارسازی چهره بدست آمد که به عنوان کاربردی برای کارهای نظری درون پروژه انتخاب شده، و تمرکز اصلی روی آن نبوده است.

دوران مطالعات برای این پایان نامه را می توان به چندین فاز مجزا جدا کرد:

۱. **بررسی ساختار فیزیولوژیکی انسان با تاکید بر سیستم بینایی:** در این بخش مغز انسان از دید یک فیزیولوژیست بررسی شده و مراحل فیزیکی طی شده توسط سیگنال بینایی در آن بررسی گردید. سپس با توجه به خصوصیات هر بخش مغز و فرضیه های موجود درباره آن، مراحل بینایی از دید واحدهای ساختاری مشخص می گردد.
۲. **بررسی ساختار کارکردی مغز انسان با تاکید بر بخش بینایی قشر خاکستری مخ:** در این بخش مغز انسان از دید یک عصب شناس مورد بررسی قرار می گیرد. ابتدا نحوه کارکرد کلی مغز بررسی می شود و سپس با تمرکز روی بخشهای قشر خاکستری مخ که در عملیات سیستم بینایی دخیل هستند، ساختار سلولی، کارکردهای شناخته شده و علل هر عملکرد بررسی می شود. بسیاری از پدیده هایی که درون سیستم عصبی انسان رخ می دهد تاکنون توجیه نشده است و بخش یزرگی از ادبیات عصب شناسی به معرفی این پدیده ها و معرفی مدلهایی برای آنها پرداخته است که تا حدودی بتوانند این پدیده ها را به همراه سایر پدیده ها توجیه کنند که تا کنون مدلهای یکپارچه با پیدا شدن مثالهای نقض و رد شدن پیش بینی های آنها راجع به نحوه مواجهه دستگاه عصبی شکست خورده اند و تنها مدلهایی در گوشه و کنار ادبیات به چشم می خورند که پابرجا مانده اند و با گسترش مطالعات تکوین می یابند. از جمله این مطالعات می توان به نظریه دو مسیر بینایی و شباهت سلولهای ساده ناحیه VI به فیلترهای گابور اشاره کرد که اولی با کشف تعاملات و مسیره های پسخور بین نواحی قشر بینایی کم اعتبار شده است و دومی از توجیه رفتارهای غیر خطی نورون عاجز مانده است. البته بدلیل ذات فرایند مدلسازی که ماهیت آن ساده سازی پدیده های پیچیده است این مدلها همچنان مورد استفاده قرار می گیرند و بر اساس آنها کشفهای جدید و پیشرفتهای علمی زیادی صورت می گیرد، ولی باید در نظر داشت که به دلیل ناقص بودن پیش فرضهای چنین دستاوردهایی، آنها نیز به نوبه خود دارای کاستیهایی هستند و تنها به صرف توجیه پدیده های بیشتر از مشاهدات نمی توان یک مدل را بر دیگری برتری داد.
۳. **بررسی قاعده مندیهایی موجود در تصاویر طبیعی:** تصاویر طبیعی تصادفی نیستند و ساختارهای موجود در آنها باعث تمایز بین انواع آن می شود. چشم انسان به صرف تعامل با چنین محیط قاعده مندی، باید در طی زمان از خود سازگاریهایی نشان بدهد و طبق نظریه تکامل، سازگاری هر چه بیشتر سیستم بینایی انسان در طول زمان با محیط اصراف خود توجیه پذیر است. توجه به عملکرد سیستم بینایی بدون در نظر داشتن محیطی که برای آن تطبیق یافته است کامل نخواهد بود و از همینرو باید این محیط نیز مورد بررسی قرار بگیرد. این محیط دارای قاعده مندیهایی درون خود است و از نظم خاصی تبعیت می کند. ادبیات زمینه تحقیقاتی فهم تصاویر طبیعی، به بررسی

این قاعده مندیها می پردازد و از آنها در جهت ساخت سیستمهای بینایی بهره می برد، چنانکه انتقال تصاویر تلویزیونی به عنوان یکی از ارکان ارتباطی روزمره انسانها نتیجه چنین بررسی هایی است. طبق یکی از معروفترین نظریه ها درباره سازگاری سیستم بینایی با محیط اطراف، سیستم بینایی در بخش اولیه خود فرایندی را آغاز می کند که طی آن سیگنالهای بینایی در طی مسیر خود به سمت نواحی سطح بالای مغز که از حاصل پردازش این سیگنالها در جهت انجام اعمال حیاتی و روزمره استفاده می کنند (نظیر تنظیم ساعت بیولوژیکی، تخمین سرعت یک جسم نزدیک شونده، استخراج اطلاعات مهم از میان سیل عظیم دادگان بینایی ورودی، شناسایی چهره های آشنا برای تنظیم روابط اجتماعی و مرتفع کردن نیازهای روانی و ...)، اطلاعات را طوری کدگذاری می کنند که با ساختار نورونی مغز تطابق داشته باشد و نحوه این کدگذاری بهترین روش موجود برای کدگذاری محرکهای موجود است. این نظریه که به کدگذاری بهینه معروف شده است مبین اینست که از میان تمام کدگذاریهای موجود برای تصاویر طبیعی، بهترین آنها در مغز وجود دارد. با توجه به این نکته پژوهشگران به تعریف معیار بهتر بودن یک کدگذاری نسبت به دیگری پرداخته اند. از این دسته معیارها می توان به کاهش افزونگی، افزایش پراکندگی، کاهش فاصله سیم کشی در مغز، تقلیل میزان مصرف انرژی، و غیره اشاره کرد. کاهش افزونگی یکی از معیارهایی است که به محیط بستگی دارد و از این رو در مبحث درک تصاویر طبیعی بسیار به آن پرداخته شده است. افزونگی در بر دارنده دانش است و این دانش با کشف افزونگیها در اختیار سیستم قرار می گیرد و می توان پس از این استخراج دانش و برای کدگذاری، این افزونگیها را از بین برد. مطالعه انواع افزونگی تصاویر طبیعی پس از اخذ تصویر از آن و گسسته سازی تصویر (پیکسلها) و در این پژوهش تبدیل آنها به تصاویر خاکستری فام، بررسی این قاعده مندی را به سروکار داشتن با ماتریسهایی از اعداد صحیح تبدیل می کند. با استفاده از علوم مرتبط مانند نظریه احتمالات، نظریه اطلاعات، پردازش سیگنال و سیستمهای خطی و غیره می توان به کشف قاعده مندیهای تصاویر طبیعی پرداخت. تعریف چارچوب آماری بررسی این قاعده مندیها و افزونگی ها را تسهیل می کند که در این پایان نامه نیز از آن بهره گرفته خواهد شد. در این راستا به مهمترین مسائل کلاسیک مرتبط مانند جداسازی کورکورانه منابع اشاره می شود که روشهایی آماری برای حل مشکل کاهش افزونگی در اختیار ما قرار می دهند و از راه حلهای موجود ایده هایی برداشته شده است.

۴. **شبکه های عصبی:** مسئله بازنمایی و یادگیری شبکه های عصبی به عنوان چارچوب اجرایی مغز، بخش دیگری از پژوهش را تشکیل می دهد. شبکه های عصبی مصنوعی به عنوان مدل ساده سازی شده ای از معادل طبیعی خود، ساختمان داده ای قوی برای انواع کارها فراهم می آورند. یکی از این وظایف یادگیری است که انواع یادگیری با توجه به هدف آنها برای کاهش افزونگی و بازنمایی مناسب باید بررسی شود.

۵. **استنتاج:** تخمین زدن مقدار مناسب پارامترها در مسائل گوناگون با توجه به ماهیت مسئله و نوع داده هایی که با آن سر و کار دارد، متفاوت است. در این پایان نامه با توجه به ماهیت ماتریسی کواریانس، روشهای گوناگونی برای استنتاج روی آن بررسی شده است و ایده هایی برای مدل کردن مسئله به صورت یک چارچوب استنتاج آورده شده است.

۶. **مدلهای نظری عملکرد قشر بینایی مغز:** پس از بررسی ساختارهای بیولوژیکی و مدلهای جسته و گریخته ای که در بخشهای مختلف مغز رفتار آنها توجیه می کند، نیار به مدلی که بتواند رفتار آنها در درک صحنه های طبیعی توجیه کند احساس می شود. این صحنه ها به عنوان محرکهای پیچیده راتر از آن هستند که مدلهای ساده که

بخشی از عملکرد قسمتی از مغز را بیان می کند بتواند کاملاً آنرا توجیه کند و نیاز به چارچوبی خاص دارد. در این راستا دو گروه مدل بوجود آمده است که عملکرد قشر بینایی را به نوبه خود توجیه می کند. دسته نخست مدل‌های پایین به بالا هستند. این مدل‌ها با مشاهده رفتار بیولوژیکی قشر بینایی انسان، منشأ پیدایش آنها را رفتار تک تک نورونها و بهم بندی آنها می دانند و سعی می کنند با ساختن سیستم‌هایی که تنها اجزای آن تعریف شده اند و نحوه اتصال این عناصر، عملکرد کلی مغز را توجیه کنند. چنین سیستم‌هایی به قوانین یادگیری هوشمندانه و عناصر کاملی نیاز دارند که پژوهشگران اغلب آنها را به طور کامل از روی مغز نمونه برداری می کنند. دسته دیگری این الگوریتم‌ها مدل‌های بالا به پایین هستند. این مدل‌ها در ابتدا یک هدف محاسباتی برای قشر بینایی فرض می کنند (مانند استقلال آماری، پراکندگی، ناهمبستگی پیکسلی، کدگذاری بهینه توزیع تصویر و غیره) و با ابزارهای آماری و ریاضی فیلترهایی برای تصاویر می سازند که نماینده نورون‌های قشر بینایی مخ می باشند و سپس سعی می کنند با استفاده از مدل خود، خصوصیات کشف شده نورونها را توجیه کنند و در مورد برخی پدیده‌هایی که در مدل دیده می شود پیش بینی‌هایی صورت دهند تا عصب شناسان را در پژوهش‌های آتی خود سمت و سو دهند. این دسته از مدل‌ها در کاربردهای بینایی رایانه ای نیز به وفور استفاده می شوند و سیر تکاملی آنها به صورت کامل بررسی شده است. جدیدترین نسل این روش‌ها مدل‌های سلسله مراتبی هستند که اخیراً مورد علاقه پژوهشگران قرار گرفته است و جایی برای خود در این ادبیات دست و پا کرده اند.

۷. آشکارسازی چهره: علیرغم اینکه این مقوله تنها کاربردی برای مدل پیشنهادی بوده است ولی خود مسئله ای جالب و چالش بر انگیز است. مقالات بیشماری در این باره به رشته تحریر درآمده است و ابزارهای متعددی از ادبیات هوش مصنوعی در آن بکار گرفته شده است. در این پایان نامه به برخی از ایده‌های بنیادین استفاده شده در این ادبیات اشاره شده و از آنها ایده‌هایی برای تولید یک سیستم آشکارسازی چهره استخراج می شود.

به طور کلی در این پایان نامه هدف اصلی تولید مدلی بوده است که بتواند خصوصیات نورون‌های ساده و پیچیده لایه VI را بخوبی مدل کند. برای اینکار ابتدا سیر ذهنی طراحان مدل تحلیل مؤلفه‌های مستقل دنبال شده است و سپس به یکی جدیدترین مدل‌های بالا به پایین یعنی مدل واریانس سلسله مراتبی پرداخته شده است. سپس توسعه ای بر این مدل پیشنهاد شده است که می تواند الگوهای ساختار همبستگی داده‌ها را مستقیماً یاد بگیرد. تا پیش از این مدل که ایده اولیه آنرا کارکلین (Karklin 2008) مطرح کرده است، مدل‌ها به بررسی وابستگی‌های بین فیلترهای خطی می پرداختند که معادل سلول‌های ساده لایه VI بودند سعی می کردند وابستگی‌های پاسخ آنها را چنان مدل کنند که کمترین افزونگی را داشته باشد. اینکه گفته شد مدل پیشنهادی مستقیماً ساختار همبستگی را یاد می گیرد بدین معنا بود که حتی فیلترهای خطی معادل سلول‌های ساده VI نیز توسط این مدل یاد گرفته می شوند و به کمک آنها می توان کوارینانس ساختار همبستگی داده‌ها را نیز تخمین زد و در کدگذاری حذف کرد. این مدل نیاز به محاسبه توزیع حاشیه ای ضرایب خطی برای محاسبه درست‌نمایی تخمین MAP برای متغیرهای مراتب بالاتر را نیز مرتفع کرده است و توصیف روشنتری از توزیع کانونی داده‌ها می دهد و در حالت پیش فرض یک گوسی چندمتغیره در نظر می گیرد. این روش همچنین حالت‌های فراکامل و فروکامل ویژگی‌های خطی را نیز پشتیبانی می کند. هدف این مدل در واقع اینست که اولاً همبستگی‌های مشاهده شده را بدون تکیه به تبدیلات خطی ثابت مدل کند و ثانیاً این همبستگی‌ها را در راستاهایی که توسط یک مجموعه ثابت از توابع پایه تعریف نشده اند ارائه دهد. این مدل بازنمایی مناسبی از تکه‌های تصویر بدست می دهد که با توجه به ساختار همبستگی و الگوی موجود در آنها (الگوهای

راستایی، فرکانسی و موقعیتی) یک کدگذاری ارائه می کند که برای ساختارهای مشابه، کدهای مشابه تولید می کند. این کدها تغییرات در توزیعهای تصویر را انتقال می دهند و به ناحیه های تصویری بعدی قشر مخ امکان تشکیل بازنمایی های مجردتر را می دهند.

همین نکته نیز در ادامه بحث که آشکارسازی چهره است نیز بکار گرفته شده است. با توجه به اینکه کدگذاری برای ساختارهای مشابه کدهای مشابه تولید می کند، و این نکته که محرک چهره نیز تصادفی نیست و بخش محدودی از فضای تصاویر را به خود تخصیص داده است که فضای چهره خوانده می شود، می توان نتیجه گرفت که یک کدگذاری با قابلیت خوشه بندی نواحی مشابه در فضای تصویر، بتواند عضویت یک تصویر را در زیرفضای چهره بررسی کند و آنرا رد کند. از آنجا که فضای تصویر در این پژوهش در واقع تمام ساختارهای ممکن برای یک تکه تصویر 20×20 پیکسلی را پوشش می دهد، زیرفضای چهره نیز عبارتست از تکه های تصویری که در تصاویر چهره احتمال دیده شدن دارند. هر چه ابعاد تکه تصویر بالاتر رود وابستگیهای بیشتری توسط مدل پیشنهادی آشکار می شود ولی هزینه محاسباتی مدلهای نیز بالاتر می رود از اینرو در ادبیات تحلیل قشر بینایی مغز، این ابعاد استفاده می شود. برای مشخص شدن وضعیت عضویت یک تکه تصویر در فضای چهره، فضای کدگذاری شده تصاویر توسط یک شبکه پرسپترون چندلایه آموزش می بیند که در نهایت این شبکه قادر به تعیین میزان عضویت (عضویت نرم) تکه تصویر به فضای چهره خواهد بود. اگر در تصویر چند تکه تصویر نزدیک به هم همگی شبیه به چهره تشخیص داده شوند، مجموعه آنها یک چهره تلقی خواهد شد.

پس از مرور رهیافت این پایان نامه، در ادامه به مسائلی که تولید این سیستم به آن برخورد کرده است، راه حل آنها و ایده های برگرفته شده از منابع دیگر اشاره می شود.

۵-۲- مسئله و راه حلها

همانگونه که در نقاط مختلفی از پایان نامه اشاره شد، در رسیدن به این الگوریتمها، از مقالات زیادی ایده گرفته شده است که در اینجا به طور مختصر به آنها اشاره می شود. پیش از هر چیز باید اشاره کرد که تمرکز این پایان نامه روی مسیر بینایی بطنی است که اشیای موجود در تصویر را تشخیص می دهد. بر طبق نظریه اطلاعات کدهایی که از نظر آماری کارتر هستند، همانهایی هستند که قاعده مندیهایی داده ها را به بهترین نحو اخذ می کنند و از سوی دیگر از آنجا که سیستمهای بیولوژیکی زیر فشار تکاملی شدیدی هستند، می توان انگاشت که کدهایی که آنها استفاده می کنند بسیار کاراست.

در مطالعه الگوریتم بیدار-خواب هینتون برای استنتاج مدل مولد، ایده بدست آوردن کدی که دارای طول توصیف کمینه باشد به ذهن می رسد. برای این منظور تلاش شد که طول بازنمایی مورد طراحی ثابت و یا وابسته به پارامتر دیگری (تعداد پیکسلهای تکه تصویر، تعداد جهتهای مورد نظر برای انقباض و انبساط توزیع و غیره) نباشد.

قاعده مندیهایی آماری تکه های تصاویر طبیعی چنانکه در فصل دوم خلاصه شد عبارتند از همبستگیهای مرتبه دوم آشکار، وجود جهتهای پراکندگی و متعامد نبودن راستاهای مهم در آنها. تعمیم عبارتست از تشخیص مشخصات مشترک یک کلاس از روی نمونه های آن. هدف مدل ارائه شده نیز یادگیری توزیع آماری ایست که مشخصات نواحی محلی تصویر را در بر بگیرد و آنها را از روی تکه های منفرد تصویر تشخیص دهد. از میان روشهایی که به بررسی وابستگیها و حذف افزونگی می

پرداختند روشی که این قاعده مندیها را به طور کامل در اختیار دارد وابستگی واریانسی است. از آنجا که وابستگی واریانسی باید برای تک تک فیلترهای تصویر رفع شود و نیازمند دورانهایی زیادی است، می توان مجموعه تمامی وابستگیهای واریانسی را یکجا مدل کرد که اینکار با تغییر ماتریس کواریانس به طوری که وابستگیها از بین برود امکانپذیر است. همبستگی در هر قسمت تصویر می تواند با یک توزیع گوسی و کواریانس آن توصیف شود. تفاوت در توزیعهای توأم پاسخفیلترهای خطی به تصویر، مشخصات الگوهای درون همبستگیها را آشکار می کند که وقتی ویژگیهای بیشتری مد نظر قرار بگیرند واضحتر نیز می شود. این الگوها می توانند برای تعمیم درون نواحی استفاده شوند، در حالی که همچنان بتوانند آنها را تمایز بدهند.

یک راه ساده برای خلاصه سازی الگوی همبستگی برای نوع تصویر داده شده ماتریس کواریانس داده هاست. یک کد نورونی برای این ساختار می تواند با شماره گذاری مجموعه کواریانسهای مشاهده شده و انتساب یک نورون به هریک از آنها ساخته شود ولی این راه حل دارای دو ایراد است: نخست اینکه کلاس تصاویر محلی از پیش معلوم نیست، و دوم اینکه با در نظر گرفتن تعداد محدود نورون در سیستم بینایی، بازنمایی تمام انواع تصاویر امکان پذیر نیست حتی اگر تنها ترکیبات ممکن مرزهای تصاویر را هم در نظر بگیریم. به جای این راه حل یک کد پراکنده در نظر گرفته شده است که در آن فعالیت سطح بندی شده جمعیت نورونها برای توصیف زنجیره الگوهای بالقوه کواریانس استفاده می شوند. این مدل همبستگی موجود در نواحی محلی تصویر را با یک توزیع گوسی چند متغیره دارای میانگین ثابت صفر و کواریانسی که تابع فعالیت نورنهاست بازنمایی می کند. این تعریف ساده آماری از سویی انعطاف پذیری در اخذ زنجیره توزیعهای تصاویر طبیعی و از سوی دیگر سادگی محاسباتی را برای تخمین راحت پارامترها به ارمغان می آورد.

یک مزیت مهم این روش اینست که بجای فرض کردن نامتغیری (ویا حساسیت) به تعداد اندکی از پارامترهای محرک مانند موقعیت و مقیاس، مدل مجموعه ای به مراتب عمومی تر از ویژگیها را یاد می گیرد که توسط ساختارهای آماری تصاویر طبیعی بدست می آیند. اگر نورونهای لایه های بالاتر بینایی بر اساس این آمارگان تعمیم یافته باشند، باید در مقابل برخی از ابعاد ورودی نامتغیر باشند و پاسخ آنها به تصاویر طبیعی باید ساختار آماری مشترکی را در نواحی محلی منعکس کند. بنابراین این مدل یک راه کمی برای تفسیر پاسخ نورونها به محرکهای پیشرفته است که به قاعده مندی آماری آنها باز می گردد.

مدل واریانسی، و مجموعه فراکامل توابع پایه A یک ماتریس کواریانس کانونی ($\vec{v} = 0$) برابر AA^T خواهد داشت در حالیکه در مدل کواریانسی توزیع پیش فرض همان گوسی یکرخت با $C=I$ است. با افزایش فراکامل بودن ویژگیهای خطی این اثر محسوستر می شود. هنگامی که یک عضو فروکامل ویژگیهای خطی استفاده می شود، مدل واریانسی تمام فضای داده را پوشش نمی دهد (ابعادی وجود خواهند داشت که واریانس در طول آنها صفر خواهد بود). از سوی دیگر مدل کواریانسی دارای پیش فرض ۱ برای تمام جهات بدون پوشش در فضاست هرچند این مدل نیز نمی تواند هیچ تغییری در این راستاها را منعکس کند.

در بررسی قانونهای یادگیری بین لایه های شبکه عصبی مشخص شد که قانون یادگیری هب، بیشینه سازی آنتروپی را تولید می کند و سیستم یادگیری سریعی خواهد داشت. همچنین روش پیشنهادی شراودولف که مجموعه ای از واحدهای رقابت کنند برای کشف کدگذاری جمعیتی را معرفی کرده است، ایده استفاده از جنبه کدگذاری جمعیتی شبکه عصبی برای تصاویر را تقویت کرد. این ایده در مقابل سایر روشهای کاهش افزونگی قرار می گیرد که سعی می کنند افزونگی بین دو

کمیت وابسته را کاهش دهند و تصویر را پس از کاهش افزودگی بازسازی، و بیان می دارد که تنها کدگذاری برای منظور این پایان نامه کافیت و لذا به خود را ملزم به ساخت فیلترهای خطی لایه اول نمی کند که شرط حذف وابستگی آنها معکوس پذیر بودن ماتریس تلفیق آنهاست. به جز پاسخ دادن به تعداد معدودی از ویژگیهای تصویری برانگیخته کننده و یا فرونشاندن، تعداد زیادی از نورونها برای تشکیل یک الگوی تنوع که در توزیع مخصوص تصویر وجود دارد شرکت می کنند. با وجود این اهمیت کارکردی این زیرواحدها اصلاح ساختار آماری توزیع کدگذاری شده است، این نورونها همچنین ویژگیهای محرکی که این نورون بیش از همه به آن حساس است را منعکس می کند. باید توجه داشت که یک نورون مدل، به جای آنکه با فعال شدن خود از حضور بهترین محرک خود خبر بدهد، توسط تمام تصاویر این توزیع فعال می شود ولی بسته به محرکی که در فضای تصویر موجود است، احتمال کمی وجود دارد که از فعالیت نورون جلوگیری شود.

بررسی مجموعه ویژگیهای خطی همچنین نشان داده است که همانند مدل سلسله مراتبی واریانس، خصوصیات جمعیتی این ویژگیها با جمعیتی که از مدل‌های خطی مانند ICA بدست می آید متفاوت است. مجموعه کامل بردارهای \vec{b}_k دارای مؤلفه های فرکانس پایین بیشتری نسبت به توابع پایه ICA یا بردارهایی است که وقتی بدست می آید که تعداد آنها با ابعاد داده برابرست.

در طول بررسی مدل‌های بالا به پایین مشخص شد که مدل آماری متغیر نهفته در تمام آنها کاربرد دارد و چارچوب مناسبی برای مدل کردن وابستگی دو تابع یا مقدار است. برای سهولت کار معمولاً این وابستگیها از نوع خطی در نظر گرفته می شوند. به ازای یک بردار از متغیرهای نهفته یک ماتریس تلفیق در نظر گرفته می شود که هر ستون آن ضرایب بردار متغیرهای نهفته را تشکیل می دهد که از ضرب آنها یکی از مقادیری حاصل می شود که در رابطه وابستگی قرار دارند. تعداد ستونهای این ماتریس نیز برابر تعداد کمیت‌هایی است که به هم وابسته هستند. استفاده از توابع غیرخطی در این ادبیات کمتر دیده شده است و لذا فرض می شود که وابستگی این کمیتها با یک تبدیل خطی از بین می رود و ماتریس تلفیق در واقع همان تبدیل به شمار می رود. این تبدیلات خطی در واقع همان پاسخ الگوریتم تعقیب افکنش هستند.

ایده استفاده از الگوریتم پیشینه درست‌نمایی برای تنظیم پارامترها در مقالات مختلفی و برای روشهای متنوعی مانند ICA آمده است. استفاده از این روش نیازمند محاسبه درست‌نمایی بود که در بعضی از این مقالات استفاده از درست‌نمایی لگاریتمی توصیه شده و مزایای آن تشریح شده بود. روش محاسبات (شامل به توان رساندن ماتریس، مشتق گیری از ماتریس به ازای یک بردار و غیره) نیز در منابع مختلفی موجود بود و برای تقریب زدن مقدارهای آن نیز راهکارهایی پیشنهاد شده بود. در یکی از منابع به یادگیری رقابتی برای درست‌نمایی پیشینه اشاره شده بود که گزینه جالبی به نظر می رسید زیرا با استفاده از آن بردارهای \vec{b}_k حاصل هر یک برای مجموعه خاصی از ویژگیهای تصویر (ترکیبی از راستا، فرکانس و موقعیت مکانی) اختصاصی می شود. همچنین رفتار رقابتی به رفتارهایی غیرخطی در پاسخ مدل می شود که با برخی از ویژگیهای سلولهای ساده مانند فرونشانی اطرافیان در V1 مطابقت دارد. این امر نشانگر آنست که رقابت عمومی و بازنمایی بسیار پراکنده ممکن است مسبب برخی از خصوصیات غیرخطی شوند. پس از پیاده سازی این روش بجای ML مشخص شد که این روش باعث می شود که تعدادی از بردارهای جهت \vec{b}_k بطور کلی هیچگاه تنظیم نشوند. از سویی کم کردن نرخ یادگیری در استنتاج روند آموزش را کند می کرد و باعث می شد که ساختار داده هایی که کمتر دیده می شوند اصلاً اخذ نشوند و از سوی دیگر کم

کردن تعداد بردارهای جهت \vec{b}_k از قدرت روش می کاست و مزایای حالت فراکامل از ست می رفت. بنابراین از این روش استفاده نشد.

یک پایه کامل، که توسط ICA یا یک مدل سلسله مراتبی با تعداد محدودی ویژگی ساخته می شود، حاوی مجموعه کاملی از ویژگیها نیست. از نگاهی دیگر می توان دید که ویژگیهای راستدار و محلی برای کدگذاری کارای تصاویر طبیعی بهینه نیستند ولی می توانند بهترین و فشرده ترین توصیفی از توزیعهای تصویر و ساختارهای تصویری مراتب بالاتر را ارائه کنند. برای همین لزوم استفاده از بازنمایی فراکامل بیش از پیش توجیه پذیر است.

برای پیش پردازش داده ها در اکثر مدلها میانگین داده ها را از آنها کم می کنند. اینکار به سرعت پردازش می افزاید زیرا مقادیر منفی در آموزشها قدرت جریمه کردن دارن و تأثیر بیشتری می گذارند. همچنین برای پیش پردازش، وایتینگ روی آنها اعمال می شود تا ساختار همبستگی عمومی آنها حذف شود. در مدل کردن وابستگیهای کواریانس اینکار باعث می شود که کواریانس تعریف شده توسط مدل تنها انحرافات از آمارگان کلی مجموعه داده ها را منعکس کند.

یکی از رایج ترین ساده سازیهای موجود در ادبیات، صفر فرض کردن میانگین توزیع هاست. اینکار به سادگی انجام می شود و تفسیر آن چندان مشکل نیست، ولی در خوانایی مدل و تعبیر مقادیر کمک شایانی می کند. در مدل حاضر نیز این فرض صفر بودن میانگین همانند سایر روشهای موجود در ادبیات، در مورد درستنمایی و تابع احتمال اولیه استفاده شده است.

در انتخاب تابع احتمال اولیه معمولاً به دنبال یک تابع زبرگوسی می گردند که متقارن نیز باشد. توزیع لاپلاسی با داشتن این ویژگیها و نیز خواص ریاضی مشخص گزینه خوبی به شمار می رود و در روشهای معروفی مانند ICA نیز از آن استفاده شده است. البته شرط تقارن اجباری نیست و ممکن است توابع بهتری نیز برای تابع احتمال اولیه موجود باشند اما در این پیاده سازی از این روش استفاده می شود و همانطور که در جدول پارامترهای مدل آمده است از این توزیع استفاده شده است (واریانس تابع احتمال اولیه برابر ۱). البته توزیع نامتقارن ضرایب DC کواریانس نشان می دهد که احتمال اولیه لاپلاسی برای واحد کنتراست DC مناسب نیست. در مدلهای خطی مانند ICA، اثر مشابهی برای ضرایب روشنایی DC نیز مشاهده شده است که آنها هم پراکنده نیستند و این مؤلفه معمولاً در آموزش مستثنی واقع می شود. بنابراین احتمالاً می توان تطبیق این مدل را با یک تابع اولیه انعطاف پذیر برای این ضرایب، و یا با اصلاح مدل برای مستثنی کردن این راستا از فضای ماتریسهای کواریانس بهبود داد. تابع احتمال اولیه متقارن برای ضرایب کواریانس به فرض بزرگی درباره تقارن الگوهای همبستگی در تصاویر طبیعی تکیه دارد. مدلی که از این فرض استفاده نکند و تنها از متغیرهای نهفته مثبت استفاده کند را می توان آسانتر به فعالیت نورونها نسبت داد (که فاقد فعالیت منفی هستند). به عنوان مثال می توان به تحلیل معنادارتری از نقش ویژگیهای تصویر در برانگیختگی یا فرونشانی سلول دست یافت.

برای محاسبه تقریبی پیشخور که پراکنده هم باشد، می توان روشهایی که برای مدلهای خطی با قیدهای پراکندگی استفاده می شوند را به خدمت گرفت. این روشها با در نظر گرفتن یک جریمه در تابع هدف برای راه حلهای ناپراکنده الگوریتم را متمایل به یافتن پاسخهای پراکنده تر می نمایند. هنگامی که تابع درستنمایی به عنوان تابع هدف استفاده می شود، همانطور که در مدلهای بالا به پایین مشاهده شد می توان جملات دیگری را نیز به آن افزود تا جواب بدست آمده خصوصیات دیگری مانند پراکندگی را نیز در خود داشته باشد. این جملات که نقش پاداش/جریمه را دارند روند استنتاج را دچار ناپایداری

می کنند. در پیاده سازی حاضر معیار پراکندگی باعث می شد که نتایج حاصله در تکرارهای حلقه استنتاج نوسان کند. ضرایب کوچک برای این جمله نیز باعث می شود که نقش چندانی در نتایج نداشته باشد. قرار دادن مجدد این معیارها در تابع هدف در کارهای آتی مجدداً صورت خواهد پذیرفت و علت این نوسانات مشخص خواهد شد.

در تحلیل تصاویر و بررسی ساختار آنها از تکه تصویرهایی از سرتاسر آن استفاده می شود. این نواحی در اکثر تحلیلها، به جای اینکه کل صحنه طبیعی را در بر بگیرند، به دلیل پیچیدگی محاسباتی به تکه های کوچک تصویر محدود می شوند. با این وجود بدلیل اینکه اندازه میدان گیرندگی در سیستم بینایی نیز محدود است، این فرض آنچنان که در نگاه اول به نظر می آید محدود کننده نیست.

برای آموزش شبکه MLP که میزان عضوین در فضای چهره را مشخص می کند، جمع آوری الگوهای نماینده چهره آسان است ولی تولید الگوهای نماینده غیرچهره ها دشوارتر است. این مسئله با یک روش بوت استرپ^{۳۲۴} تخفیف داده شده است که در آن تصاویر را با انتخاب خود در طول آموزش به داده های آموزشی اضافه می کند. با شروع از یک مجموعه کوچک نمونه های غیرچهره در مجموعه آموزشی، دسته بندی کننده با این پایگاه داده از مثالها آموزش می بیند. سپس تشخیص دهنده چهره روی یک دنباله تصادفی از تصاویر اجرا می شود و تمامی الگوهایی که سیستم فعلی اشتبهاً چهره اعلام کرده است را جمع آوری می کند. داده های اشتبهاً مثبت سپس به پایگاه داده به عنوان نمونه های جدید غیرچهره اضافه می شوند. این روش بوت استرپ از مشکلات جمع آوری صریح نمونه های الگوهای غیرچهره اجتناب می کند و در کارهای بعدی نیز از آن استفاده شده است

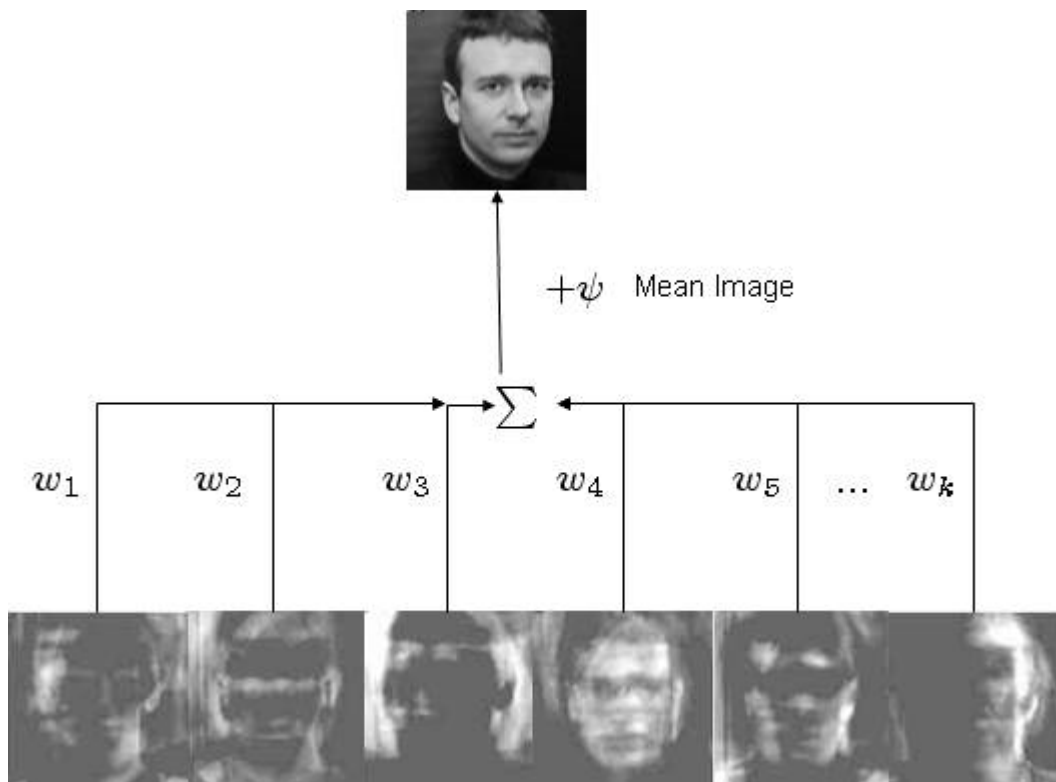
برای آشکارسازی چهره در هر جای تصویر، شبکه عصبی به تمام نقاط تصویر اعمال می شود. برای آشکارسازی چهره هایی بزرگتر از 20×20 پیکسل، تصویر ورودی باید مکرراً زیرنمونه گیری شود و شبکه به هر مقیاس اعمال شود. برای کسب نامتغیری، تعداد زیادی نمونه چهره از اندازه ها، راستاها، موقعیتها و نورپردازیها برای آموزش شبکه استفاده شده اند.

۵-۳- مقایسه با سایر روشها

۵-۳-۱- استفاده از تحلیل مؤلفه های اصلی در آشکارسازی چهره

در این بخش خلاصه ای از یکی از موفقترین و معروفترین روشهای آشکارسازی و شناسایی چهره ارائه می شود. این روش که چهره های یگه نام دارد (Turk & Pentland 91) سعی دارد که مؤلفه های اصلی اوزیع چهره ها را بیابد که در واقع بردارهای یگه ماتریس کوارینانس مجموعه تصاویر چهره می باشد. این بردارهای یگه را می توان به عنوان مجموعه ای از ویژگیها تلقی کرد که با هم تنوع بیت چهره ها را مشخص می کنند. هر مکان تصویر کمابیش در هر بردار تاثیری دارند بنابراین می توان بردارهای یگه را به عنوان چهره هایی شبیح مانند دانست که به آنها چهره های یگه اطلاق می شود.

هر چهره در مجموعه آموزشی را می توان دقیقاً به صورت ترکیب خطی چهره های یکه آن نشان داد (البته می توان آنرا به صورت ترکیب خطی هر مجموعه ای از تصاویر نشان داد که از آن جمله می توان به چهره های یکه افراد دیگر نیز اشاره نمود، شکل ۵-۱). تعداد چهره های یکه ممکن برابر تصاویر چهره در مجموعه آموزشی است. با اینحال چهره می تواند توسط بهترین چهره های یکه تقریب زده شوند. منظور از بهترین، آنهایی است که بزرگترین مقدار یکه متناظر را داشته باشند و از همین رو دارای بیشترین واریانس بین مجموعه تصاویر باشند (شکل ۵-۲) و دلیل اصلی استفاده از چهره های یکه کمتر، کارایی محاسباتی است. بهترین m چهره یکه در زیرفضای m بعدی («فضای چهره ها») از میان تمام تصاویر قرار دارند. همانطور که توابع سینوسی با فرکانسها و فازهای متفاوت پایه های یک تجزیه سینوسی را می سازند (و در واقع توابع یکه سیستم خطی هستند)، چهره های یکه نیز بردارهای پایه تجزیه چهره های یکه هستند.



شکل ۵-۱: بازسازی تصویر چهره توسط ترکیب خطی چهره های یکه

تصاویر چهره در فضای بسیار بزرگ تصاویر به طور تصادفی توزیع نشده اند و می توان آنها را با ابعاد نسبتاً اندکی توصیف کرد. ایده اصلی تحلیل مؤلفه های اصلی یافتن بردارهایی است که به توزیع چهره ها در این فضای تصاویر مربوطند. که همانگونه که اشاره شد این فضا را فضای چهره می نامند. هر تصویر به طول n^2 ترکیبی خطی از تصویرهای اصلی چهره است. از آنجایی که این چهره ها بردارهای یکه ماتریس کواریانس تصاویر اصلی چهره هستند و از آنجایی که ظاهری شبیه به چهره دارند به آنها چهره های یکه اطلاق می شود. کاهش ابعاد فضا از n^2 به m بعد به کمک انتخاب بردارهای یکه متناظر با بزرگترین مقدارهای یکه، محاسبات را تا حد زیادی کاهش می دهد.



شکل ۵-۲: (چپ) تصویر چهره اصلی (راست) نمایش سه بردار یکه متناظر با سه مقدار یکه بزرگ حاصل از تجزیه بردار یکه تصویر چپ

هر تصویر ورودی با کم کردن میانگین تمام تصاویری که سیستم با آن آموزش دیده است و سپس ضرب در m بردار یکه متناظر با بزرگترین مقادیر یکه به یک بردار ویژگی تبدیل می شود. از این بردار برای این استفاده می شود که مشخص شود کدامیک از کلاسهای از پیش تعیین شده چهره، بهتر تصویر ورودی را توصیف می کند که این تصمیم گیری از روی فاصله اقلیدسی بردار تصویر ورودی و بردار نماینده هر کلاس انجام می شود.

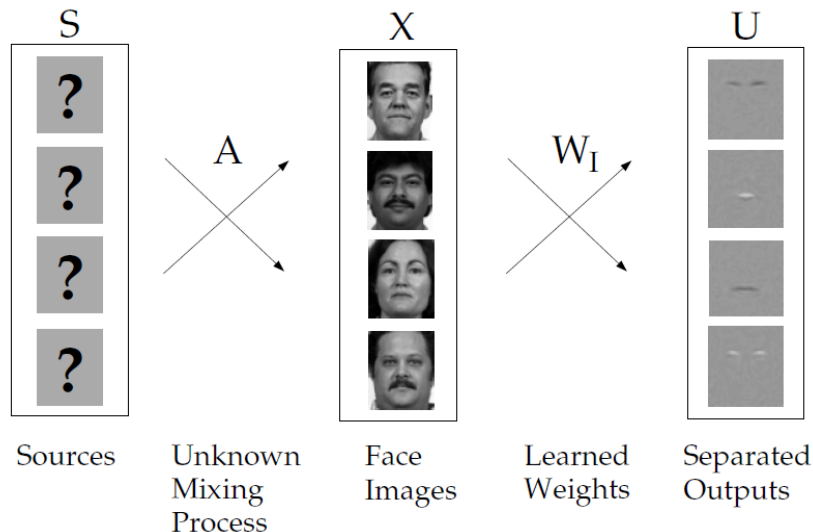
می توان از دانش موجود از فضای چهره استفاده کرد و چهره را در یک تصویر آشکارسازی کرد. ایجاد بردار وزنها برای یک چهره معادل افکنش آنها روی فضای کم بعدتر چهره است. فاصله بین تصویر و افکنش آن روی فضای تصویر بسیار راحت محاسبه شده و \mathcal{E} نامیده می شود. طبق بررسیهای مقاله (Turk & Pentland 91) تصاویر چهره وقتی که بر روی فضای تشویر بازتاب می شوند تغییر چندانی نمی کنند ولی تصاویر غیر چهره پس از این افکنش بسیار متفاوت خواهند بود. این ایده ساده برای آشکارسازی چهره در تصویر استفاده می شود: در هر مختصات تصویر فاصله \mathcal{E} بین زیر تصویر محلی و فضای چهره حساب می شود. این فاصله به عنوان معیاری برای چهره بودن استفاده می شود و نتایج محاسبات مربوط به \mathcal{E} هنگامی که در آرایه ای زهیره شود یک نقشه چهره را می سازد. کمینه های آشکار در این نقشه چهره مکان چهره را در تصویر نشان می دهند. مکان دقیق چهره ها با آستانه گیری از تصویر مشخص می شود. متأسفانه محاسبه مستقیم معیار فاصله هزینه محاسباتی زیادی دارد و روشهای متعددی برای محاسبه کارای آن پیشنهاد شده است.

۵-۳-۲- استفاده از تحلیل مؤلفه های مستقل در آشکارسازی چهره

در این بخش خلاصه ای از دو تا از موفقترین و معروفترین روشهای آشکارسازی و شناسایی چهره با استفاده از تحلیل مؤلفه های مستقل ارائه می شود (Bartlett 2002).

۵-۳-۲-۱- تصاویر پایه مستقل آماری (معماری ۱)

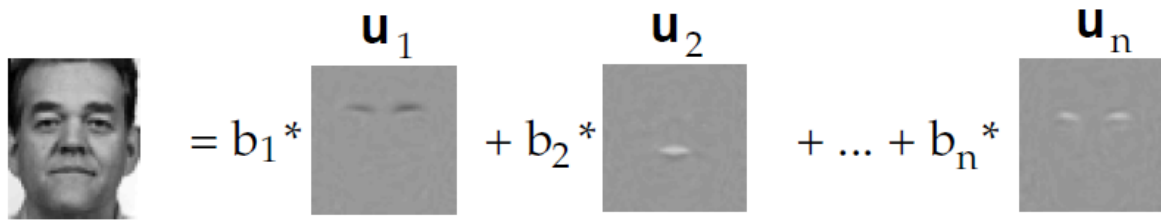
جهت پیدا کردن تصاویر پایه مستقل آماری برای مجموعه‌ی چهره‌ها، بر اساس مدل شکل ۵-۳، اجزای مستقل نمایش تصویر چهره پراکنده می‌شود. تصویر چهره در X به عنوان یک ترکیب خطی از مجموعه‌ای مجهول از منبع تصاویر مستقل آماری S فرض شده است، در جایی که A نیز یک ماتریس مخلوط مجهول می‌باشد. منابع با کمک ماتریس فیلترهای خطی W_I بازیابی می‌شوند، که خروجی‌های مستقل آماری U را تولید می‌کند.



شکل ۵-۳: مدل ساخت چهره با معماری ۱. برای یافتن مجموعه‌ای از مؤلفه‌های مستقل تصویری، تصاویر درون X به عنوان ترکیبی خطی از تصاویر پایه مستقل آماری، S ، در نظر گرفته می‌شوند که در آن A یک ماتریس مخلوط مجهول است. تصاویر پایه با ماتریسی از فیلترهای یادگیری شده W_I بازیابی می‌شوند که خروجی‌های مستقل آماری U را می‌سازد.

تصاویر سطرهای ماتریس ورودی، X ، را تشکیل می‌دهند. با داشتن تصاویر ورودی در سطرهای X ، خروجی‌های ICA در سطرهای $U = W_I X$ نیز تصویر می‌باشند، و مجموعه‌ای از تصاویر پایه‌ی مستقل برای چهره را تشکیل می‌دهند (شکل ۵-۴). این تصاویر پایه می‌توانند به عنوان یک مجموعه از ویژگی‌های مستقل آماری چهره در نظر گرفته شوند، که مقادیر پیکسل‌های هر تصویر ویژگی را نمی‌توان از روی مقادیر پیکسل‌های تصویر ویژگی دیگری پیش‌بینی کرد. همانطور که در شکل ۵-۴ آمده است، نمایش ICA از ضرایب ترکیب خطی تصاویر پایه‌ی مستقل در U تشکیل شده است که هر تصویر چهره را تشکیل می‌دهد. ماتریس ضرایب، B ، از ماتریس مخلوط $A = W_I^{-1}$ بدست می‌آید.

تعداد مؤلفه‌های مستقلی که توسط الگوریتم ICA پیدا می‌شود، با ابعاد ورودی مطابقت دارد. برای اینکه به تعداد مؤلفه‌های مستقلی که توسط الگوریتم استخراج می‌شوند کنترل داشته باشیم، به جای اجرای ICA بر روی تمامی n تصویر اصلی، ICA را بر روی مجموعه‌ای از m ترکیب خطی آن تصاویر اجرا می‌کنیم که $m < n$ برقرار باشد. باید خاطر نشان کرد که مدل ساخت تصویر، تصاویر X را به عنوان ترکیبی خطی از مجموعه‌ای مجهول از منبع تصاویر مستقل آماری فرض کرده است. مدل ساخت تصویر در اثر تعویض تصاویر اصلی با تعدادی دیگر از تصاویر ترکیب خطی تأثیر نمی‌پذیرد.



شکل ۴-۵: بازنمایی تصاویر پایه مستقل از ضرایب b برای ترکیب خطی تصاویر پایه مستقل u تشکیل شده است که هر تصویر چهره X را می سازد. $ICA Representation=(b_1, b_2, \dots, b_n)$

با الهام از یک روش که بر روی تحلیل اجزای مستقل از داده‌های fMRI اجرا شده است، اولین m بردار یگانه تصویر برای این ترکیب‌های خطی استفاده می‌شود. تحلیل مولفه اصلی بر روی مجموعه تصویری که در آن، محل پیکسل‌ها مانند یک مشاهده تلقی می‌شوند و هر تصویر چهره‌به عنوان یک اندازه گیری محسوب می شود، ترکیبی خطی از پارامترها (تصاویر) بدست می دهد که مسئول لیست‌ترین تنوع در مشاهده ها (پیکسلها) می باشند. استفاده از بردار PCA در ورودی رابطه های مرتبه بالا را از بین نمی برد. این رابطه ها همچنان در میان داده‌ها باقی مانده است ولی جدا نمی شود.

فرض کنید P_m ماتریسی است که اولین m محور مولفه های اصلی ستون‌هایش را تشکیل دهد. با اجرای ICA بر روی P_m^T ، ماتریسی از m منبع تصاویر مستقل در سطرهای U تولید می‌شود. ضرایب، \vec{b} ، برای ترکیب خطی تصاویر پایه در U که شامل تصاویر چهره در X می‌باشد، به صورت زیر تعیین می شوند:

بازنمایی مولفه های اصلی از مجموعه‌ای از تصاویر دارای میانگین صفر در X بر اساس P_m به صورت $R_m = X_{rec} \times P_m$ تعریف شده است. تقریب کمینه خطای مربعی X از رابطه $X_{rec} = R_m \times P_m^T$ بدست می‌آید.

الگوریتم ICA ماتریس $W_I = W \times W_z$ را چنان تولید می‌کند که

$$W_I \times P_m^T = U \Rightarrow P_m^T = W_I^{-1} U \tag{۱-۵}$$

بنابراین

$$X_{rec} = R_m \times P_m^T \Rightarrow X_{rec} = R_m \times W_I^{-1} U \tag{۲-۵}$$

که در آن W_z ماتریس وایت‌نی‌تگ است. بنابراین سطرهای $W_I^{-1} * R_m$ حاوی ضرایب ترکیب خطی منابع مستقل آماری U می باشد که X_{rec} را تشکیل داده است، که X_{rec} تقریب کمینه خطای مربعی X است، مانند آنچه در PCA داشتیم. بنابراین بازنمایی مولفه های مستقل تصاویر چهره بر اساس مجموعه‌ای از m تصویر ویژگی مستقل آماری U ، توسط سطرهای ماتریس زیر داده شده است.

$$B = R_m \times W_I^{-1} \tag{۳-۵}$$

یک بازنمایی برای تصاویر آزمایشی بوسیله‌ی استفاده از بازنمایی مولفه‌های اصلی براساس تصاویر آموزش داده شده برای بدست آوردن $R_{test} = X_{test} \times P_m$ فراهم آمده است، و سپس $B_{test} = R_{test} \times W_I^{-1}$ محاسبه شده است.

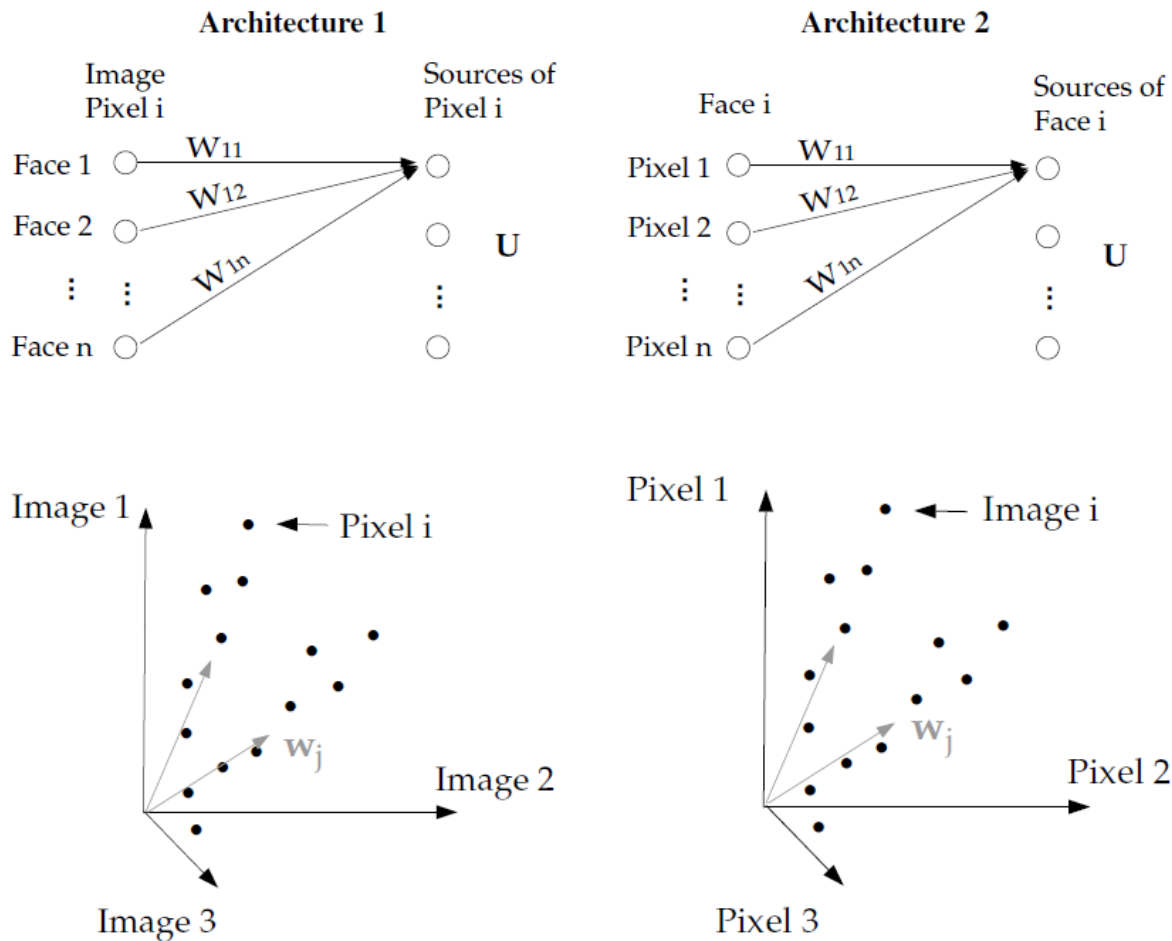
۵-۳-۲-۲-استقلال در فضای چهره در مقابل فضای پیکسل

تحلیل‌ها در بخش ۵-۳-۲-۱ تصاویر پایه مستقل آماری را تولید می‌کند. الگوریتم ICA تصاویر را در میان محل پیکسل‌ها تفکیک می‌کند (شکل ۵-۵ چپ بالا). هر محل پیکسل یک مشاهده است که برای هر یک از چهره‌ها، مقادیر متفاوتیاز سطوح خاکستری را اختیار می‌کند. این امر در شکل ۵-۵ نمایش داده شده است (چپ پایین) که در آن هر پیکسل برای تصویر هر چهره براساس مقدار خاکستری رسم می‌شود. روش ICA در معماری اول، بردارهای وزن را در جهت مستقل‌های آماری جمعیت تصاویر چهره‌ها بر اساس محل پیکسل‌ها پیدا می‌کند. افکنش داده‌ها روی این وزن‌ها مجموعه‌ای از تصاویر مستقل را تولید می‌کند، در حالی که شدت روشنایی پیکسل‌ها در یک تصویر نمی‌تواند از مقادیر شدت روشنایی دیگر تصاویر پیشبینی شود. این تصاویر مستقل فضای تصاویر چهره را پوشش می‌دهد، و هر چهره با ضریبی برای ترکیب خطی این الگوی تصویرهای مستقل که شامل تصویر هر چهره است بازنمایی می‌شود.

تشابه بازنمایی پایه‌های تحلیل مولفه‌های مستقل (معماری ۱) با تحلیل مولفه‌های اصلی اینست که چهره‌های یکه از لحاظ شدت روشنایی پیکسل‌ها ناهمبسته اند. تحلیل مولفه‌های اصلیحهره‌ها را به عنوان ترکیب خطی از الگوی تصاویر غیر همبسته نمایش می‌دهد، در حالی که ICA چهره‌ها را به عنوان ترکیب خطی تصاویر ناهمبسته کلیشه نشان می‌دهد. چهره‌های یگهناهمبسته اند زیرا: ۱- آن‌ها طبق تعریف متعامدند، و ۲- خود چهره‌های یکه مجموعه‌ای از ضرایب PCA هستند (با فرض اینکه هر محل پیکسل به عنوان یک مشاهده و هر تصویر چهره یک اندازه‌گیری باشد). فرض می‌شود که \vec{v}_i معرف بردارهای یکه از $X^T X$ باشد. بردارهای یکه \vec{u}_i از XX^T را می‌توان از طریق $X\vec{v}_i$ بدست می‌آید.

اگرچه تصاویر پایه‌ای که در معماری ۱ بدست آمده‌اند از لحاظ مکانی مستقل می‌باشند، ولی ضرایبی که هر چهره را کد کرده مستقل نیستند. با تغییری در معماری ۱ تحلیل مولفه مستقل، بازنمایی دیگری تعریف می‌شود که در آن ضرایب، به صورت آماری مستقل می‌باشند. به گفته‌ای دیگر دومین معماری ICA کد فاکتوریل^{۳۲۵} برای تصاویر چهره‌ها یافته است. تغییر در معماری با ترانهاده کردن ورودی متناظر است (شکل ۵-۵ راست بالا). با هر تصویر چهره به عنوان یک مشاهده که توسط مقادیر شدت روشنایی در محل هر پیکسل کدگذاری شده است، برخورد می‌شود. در معماری ۲، ICA، بردارهای وزن را در جهت استقلال آماری کدگذاری چهره‌ها در میان جمعیت چهره‌ها پیدا می‌کند. افکنش داده‌ها روی این وزن‌ها مجموعه‌ای از متغیرهای کدگذاری مستقل را بجای محل پیکسل‌ها تولید می‌کند، بطوری که مقدار هر متغیر کدگذاری از روی سایر متغیرهای کدگذاری قابل پیشبینی نمی‌باشد. هر چهره توسط مقداری که از این مجموعه‌ی جدید متغیرهای مستقل کدگذاری برداشته شده، بازنمایی می‌شود (شکل ۵-۶).

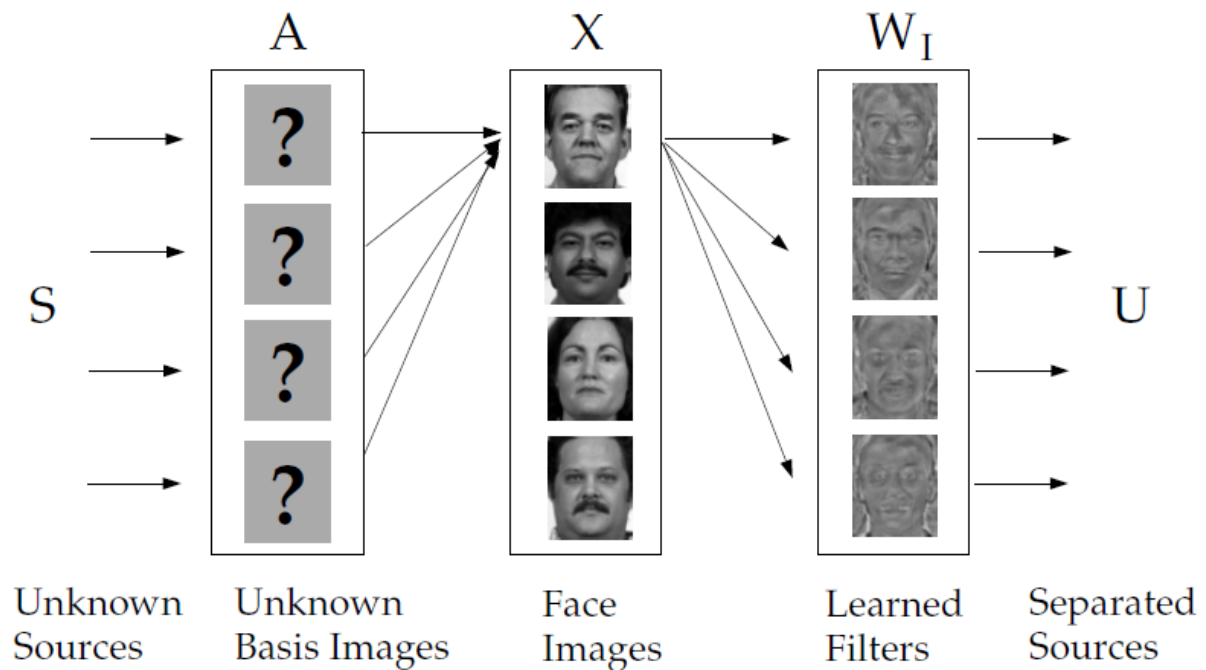
تشابه بازنمایی فاکتوریل تحلیل مولفه‌های مستقل (معماری ۲) نیز با بازنمایی تحلیل مولفه‌های اصلی مستقیم است. ضرایب تحلیل مولفه اصلی یک کد چهره‌ناهمبسته را تشکیل می‌دهند، در حالی که ضرایب معماری ۲ یک کد چهره مستقل را تشکیل می‌دهند.



شکل ۵-۵: دو مدل برای اعمال ICA بر روی تصاویر. (چپ) یک معماری برای یافتن تصاویر پایه مستقل آماری (بالا چپ) اجرای تجزیه منابع بر روی تصاویر چهره مولفه های مستقل در سطرهاى U تولید می کند. (پایین چپ) سطح خاکستری در موقعیت پیکسل برای هر تصویر چهره رسم شده است. در معماری نخست، ICA بردارهای وزن در جهت های استقلال آماری بین موقعیتهای پیکسل پیدا می شود. (راست) معماری یافتن کد فاکتوریل (بالا راست) اجرای تجزیه منابع بر روی پیکسلها یک کد جدا پذیر در ستونهای ماتریس خروجی U تولید می کند. (پایین راست) هر تصویر چهره با توجه به سطوح خاکستری که در هر موقعیت پیکسل اخذ شده است رسم شده است. در معماری دوم، ICA بردارهای وزن را در راستای استقلال آماری بین تصاویر چهره پیدا می کند.

۵-۳-۲-۳-۵- کد چهره فاکتوریل

یک کد فاکتوریل چهره با اجرای تفکیک منابع روی تصاویر چهره تحت معماری ۲ بدست می آید. تغییر در معماری مطابق ترانهاده کردن ماتریس ورودی X است و مانند آنست که تصاویر در ستونها باشند و پیکسلها در سطرها (شکل ۵-۵ راست). تحت این معماری، فیلترها (سطرهاى W_I) تصاویر هستند که بالطبع ستونهای $A = W_I^{-1}$ می باشند. ستونهای A یک مجموعه ی جدید از تصاویر پایه برای چهره ها، و ضرایب آنها برای بازسازی هر چهره (که در ستونهای خروجی های ICA، U ، موجود است) را تشکیل می دهند.



شکل ۵-۶: مدل ساخت چهره با معماری ۲ بر مبنای (Olshausen & Field 96) و (Bell & Sejnowski 97). هر تصویر در مجموعه تصاویر به عنوان یک ترکیب خطی از تصاویر پایه ماتریس به حساب می آید. تصاویر پایه هر یک به مجموعه ای از عوامل مستقل منتسب می شود که بردار ضرایب آن در A آمده است. این عوامل با استفاده از ماتریسی از فیلترهای یاد گرفته شده S بازیابی می شود که تلاش می کند که تابعهای پایه مجهول را برای تولید خروجیهای مستقل آماری U معکوس می شود.

معماری ۲ به تصویر مدل ترکیبی در شکل ۵-۷ وابسته است. این مدل همانند مدل موجود در شکل ۵-۳ می باشد، به جز این که اکنون فرض میشود که چهره ها از مجموعه ای از ضریب های مستقل، S ، برای یک مجموعه از تصاویر پایه A ، متشکل شده اند در حالی که در مدل موجود در شکل ۵-۳ به صورت دیگری بود: منابع مستقل S تصاویر پایه بودند، و ضرایب در A قرار داشتند. این مدل براساس تصویر مدل ترکیبی از اولشاسن و فیلد (Olshausen & Field) می باشد، و همچنین توسط بل و سجویسکی (Bell & Sejnowski 97) برای پیدا کردن فیلترهایی که از صحنه های طبیعی خروجی های مستقل آماری تولید کند، به کار رفته است. تلاش الگوریتم ICA بر این است که منبع ضرایب را بوسیله ی یافتن مجموعه ای از صافی ها W_I ، که خروجی های مستقل آماری U را تولید می کنند، بازیابی کند.

$$\text{Face Image} = u_1 * a_1 + u_2 * a_2 + \dots + u_n * a_n$$

شکل ۵-۷: بازنمایی کد فاکتوریل از ضرایب مستقل u تشکیل می شود تا ترکیبی خطی از تصاویر پایه در A که هر تصویر چهره X را تشکیل داده است بدست دهد. $ICA\ Representation = (u_1, u_2, \dots, u_n)$

ستون‌های ماتریس خروجی ICA، $W_I X = U$ یک کد فاکتوریل برای تصاویر آموزشی در X تهیه می‌کند. هر ستون از U شامل ضرایب تصاویر پایه درون A برای بازسازی هر تصویر در X می‌باشد (شکل ۵-۷). کد نمایش برای تصاویر آزمایشی بوسیله $W_I X_{test,i} = U_{test,i}$ پیدا می‌شود، در حالی که X_{test} ماتریس دارای میانگین صفر تصاویر آزمایشی می‌باشد، و W_I ماتریس وزن، که توسط اجرای ICA بر روی تصاویر آموزشی پیدا شده می‌باشد.

این دو روش بازنمایی امضاهایی را برای تصویر بدست می‌دهند. همانگونه که گفته شد امضای یک تصویر هنگامی که بر روی بردارهای مشخص شده توسط این روشها افکنده می‌شود بسیار شبیه سایر چهره هاست (به طوری که حتی می‌توان توسط آن صاحب چهره را نیز شناسایی کرد) ولی در مورد تصاویر غیر چهره اینگونه نیست. برای آشکارسازی چهره به کمک این دو معماری ابتدا تمام تکه‌های تصویر استخراج می‌شود و امضاهای آن با افکنش بر روی فضای مربوط بدست می‌آید. سپس با استفاده از روش نزدیکترین همسایه روی امضای تکه تصاویری که از چهره‌ها و غیرچهره‌های متنوعی موجود است درباره چهره بودن یا نبودن آن تصمیم‌گیری می‌شود.

۵-۳-۳- مقایسه روشها

برای مقایسه روشها چند آزمایش ترتیب داده می‌شود.

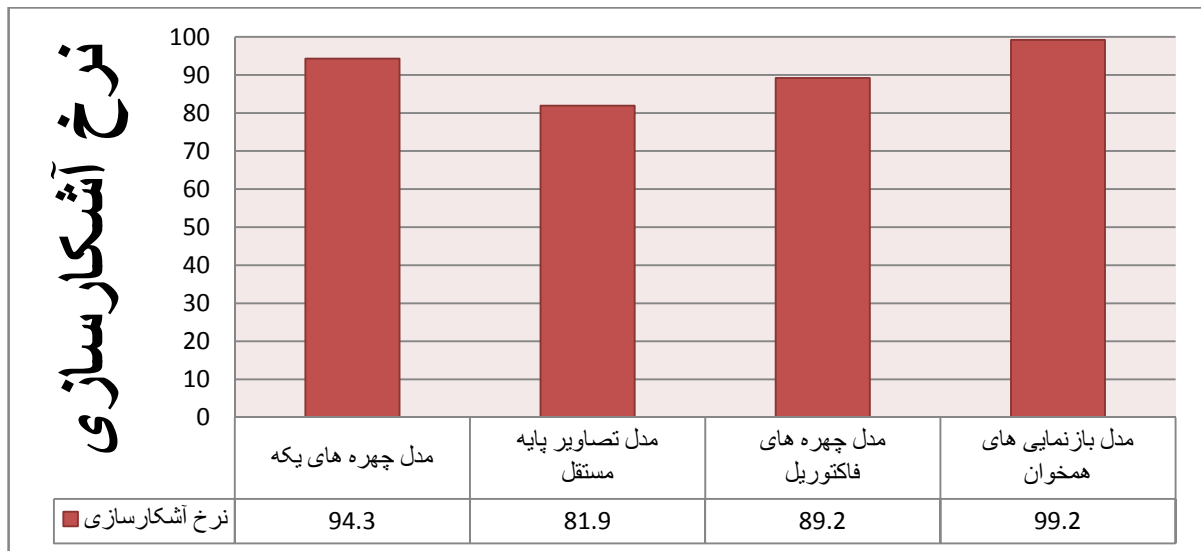
۵-۳-۳-۱- آشکارسازی چهره تمام رخ

در این آزمایش از مجموعه داده‌های AT&T (Olivetti) Database استفاده می‌شود که شامل ۴۰ فرد مختلف است که از هر کدام از آنها ۱۰ تصویر تمام رخ موجود است (شکل ۵-۸).



شکل ۵-۸: نمونه‌ای از مجموعه داده‌های AT&T (Olivetti) Database

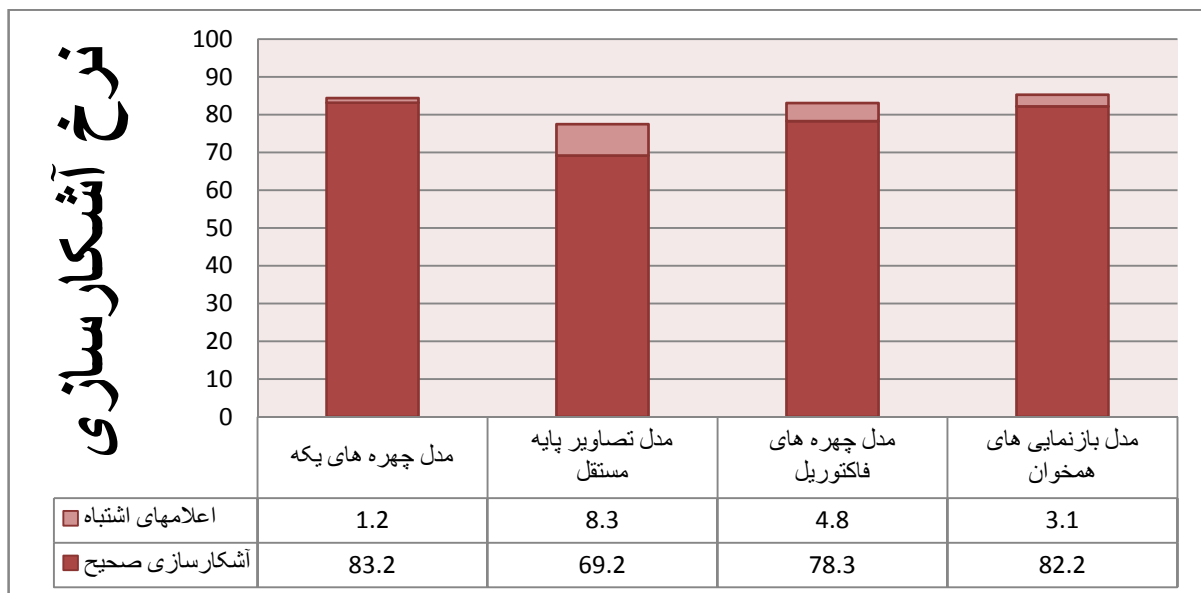
در این آزمایش از ۱۵۰ تصویر نیم‌رخ که وجود دارد استفاده نمی‌شود. هر چهار مدل در آموزش خود تنها به یک تصویر از ۱۰ تصویر هر فرد (علاوه بر مجموعه آموزشی FERET که برای هر چهار استفاده شده است) دسترسی داشته‌اند. هدف از این آزمایش مقایسه نرخ آشکارسازی چهره در تصویری است که تنها یک چهره تمام رخ تمام فضای آنرا پوشانده است. نمودار شکل ۵-۹ نتیجه هر چهار مدل را مقایسه می‌کند.



شکل ۵-۹: نتیجه هر چهار مدل برای آزمایش صحت آشکارسازی. دوران مختصر چهره و مشخص نبودن سوراخ بینی در برخی از تصاویر و بالا بودن آستانه را می توان علت آشکار نشدن برخی از چهره ها دانست. آستانه مدل پیشنهادی برای حفظ عدالت مانند آزمایش بعدی تنظیم شده است و گرنه با پایین رفتن آن نرخ آشکارسازی بالاتر هم خواهد رفت.

۵-۳-۲- آشکارسازی چهره تمام رخ با حضور غیر چهره ها

این آزمایش با شرایطی مشابه آزمایش بخش ۴-۵-۱ صورت می گیرد. هر چهار مدل در آموزش خود به هیچ یک از تصاویر افراد (بجز مجموعه آموزشی FERET که برای هر چهار استفاده شده است) دسترسی نداشته اند.



شکل ۵-۱۰: نتیجه هر چهار مدل برای آزمایش دقت آشکارسازی. مدل چهره های یکه هم از لحاظ دقت آشکارسازی و هم کمی اعلامهای اشتباه برتر از بقیه روشهاست. روش پیشنهادی نرخ آشکارسازی نسبتاً مشابهی دارد ولی به دلیل اعلامهای اشتباه (اغلب اشتباهاً مثبت) از کیفیت آن کاسته شده است. البته با کاهش پارامتر آستانه می توان از این اعلامهای اشتباه جلوگیری کرد ولی از نرخ

آشکارسازی نیز کاسته خواهد شد.

هدف از این آزمایش مقایسه نرخ آشکارسازی چهره در تصویری است یک چهره تمام رخ در حالت عادی روزمره و در پس زمینه ای غیر ساده قرار دارد. نمودار شکل ۵-۱۰ نتیجه هر چهار مدل را مقایسه می کند.

۵-۴- کارهای آتی

چنانکه تمامی تحقیقات علمی برمی تابد، پژوهش حاضر نیز کامل نیست و نویسندگان طی مراحل آزمایش و نگارش به ایده‌هایی جهت ادامه تحقیقات در این زمینه دست یافته است که ذکر آنها وظیفه هر پژوهشگری به شمار می‌رود. این ایده‌ها در چند دسته طبقه‌بندی می شوند که در ادامه آورده شده‌اند:

- میانگین توزیع درست‌نمایی صفر فرض شده است. این ساده‌سازی برای راحت کردن محاسبات بسیار مطلوب است و می‌توان در کارهای آینده به آن پرداخت.
- تابع احتمال اولیه متقارن به طور ضمنی فرض می‌کند که برای یک الگوی همبستگی داده شده، ضرایب مثبت و منفی دارای احتمال یکسان هستند؛ و آشکارتر از آن فرض می‌کند که هر دو این مجموعه ضرایب های مثبت و منفی یک مجموعه بهینه از مؤلفه های کواریانس را می‌سازد. چنین فرضی درباره تقارن در مدل‌های خطی نیز وجود داشته است، ولی در این مدل ساختار تصویر که توسط دو قطب ضرایب نشان داده می‌شوند بسیار متفاوتند. بنابراین مدل تقارنی را فرض می‌کند که ممکن است در داده‌ها وجود نداشته باشد. در واقع مدل هیچ راهی برای صفر نگاه داشتن یکی از ضرایب ندارد و بازنمایی یادگرفته شده ممکن است برای داده‌های آموزش بهینه نباشد. به همین سبب بررسی بیشتر روی توابع احتمال اولیه مناسب، لازم به نظر می‌رسد.
- علاوه بر روش به کار برده شده در این پایان‌نامه، تقریبهای سریع دیگری برای به توان رساندن ماتریس موجود است که به کمک آنها می‌توان درست‌نمایی را از روشهایی سریعتر یا دقیقتر نیز تقریب زد.
- برای پیاده سازی روش به شکل شبکه عصبی مطلوب است که راه حل را با یک شیب نوردی صعودی تکراری با گامهای پیشخور تقریب زد. این کار بخصوص وقتی مفید است که بخواهیم استنتاج بسیار سریع انجام شود (مثلاً در کاربردهای برخط) و ارتباط مستقیم تری بین محاسبات مدل و پردازش نورونی پیشخوری سریع اطلاعات بینایی برقرار باشد. می‌توان بجای استفاده از تقریب MAP از روشهای نمونه‌گیری مانند روش زنجیره مارکوفی مونت کارلو استفاده کرد و با آنها توزیع ثانویه هم متغیرهای نهفته و هم پارامترهای مدل را تقریب زد. تمام تحلیل‌های موجود در این پژوهش را می‌توان با داشتن تمام مقادیر احتمال ثانویه (نه فقط تخمینی از مقدار بیشینه آن) قوت بخشید. این روش جایگزین به کارهای آینده سپرده می‌شود.
- برای بدست آوردن تخمین‌های درست‌نمایی بیشینه از شیب نوردی صعودی استفاده شده است. همانند تخمین متغیرهای نهفته، می‌توان از تعریف رسمی گرادینان استفاده کرد که به تجزیه طیفی A در هر تکرار نیازمند است، و یا می‌توان از تقریب سریعتری که روی بسط سریها استوار است بهره برد. این آزمایشها همچنان جای کار دارد.
- اثبات نظری اینکه تقریبهای مبتنی بر نگاشت مستقل و محاسبه واریانس لگاریتمی، مقادیر MAP را بسیار بهتر از تخمین کواریانس لگاریتمی تک نمونه ای تقریب می‌زند نیز به کارهای آینده موکول شده است.

- یکی از ضعفهای مدل مؤلفه های کواریانسی این فرض است که توزیع شرطی داده ها گوسی است در حالی که مدل واریانسی عمومی تر است و می تواند شامل توزیعهای متنوعی برای ضرایب خطی باشد و نیز می توان لگاریتم پارامتر مقیاس را مدل کرد و مقادیر مطلوب را تخمین زد. چگونگی بهبود این فرض زمینه مناسبی برای تحقیق به شمار می آید.
- تمام تکه های تصویر وایتنینگ می شوند تا همبستگی عمومی از بین برود و واریانس آن نرمال شود. این امر باعث می شود که مدل تنها انحرافات توزیع هر تصویر از آمارگان اصلی را فارغ از ساختار کواریانس سفید حذف کند. آیا اینکار واقعاً در عمل صورت می گیرد؟
- اهداف محاسباتی و قیود نورونی دیگری که برخی از آنها به کاهش افزونگی مرتبطند در روشهای دیگر به تصاویر طبیعی اعمال شده اند. مدل‌های خطی که هزینه انرژی را کمینه می کنند، فعالیت نورونها را ملزم می کنند که در طول زمان به نرمی تغییر کنند، یا شبه تعامد یک پایه فراکامل را بیشینه می کنند نیز به مؤلفه های راستادار و محلی مشابهی منجر می شوند. همچنین معیار پراکندگی که تعبیه آن در این پیاده سازی موفق بوده است چنین مؤلفه هایی را می سازد. مطالعات و آزمایشات در این زمینه نیز مناسب به نظر می رسد.
- در کارهای آینده می توان با استفاده از چارچوب یادگیری بازنمایی تکاملی، سیر تکامل وزنه‌های لایه VI را شبیه سازی کرد و از معیارهایی مانند کورتوسیس، نگانتروپی، اطلاعات متقابل، پراکندگی، درستنمایی استفاده شده در این پژوهش و ... به عنوان تابع برازش استفاده کرد.
- پیاده سازی موازی شبکه عصبی حاصل برای استفاده گسترده از بازنمایی بدست آمده در این پایان نامه لازم به نظر می رسد. ذات پردازش موازی سیستم های عصبی در این زمینه کمک شایانی کرده است.

مراجع

- Abramovich, F., T. Sapatinas, et al. (1998). "Wavelet thresholding via a Bayesian approach." Journal of the Royal Statistical Society: Series B (Statistical Methodology) **60**(4): 725-749.
- Adams, N. and C. Williams (2003). "Dynamic trees for image modelling." Image and Vision Computing **21**(10): 865-877.
- Adelson, E. and J. Bergen (1985). "Spatiotemporal energy models for the perception of motion." Journal of the Optical Society of America A **2**(2): 284-299.
- Anderson, C. and G. DeAngelis (2004). "Population codes and signal to noise ratios in primary visual cortex." Society for Neuroscience Abstract: 822.
- Andrews, D. and C. Mallows (1974). "Scale mixtures of normal distributions." Journal of the Royal Statistical Society. Series B (Methodological) **36**(1): 99-102.
- Asai, M., M. McAleer, et al. (2006). "Multivariate stochastic volatility: a review." Econometric Reviews **25**(2): 145-175.
- Atick, J. (1992). "Could information theory provide an ecological theory of sensory processing?" Network: Computation in Neural Systems **3**(2): 213-251.
- Atick, J. and A. Redlich (1990). "Towards a theory of early visual processing." Neural Computation **2**(3): 308-320.
- Atick, J. and A. Redlich (1992). "What does the retina know about natural scenes?" Neural Computation **4**(2): 196-210.
- Atick, J. and A. Redlich (1993). "Convergent algorithm for sensory receptive field development." Neural Computation **5**(1): 45-60.
- Attneave, F. (1954). "Some informational aspects of visual perception." Psychological review **61**(3): 183-193.
- Barlow, H. (1961). "Possible principles underlying the transformation of sensory messages." Sensory Communication: 217-234.
- Barlow, H. (1989). "Unsupervised Learning." Neural Computation **1**: 295-311.
- Barlow, H. (1994). What is the computational goal of neocortex? Large scale neuronal theories of the brain. C. Koch. Cambridge, MA, MIT Press: 1-22.
- Barlow, H. (2001). "Redundancy reduction revisited." Network: Computation in Neural Systems **12**(3): 241-253.
- Barnard, J., R. McCulloch, et al. (2000). "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage." Statistica Sinica **10**(4): 1281-1312.
- Bartlett, M. (1998). Face image analysis by unsupervised learning and redundancy reduction. Cognitive Science and Psychology. San Diego, CA, University of California. **PhD**: 151.
- Bartlett, M., J. Movellan, et al. (2002). "Face recognition by independent component analysis." IEEE Transactions on neural networks **13**(6): 1450-1464.
- Baylor, D., B. Nunn, et al. (1987). "Spectral sensitivity of cones of the monkey *Macaca fascicularis*." The

- Journal of Physiology **390**(1): 145.
- Bell, A. and T. Sejnowski (1997). "The "independent components" of natural scenes are edge filters." Vision research **37**(23): 3327-3338.
- Berkes, P. and L. Wiskott (2005). "Slow feature analysis yields a rich repertoire of complex cell properties." Journal of Vision **5**(6): 579-602.
- Boik, R. (2002). "Spectral models for covariance matrices." Biometrika **89**(1): 159.
- Bollerslev, T., K. Engle, et al. (1994). ARCH Models. B. Engle and D. McFadden. Handbook of Econometrics VI & V, North-Holland: Amsterdam.
- Bonds, A. (1989). "Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex." Visual Neuroscience **2**(01): 41-55.
- Boynton, G. and J. Hegd  (2004). "Visual cortex: The continuing puzzle of area V2." Current Biology **14**(13): R523-R524.
- Brehm, H. and W. Stammers (1987). "Description and generation of spherically invariant speech-model signals." Signal Processing **12**(2): 119-141.
- Buccigrossi, R. and E. Simoncelli (1999). "Image compression via joint statistical characterization in the wavelet domain." IEEE Transactions on image processing **8**(12): 1688-1701.
- Buchsbaum, G. and A. Gottschalk (1983). "Trichromacy, opponent colours coding and optimum colour information transmission in the retina." Proceedings of the Royal Society of London. Series B, Biological Sciences **220**(1218): 89-113.
- Burt, P. and E. Adelson (1983). "The Laplacian pyramid as a compact image code." IEEE Transactions on communications **31**(4): 532-540.
- Cadieu, C., M. Kouh, et al. (2007). "A model of V4 shape selectivity and invariance." Journal of Neurophysiology **98**(3): 1733-1750.
- Campbell, F. and J. Robson (1968). "Application of Fourier analysis to the visibility of gratings." The Journal of Physiology **197**(3): 551.
- Carandini, M. (2004). "Receptive fields and suppressive fields in the early visual system." The cognitive neurosciences: 313.
- Carandini, M., D. Heeger, et al. (1997). "Linearity and normalization in simple cells of the macaque primary visual cortex." Journal of Neuroscience **17**(21): 8621.
- Cardoso, J. (1989). Source separation using higher order moments. Int'l Conference of Acoustic Speech Signal Processing, Glasgow, IEEE.
- Cavanaugh, J., W. Bair, et al. (2002). "Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons." Journal of Neurophysiology **88**(5): 2530.
- Chang, S., B. Yu, et al. (1998). Spatially adaptive wavelet thresholding with context modeling for image denoising. Fifth IEEE Int'l Conf on Image Processing, Chicago, IEEE Computer Society.

- Chen, S., D. Donoho, et al. (2001). "Atomic decomposition by basis pursuit." SIAM review **43**(1): 129-159.
- Chen, X., F. Han, et al. (2007). "Excitatory and suppressive receptive field subunits in awake monkey primary visual cortex (V1)." Proceedings of the National Academy of Sciences **104**(48): 19120.
- Chichilnisky, E. (2001). "A simple white noise analysis of neuronal light responses." Network: Computation in Neural Systems **12**(2): 199-213.
- Chipman, H., E. Kolaczyk, et al. (1997). "Adaptive Bayesian wavelet shrinkage." Journal of the American Statistical Association **92**(440): 1413-1421.
- Chiu, T., T. Leonard, et al. (1996). "The Matrix-Logarithmic Covariance Model." Journal of the American Statistical Association **91**(433).
- Comon, P. (1992). Independent component analysis. Int'l signal processing workshop on higher-order statistics. Chamrousse, France: 111-120.
- Connor, C., S. Brincat, et al. (2007). "Transformation of shape information in the ventral pathway." Current opinion in neurobiology **17**(2): 140-147.
- Cottrell, G. and J. Metcalfe (1991). Face, gender and emotion recognition using holons. Advances in neural information processing systems. D. Tourezky. San Mateo, CA, Morgan Kaufmann. **3**: 564-571.
- Croner, L. and E. Kaplan (1995). "Receptive fields of P and M ganglion cells across the primate retina." Vision research **35**(1): 7-24.
- Crouse, M., R. Nowak, et al. (1998). "Wavelet-based statistical signal processing using hidden Markov models." IEEE Transactions on Signal Processing **46**(4): 886-902.
- Dai, Y. and Y. Nakano (1996). "Face-texture model based on SGLD and its application in face detection in a color scene." Pattern Recognition **29**(6): 1007-1017.
- Daniels, M. and R. Kass (1999). "Nonconjugate Bayesian Estimation of Covariance Matrices and Its Use in Hierarchical Models." Journal of the American Statistical Association **94**(448).
- Daugman, J. (1985). "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters." Journal of the Optical Society of America A **2**(7): 1160-1169.
- Daugman, J. (1988). "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression." Acoustics, Speech and Signal Processing, IEEE Transactions on **36**(7): 1169-1179.
- Daugman, J. (1989). "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields." Biomedical Engineering, IEEE Transactions on **36**(1): 107-114.
- David, S. and J. Gallant (2005). "Predicting neuronal responses during natural vision." Network: Computation in Neural Systems **16**(2-3): 239-260.
- Dayan, P., G. Hinton, et al. (1995). "The helmholtz machine." Neural Computation **7**(5): 889-904.
- De Boer, E. and P. Kuyper (1968). "Triggered correlation." Biomedical Engineering, IEEE Transactions on **3**(3): 169-179.

- De Valois, R., D. Albrecht, et al. (1982). "Spatial frequency selectivity of cells in macaque visual cortex." Vision research **22**(5): 545-559.
- Dempster, A. (1972). "Covariance selection." Biometrics **28**(1): 157-175.
- Deriugin, N. (1957). "The power spectrum and the correlation function of the television signal." Telecommunications **1**(7): 1-12.
- Derrington, A. and P. Lennie (1984). "Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque." The Journal of Physiology **357**(1): 219.
- Dhingra, N. and R. Smith (2004). "Spike generator limits efficiency of information transfer in a retinal ganglion cell." Journal of Neuroscience **24**(12): 2914.
- Doi, E., D. Balcan, et al. (2006). "A theoretical analysis of robust coding over noisy overcomplete channels." Advances in neural information processing systems **18**: 307.
- Doi, E. and T. Inui (1999). Self-organization of spatio-chromatic receptive fields in the early visual system by ICA. Technical report of IEICE. Tokyo, Japan, IEICE **NC98-170**: 131-138.
- Doi, E., T. Inui, et al. (2003). "Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes." Neural Computation **15**(2): 397-417.
- Doi, E. and M. Lewicki (2005). Relations between the statistical regularities of natural images and the response properties of the early visual system. SIG P&P, Kyoto, Japan, Japanese Cognitive Science Society.
- Doi, E. and M. Lewicki (2005). Sparse coding of natural images using an overcomplete set of limited capacity units. Advances in neural Information processing systems, The MIT Press.
- Dong, D. and J. Atick (1995). "Statistics of natural time-varying images." Network: Computation in Neural Systems **6**(3): 345-358.
- Duda, R., P. Hart, et al. (2001). Pattern classification (2nd Ed), Wiley.
- Elder, J. and R. Goldberg (2002). "Ecological statistics of Gestalt laws for the perceptual organization of contours." Journal of Vision **2**(4).
- Evans, I. (1965). "Bayesian estimation of parameters of a multivariate normal distribution." Journal of the Royal Statistical Society. Series B (Methodological): 279-283.
- Fairhall, A., G. Lewen, et al. (2001). "Efficiency and ambiguity in an adaptive neural code." Nature **412**(6849): 787-792.
- Felleman, D. and D. Van Essen (1991). "Distributed hierarchical processing in the primate cerebral cortex." Cerebral cortex **1**(1): 1.
- Field, D. (1987). "Relations between the statistics of natural images and the response properties of cortical cells." Journal of the Optical Society of America A **4**(12): 2379-2394.
- Field, D. (1994). "What is the goal of sensory coding?" Neural Computation **6**(4): 559-601.

- Földiák, P. (1991). "Learning invariance from transformation sequences." Neural Computation **3**(2): 194-200.
- Friedman, J. (1987). "Exploratory projection pursuit." Journal of the American Statistical Association **82**(397): 249-266.
- Friedman, J. and J. Tukey (1971). "A projection pursuit algorithm for exploratory data analysis." Computers, IEEE Transactions on **100**(9): 881-890.
- Gaeta, M. and J. Lacoume (1990). Source separation a priori knowledge: the maximum likelihood solution. EUSIPCO, Barcelona, Elsevier.
- Gallant, J., J. Braun, et al. (1993). "Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex." Science **259**(5091): 100.
- Gallant, J., C. Connor, et al. (1996). "Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey." Journal of Neurophysiology **76**(4): 2718-2739.
- Gawne, T. and J. Martin (2002). "Responses of primate visual cortical V4 neurons to simultaneously presented stimuli." Journal of Neurophysiology **88**(3): 1128.
- Geisler, W., J. Perry, et al. (2001). "Edge co-occurrence in natural images predicts contour grouping performance." Vision research **41**(6): 711-724.
- Harris, M. (2010). Can brain scans show whether people are telling the truth? IEEE Spectrum, IEEE. **47**: 38-43.
- Hastie, T., R. Tibshirani, et al. (2009). Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Ed). Springer.
- Heeger, D. (1992). "Normalization of cell responses in cat striate cortex." Visual Neuroscience **9**(2): 181-197.
- Heeger, D. and J. Bergen (1995). Pyramid-based texture analysis/synthesis. Proc. ACM SIGGRAPH, Association for Computing Machinery.
- Heeger, D., E. Simoncelli, et al. (1996). "Computational models of cortical visual processing." Proceedings of the National Academy of Sciences of the United States of America **93**(2): 623-627.
- Hegd , J. and D. Van Essen (2000). "Selectivity for complex shapes in primate visual area V2." Journal of Neuroscience **20**(5): 61.
- Hegd , J. and D. Van Essen (2003). "Strategies of shape representation in macaque visual area V2." Visual Neuroscience **20**(03): 313-328.
- Hegd , J. and D. Van Essen (2007). "A comparative study of shape representation in macaque visual areas V2 and V4." Cerebral cortex **17**(5): 1100.
- Hinton, G., P. Dayan, et al. (1995). "The "wake-sleep" algorithm for unsupervised neural network." Science **268**(5214): 1158-1161.
- Hinton, G. and Z. Ghahramani (1997). "Generative models for discovering sparse distributed representations." Philosophical Transactions of the Royal Society B: Biological Sciences **352**(1358): 1177.
- Hinton, G., Z. Ghahramani, et al. (2000). Learning to parse images. Advances in neural information

- processing systems. M. Kearns, S. Solla and D. Cohn. Cambridge, MA, MIT Press. **11**: 463-469.
- Hoyer, P. and A. Hyvärinen (2002). "A multi-layer sparse coding network learns contour coding from natural images* 1." Vision research **42**(12): 1593-1605.
- Hoyer, P. and A. Hyvärinen (2000). "Independent component analysis applied to feature extraction from colour and stereo images." Network: Computation in Neural Systems **11**(3): 191-210.
- Huang, J. and D. Mumford (1999). Statistics of natural images and models. Computer Vision and Pattern Recognition, IEEE Computer Society.
- Hubel, D. and T. Wiesel (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." The Journal of Physiology **160**(1): 106.
- Hubel, D. and T. Wiesel (1968). "Receptive fields and functional architecture of monkey striate cortex." The Journal of Physiology **195**(1): 215-243.
- Huber, P. (1985). "Projection pursuit." The Annals of Statistics **13**(2): 435-475.
- Hurri, J. and A. Hyvärinen (2003). "Simple-cell-like receptive fields maximize temporal coherence in natural video." Neural Computation **15**(3): 663-691.
- Hyvärinen, A. and P. Hoyer (2000). "Emergence of phase-and shift-invariant features by decomposition of natural images into independent feature subspaces." Neural Computation **12**(7): 1705-1720.
- Hyvärinen, A. and P. Hoyer (2000). Emergence of topography and complex cell properties from natural images using extensions of ICA. Advances in neural information processing systems (NIPS 99). S. Solla, T. Leen and K. Muller. Cambridge, MA, MIT Press. **12**: 827-833.
- Hyvärinen, A. and P. Hoyer (2001). "A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images." Vision research **41**(18): 2413-2423.
- Hyvärinen, A., P. Hoyer, et al. (2001). "Topographic independent component analysis." Neural Computation **13**: 1527-1558.
- Hyvärinen, A., J. Hurri, et al. (2009). Natural Image Statistics: A probabilistic approach to early computational vision, Springer.
- Hyvärinen, A., J. Hurri, et al. (2003). "Bubbles: a unifying framework for low-level statistical properties of natural image sequences." Journal of Optic Society of America A **20**(7): 1237-1252.
- Hyvärinen, A., J. Karhunen, et al. (2001). Independent Component Analysis, Wiley.
- Hyvärinen, A. and E. Oja (2000). "Independent component analysis: algorithms and applications." Neural Networks **13**(4-5): 411-430.
- Inouye, Y. and T. Matsui (1989). Cumulant based parameter estimation of linear systems. Workshop on higher order spectral analysis, Vail, Colorado.
- Ito, M. and H. Komatsu (2004). "Representation of angles embedded within contour stimuli in area V2 of macaque monkeys." Journal of Neuroscience **24**(13): 3313.

- Jaynes, E. (1978). Where do we stand on maximum entropy. The maximum entropy formalism. R. Levine and M. Tribus. Cambridge, MA: 15-118.
- Jones, H., W. Wang, et al. (2002). "Spatial organization and magnitude of orientation contrast interactions in primate V1." Journal of Neurophysiology **88**(5): 2796.
- Jones, J. and L. Palmer (1987). "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex." Journal of Neurophysiology **58**(6): 1233.
- Jones, J. and L. Palmer (1987). "The two-dimensional spatial structure of simple receptive fields in cat striate cortex." Journal of Neurophysiology **58**(6): 1187.
- Kandel, E., J. Schwartz, et al. (2001). Principles of Neural Science (4th Ed), McGraw-Hill.
- Kanwisher, N., D. Stanley, et al. (1999). "The fusiform face area is selective for faces not animals." Neuroreport **10**(1): 183.
- Karklin, Y. and M. Lewicki (2003). "Learning higher-order structures in natural images." Network: Computation in Neural Systems **14**(3): 483-499.
- Karklin, Y. and M. Lewicki (2003). A Model for learning variance components of natural images on natural images. Advances in neural information processing systems. S. Becker, S. Thrun and K. Obermayer. Cambridge, MA, MIT Press. **2**: 1367-1374.
- Karklin, Y. and M. Lewicki (2005). "A hierarchical Bayesian model for learning nonlinear statistical regularities in nonstationary natural signals." Neural Computation **17**(2): 397-423.
- Karklin, Y. and M. Lewicki (2008). "Emergence of complex cell properties by learning to generalize in natural scenes." Nature **457**(7225): 83-86.
- Kiani, R., H. Esteky, et al. (2007). "Object category structure in response patterns of neuronal population in monkey inferior temporal cortex." Journal of Neurophysiology **97**(6): 4296.
- Knierim, J. and D. van Essen (1992). "Neuronal responses to static texture patterns in area V1 of the alert macaque monkey." Journal of Neurophysiology **67**(4): 961.
- Kobatake, E. and K. Tanaka (1994). "Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex." Journal of Neurophysiology **71**(3): 856.
- Lamme, V. (1995). "The neurophysiology of figure-ground segregation in primary visual cortex." Journal of Neuroscience **15**(2): 1605.
- Lampl, I., D. Ferster, et al. (2004). "Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex." Journal of Neurophysiology **92**(5): 2704-2713.
- Laughlin, S. (1981). "A simple coding procedure enhances a neuron's information capacity." Z. Naturforsch **36**(9-10): 910-912.
- Laughlin, S. and T. Sejnowski (2003). "Communication in neuronal networks." Science **301**(5641): 1870.
- Lee, B., J. Kremers, et al. (1998). "Receptive fields of primate retinal ganglion cells studied with a novel technique." Visual Neuroscience **15**(01): 161-175.
- Lee, J. (1980). "Digital image enhancement and noise filtering by use of local statistics." IEEE Transactions

- on pattern analysis and machine intelligence(2): 165-168.
- Lee, J. (2009). "Digital image enhancement and noise filtering by use of local statistics." Pattern Analysis and Machine Intelligence, IEEE Transactions on(2): 165-168.
- Lee, T., D. Mumford, et al. (1998). "The role of the primary visual cortex in higher level vision." Vision research **38**(15-16): 2429-2454.
- Lennie, P. (1998). "Single units and visual cortical organization." PERCEPTION-LONDON- **27**: 889-936.
- Leonard, T. and J. Hsu (1992). "Bayesian inference for a covariance matrix." The Annals of Statistics **20**(4): 1669-1696.
- Levitt, J., D. Kiper, et al. (1994). "Receptive fields and functional architecture of macaque V2." Journal of Neurophysiology **71**(6): 2517.
- Lewicki, M. (2002). "Efficient coding of natural sounds." Nature neuroscience **5**(4): 356-363.
- Li, Z. (2002). "A saliency map in primary visual cortex." Trends in Cognitive Sciences **6**(1): 9-16.
- Li, Z. and J. Atick (1994). "Toward a theory of the striate cortex." Neural Computation **6**(1): 127-146.
- Linkser, R. (1988). "Self-organization in a perceptual network." Computer **21**(3): 105-117.
- Logothetis, N., J. Pauls, et al. (1995). "Shape representation in the inferior temporal cortex of monkeys." Current Biology **5**(5): 552-563.
- LoPresto, S., K. Ramchandran, et al. (1997). Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework. Data Compression, Snowbird, Utah, Citeseer.
- Lyu, S. and E. Simoncelli (2007). Statistically and perceptually motivated nonlinear image representation. Conference on Human Vision and Electronic Imaging XII, San Jose, CA, SPIE.
- Lyu, S. and E. Simoncelli (2008). "Reducing statistical dependencies in natural signals using radial Gaussianization." Advances in neural information processing systems **21**: 1009-1016.
- Lyu, S. and E. Simoncelli (2009). "Modeling multiscale subbands of photographic images with fields of Gaussian scale mixtures." Pattern Analysis and Machine Intelligence, IEEE Transactions on **31**(4): 693-706.
- Malfait, M. and D. Roose (1997). "Wavelet-based image denoising using a Markov random field a priori model." Image Processing, IEEE Transactions on **6**(4): 549-565.
- Mallat, S. (1989). "A theory for multiresolution signal decomposition: the wavelet representation." IEEE Transactions on pattern analysis and machine intelligence **11**(7): 674-693.
- Malmir, M. and S. Shiry (2009). Class specific redundancies in natural images: a theory of extrastriate visual processing. Int'l Joint Conference on Neural Networks (IJCNN). Atlanta, GA.
- Malmir, M. and S. Shiry (2009). A Model of Angle Selectivity in Area V2 with Local Divisive Normalization. IEEE Symposium on Computational Intelligence for Multimedia Signal and Vision Processing (CIMSVP). Nashville, TN, IEEE.

- MartõÁnez, A. and A. Kak (2001). "PCA versus LDA." IEEE Transactions on pattern analysis and machine intelligence **23**(2).
- McGill-University. (2006). "The Brain from top to bottom." 2010, from <http://thebrain.mcgill.ca>.
- McMahon, M., M. Lankheet, et al. (2000). "Fine structure of parvocellular receptive fields in the primate fovea revealed by laser interferometry." Journal of Neuroscience **20**(5): 2043.
- Mechler, F. and D. Ringach (2002). "On the classification of simple and complex cells." Vision research **42**(8): 1017-1033.
- Mihcak, K., I. Kozintsev, et al. (1999). "Low-complexity image denoising based on statistical modeling of wavelet coefficients." Signal Processing Letters, IEEE **6**(12): 300-303.
- Miller, E. and J. Cohen (2001). "An integrative theory of prefrontal cortex function." Neuroscience **24**(1): 167-202.
- Mishkin, M., L. Ungerleider, et al. (1983). "Object vision and spatial vision: Two cortical pathways." Trends in neurosciences **6**: 414-417.
- Morrone, M., D. Burr, et al. (1982). "Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence." Proceedings of the Royal Society of London. Series B, Biological Sciences **216**(1204): 335-354.
- Moulin, P. and J. Liu (1999). "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors." Information Theory, IEEE Transactions on **45**(3): 909-919.
- Movshon, J., I. Thompson, et al. (1978). "Receptive field organization of complex cells in the cat's striate cortex." The Journal of Physiology **283**(1): 79.
- Movshon, J., I. Thompson, et al. (1978). "Spatial summation in the receptive fields of simple cells in the cat's striate cortex." The Journal of Physiology **283**(1): 53-77.
- Mutch, J. and D. Lowe (2006). Multiclass object recognition with sparse, localized features, IEEE.
- Nadal, J. and N. Parga (1997). "Redundancy reduction and independent component analysis: Conditions on cumulants and adaptive approaches." Neural Computation **9**(7): 1421-1456.
- Najfeld, I. and T. Havel (1995). "Derivatives of the matrix exponential and their computation." Advances in Applied Mathematics **16**(3): 321-375.
- Neisser, U. (1967). Cognitive psychology. Englewood Cliffs, NJ, Prentice Hall.
- Nishino, K., S. Nayar, et al. (2005). "Clustered blockwise PCA for representing visual data." IEEE Transactions on pattern analysis and machine intelligence: 1675-1679.
- Nothdurft, H., J. Gallant, et al. (2000). "Response profiles to texture border patterns in area V1." Visual Neuroscience **17**(03): 421-436.
- Nowlan, S. (1990). Maximum likelihood competitive learning. Neural information processing systems. D. Tourezky. San Mateo, CA, Morgan Kaufmann. **2**: 574-582.

- Olshausen, B. (1996). "Emergence of simple-cell receptive field properties by learning a sparse code for natural images." Nature **381**(6583): 607-609.
- Olshausen, B. (2002). Sparse codes and spikes. Probabilistic models of the brain: Perception and neural function. R. Rao, B. Olshausen and M. Lewicki, MIT Press: 257-272.
- Olshausen, B. and D. Field (1997). "Sparse coding with an overcomplete basis set: A strategy employed by V1?" Vision research **37**(23): 3311-3325.
- Olshausen, B. and D. Field (2005). "How close are we to understanding V1?" Neural Computation **17**(8): 1665-1699.
- Oppenheim, A. and J. Lim (1981). "The importance of phase in signals." Proceedings of the IEEE **69**(5): 529-541.
- Paninski, L. (2003). "Convergence properties of three spike-triggered analysis techniques." Network: Computation in Neural Systems **14**(3): 437-464.
- Park, H. and T. Lee (2005). Modeling nonlinear dependencies in natural images using mixture of Laplacian distribution. Advances in neural information processing systems. L. Saul, Y. Weiss and L. Bottou. Cambridge, MA, MIT Press. **17**: 1041-1048.
- Pasupathy, A. and C. Connor (2001). "Shape representation in area V4: position-specific tuning for boundary conformation." Journal of Neurophysiology **86**(5): 2505.
- Penev, P. and J. Atick (1996). "Local feature analysis: a general statistical theory for object representation." Network: Computation in Neural Systems **7**(3): 477-500.
- Pham, D. and P. Garat (1992). Blind Separation of Mixture of Independent Sources Through a Maximum Likelihood Approach. EUSIPCO.
- Plebe, A. (2007). "A model of angle selectivity development in visual area V2." Neurocomputing **70**(10-12): 2060-2063.
- Plebe, A. and R. Domenella (2007). "Object recognition by artificial cortical maps." Neural Networks **20**(7): 763-780.
- Portilla, J. and E. Simoncelli (2000). "A parametric texture model based on joint statistics of complex wavelet coefficients." International Journal of Computer Vision **40**(1): 49-70.
- Portilla, J., V. Strela, et al. (2001). Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain. 8th IEEE Int'l. Conference on Image Processing, Piscataway, NJ, IEEE Computer Society.
- Portilla, J., V. Strela, et al. (2003). "Image denoising using scale mixtures of Gaussians in the wavelet domain." IEEE Transactions on image processing **12**(11): 1338-1351.
- Pourahmadi, M. (1999). "Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation." Biometrika **86**(3): 677.
- Pourahmadi, M. (2000). "Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix." Biometrika **87**(2): 425.

- Pourahmadi, M. (2004). "Simultaneous Modelling of Covariance Matrices: GLM, Bayesian and Nonparametric Perspectives." Correlated Data Modelling.
- Rao, R., B. Olshausen, et al. (2002). Probabilistic models of the brain: Perception and neural function, The MIT Press.
- Reynolds, J., L. Chelazzi, et al. (1999). "Competitive mechanisms subserve attention in macaque areas V2 and V4." Journal of Neuroscience **19**(5): 1736.
- Riesenhuber, M. and T. Poggio (1999). "Hierarchical models of object recognition in cortex." Nature neuroscience **2**: 1019-1025.
- Ringach, D. (2002). "Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex." Journal of Neurophysiology **88**(1): 455-463.
- Rolls, E., A. Cowey, et al. (1992). "Neurophysiological Mechanisms Underlying Face Processing within and beyond the Temporal Cortical Visual Areas [and Discussion]." Philosophical Transactions: Biological Sciences **335**(1273): 11-21.
- Rolls, E. and S. Stringer (2006). "Invariant visual object recognition: A model, with lighting invariance." Journal of Physiology-Paris **100**(1-3): 43-62.
- Romberg, J., H. Choi, et al. (1999). Bayesian wavelet-domain image modeling using hidden Markov trees. IEEE Int'l Conf on Image Processing, Kobe, Japan, IEEE.
- Romberg, J., H. Choi, et al. (2001). "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models." Image Processing, IEEE Transactions on **10**(7): 1056-1068.
- Romberg, J., H. Choi, et al. (2002). Bayesian wavelet-domain image modeling using hidden Markov trees, IEEE.
- Roorda, A. and D. Williams (1999). "The arrangement of the three cone classes in the living human eye." Nature **397**(6719): 520-522.
- Rossi, A., R. Desimone, et al. (2001). "Contextual modulation in primary visual cortex of macaques." Journal of Neuroscience **21**(5): 1698.
- Rowley, H., S. Baluja, et al. (1998). "Neural Network-Based Face Detection." IEEE Transactions on pattern analysis and machine intelligence **20**(1).
- Ruderman, D. (1996). "The statistics of natural images." Network: Computation in Neural Systems **5**(4): 517-548.
- Ruderman, D. and W. Bialek (1994). "Statistics of natural images: Scaling in the woods." Physical Review Letters **73**(6): 814-817.
- Ruderman, D., T. Cronin, et al. (1998). "Statistics of cone responses to natural images: implications for visual coding." Journal of Optical Society of America **15**: 2036-2045.
- Rust, N. and J. Movshon (2005). "In praise of artifice." Nature neuroscience **8**(12): 1647-1650.
- Rust, N., O. Schwartz, et al. (2005). "Spatiotemporal elements of macaque v1 receptive fields." Neuron **46**(6):

945-956.

- Salinas, E. and L. Abbott (1997). "Invariant visual responses from attentional gain fields." Journal of Neurophysiology **77**(6): 3267.
- Schraudolph, N. and T. Sejnowski (1992). Competitive anti-hebbian learning of invariants. Advances in neural information processing systems. J. Moody, S. Hanson and R. Lippman. San Fransisco, Morgan Kaufmann. **4**: 1017-1024.
- Schwartz, O., J. Pillow, et al. (2006). "Spike-triggered neural characterization." Journal of Vision **6**(4).
- Schwartz, O. and E. Simoncelli (2001). "Natural signal statistics and sensory gain control." Nature neuroscience **4**(8): 819-825.
- Serre, T., L. Wolf, et al. (2007). "Robust object recognition with cortex-like mechanisms." Pattern Analysis and Machine Intelligence, IEEE Transactions on **29**(3): 411-426.
- Shannon, C. (1948). "A mathematical theory of communication." Bell System Technical Journal **27**: 379-423.
- Shannon, C. and W. Weaver (1949). The mathematical theory of communication.
- Shannon, C. E. and W. Weaver (1949). Nonlinear problems in random theory. University of Illinois Press. Urbana.
- Shapiro, J. (1993). "Embedded image coding using zerotrees of wavelet coefficients." Signal Processing, IEEE Transactions on **41**(12): 3445-3462.
- Sigman, M., G. Cecchi, et al. (2001). "On a common circle: Natural scenes and Gestalt rules." Proceedings of the National Academy of Sciences of the United States of America **98**(4): 1935.
- Sillito, A. (1975). "The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the cat." The Journal of Physiology **250**(2): 305.
- Simoncelli, E. (1997). Statistical models for images: Compression, restoration and synthesis. 31st Asilomar Conf on Signals, Systems and Computers, Pacific Grove, CA, IEEE.
- Simoncelli, E. (1999). Bayesian denoising of visual images in the wavelet domain. Bayesian Inference in Wavelet Based Models. P. Muller and B. Vidakovic. New York, NY, Springer-Verlag: 291-308.
- Simoncelli, E. (2005). Statistical modeling of photographic images. Handbook of image and video processing. A. Bovik, Academic Press: 961-974.
- Simoncelli, E. and E. Adelson (1996). Noise removal via Bayesian wavelet coring. Third Int'l Conf on Image Processing, Lausanne, IEEE Sig Proc Society.
- Simoncelli, E. and B. Olshausen (2001). "Natural image statistics and neural representation." Annual review of neuroscience **24**(1): 1193-1216.
- Sirosh, J. and R. Miikkulainen (1997). "Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex." Neural Computation **9**(3): 577-594.
- Smith, E. and M. Lewicki (2005). "Learning efficient auditory codes using spikes predicts cochlear filters." Advances in neural information processing systems **17**: 1289-1296.

- Smith, E. and M. Lewicki (2006). "Efficient auditory coding." Nature **439**(7079): 978-982.
- Smith, M. and R. Kohn (2002). "Parsimonious covariance matrix estimation for longitudinal data." Journal of the American Statistical Association **97**(460): 1141-1153.
- Smyth, D., B. Willmore, et al. (2003). "The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation." Journal of Neuroscience **23**(11): 4746.
- Song, Y. and C. Baker (2007). "Neuronal response to texture-and contrast-defined boundaries in early visual cortex." Visual Neuroscience **24**(01): 65-77.
- Srivastava, A., A. Lee, et al. (2003). "On advances in statistical modeling of natural images." Journal of mathematical imaging and vision **18**(1): 17-33.
- Srivastava, A., X. Liu, et al. (2002). "Universal analytical forms for modeling image probabilities." Pattern Analysis and Machine Intelligence, IEEE Transactions on **24**(9): 1200-1214.
- Stettler, D., A. Das, et al. (2002). "Lateral connectivity and contextual interactions in macaque primary visual cortex." Neuron **36**(4): 739-750.
- Strela, V., J. Portilla, et al. (2000). Image denoising using a local Gaussian scale mixture model in the wavelet domain. SPIE 45th Annual Meeting, Bellingham, WA, International Society for Optical Engineering.
- Taylor, D., L. Finkel, et al. (2000). "Color-opponent receptive fields derived from independent component analysis of natural images." Vision research **40**(19): 2671-2676.
- Tanaka, K. (1996). "Inferotemporal cortex and object vision." Annual review of neuroscience **19**(1): 109-139.
- Tolhurst, D., Y. Tadmor, et al. (1992). "Amplitude spectra of natural images." Ophthalmic and Physiological Optics **12**(2): 229-232.
- Torralba, A. and A. Oliva (2003). "Statistics of natural image categories." Network: Computation in Neural Systems **14**(3): 391-412.
- Touryan, J., G. Felsen, et al. (2005). "Spatial structure of complex cell receptive fields measured with natural images." Neuron **45**(5): 781-791.
- Turiel, A., G. Mato, et al. (1998). "Self-similarity properties of natural images resemble those of turbulent flows." Physical Review Letters **80**(5): 1098-1101.
- Turiel, A. and N. Parga (2000). "The multifractal structure of contrast changes in natural images: from sharp edges to textures." Neural Computation **12**(4): 763-793.
- Turk, J. and A. Pentland (1991). "Eigenfaces for recognition." Cognitive Neuroscience **3**(1): 71-86.
- Valpola, H., M. Harva, et al. (2004). "Hierarchical models of variance sources." Signal Processing **84**(2): 267-282.
- Van der Schaaf, A. and J. Van Hateren (1996). "Modelling the power spectra of natural images: statistics and information." Vision research **36**(17): 2759-2770.
- Van Hateren, J. and H. Snippe (2001). "Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells." Vision research **41**(14): 1851-1865.

- Van Hateren, J. and A. Van der Schaaf (1998). "Independent component filters of natural images compared with simple cells in primary visual cortex." Proceedings of the Royal Society B: Biological Sciences **265**(1394): 359.
- Van Steveninck, R. and W. Bialek (1988). "Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences." Proceedings of the Royal Society of London. Series B, Biological Sciences **234**(1277): 379-414.
- Von der Twer, T. and D. MacLeod (2001). "Optimal nonlinear codes for the perception of natural colours." Network: Computation in Neural Systems **12**(3): 395-407.
- Wachtler, T., T. Lee, et al. (2001). "Chromatic structure of natural scenes." JOURNAL-OPTICAL SOCIETY OF AMERICA A **18**(1): 65-77.
- Wainwright, M. and E. Simoncelli (2000). Scale mixtures of Gaussians and the statistics of natural images. Advances in neural information processing systems (NIPS 99). S. Solla, T. Leen and K. Muller. Cambridge, MA, MIT Press. **12**: 855-861.
- Wainwright, M., E. Simoncelli, et al. (2001). "Random cascades on wavelet trees and their use in analyzing and modeling natural images." Applied and Computational Harmonic Analysis **11**(1): 89-123.
- Wang, R. (1999). The Primary Visual Cortex (Striate Cortex, V1), <http://fourier.eng.hmc.edu/e180/lectures/v1/node10.html>.
- Wax, M., T. Shan, et al. (1984). "Spatio-temporal spectral analysis by eigenstructure methods." IEEE Transactions on Acoustics, Speech, and Signal Processing **32**: 817-827.
- Wegmann, B. and C. Zetsche (1990). Statistical dependence between orientation filter outputs used in a human-vision-based image code. SPIE Visual Comm. and Image Processing, Bellingham, WA, SPIE.
- Wong, F., C. Carter, et al. (2003). "Efficient estimation of covariance selection models." Biometrika **90**(4): 809.
- Yang, R. and J. Berger (1994). "Estimation of a covariance matrix using the reference prior." The Annals of Statistics **22**(3): 1195-1211.
- Yen, S., J. Baker, et al. (2007). "Heterogeneity in the responses of adjacent neurons to natural stimuli in cat striate cortex." Journal of Neurophysiology **97**(2): 1326.
- Zetsche, C. (1990). Sparse coding: the link between low level vision and associative memory. Parallel processing in neural systems and computers. R. Eckmiller, G. Hartmann and G. Hauske, Elsevier Science: 273-276.
- Zetsche, C. and E. Barth (1990). "Fundamental limits of linear filters in the visual processing of two dimensional signals." Vision Research **30**(7): 1111-1117.
- Zetsche, C. and F. Rhrbein (2001). "Nonlinear and extra-classical receptive field properties and the statistics of natural scenes." Network: Computation in Neural Systems **12**(3): 331-350.
- Zetsche, C., B. Wegmann, et al. (1993). Nonlinear aspects of primary vision: Entropy reduction beyond decorrelation. Int'l. Symposium of Society for Information Display, Playadel Ray, CA, Society for Information Display.

- Zhang, C. and Z. Zhang (2010). A Survey of Recent Advances in Face Detection. Redmond, WA, Microsoft Corporation.
- Zhu, S. (2003). "Statistical modeling and conceptualization of visual patterns." IEEE Transactions on pattern analysis and machine intelligence: 691-712.
- Zhu, S., Y. Wu, et al. (1997). "Minimax entropy principle and its application to texture modeling." Neural Computation **9**(8): 1627-1660.
- Zhu, S., Y. Wu, et al. (1998). "Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling." International Journal of Computer Vision **27**(2): 107-126.
- Zoccolan, D., D. Cox, et al. (2005). "Multiple object response normalization in monkey inferotemporal cortex." Journal of Neuroscience **25**(36): 8150.
- مالمیر، م. (۱۳۸۸). یادگیری مفهوم در روبات با الهام از مغز انسان. دانشکده مهندسی کامپیوتر و فناوری اطلاعات تهران، دانشگاه صنعتی امیرکبیر
کارشناسی ارشد.

واژه نامه

[واژگان معادل]

۱. واژه نامه

۲. مخفف ها

واژه نامه:

فارسی	English
عصب شناسی	
عصب شناس	neuroscientist
مکانیسم دروازه بندی	gating mechanism
مخچه	cerebellum
سازمان عضلانی	musculature
خروجی حرکتی	motor output
ورودی حسی	sensory input
بیهوشی	epilepsy
ریتم خواب	sleep rythme
باشعور، شناختی	cognitive
ادراکی	perceptual
یادگیری خود نظارتی	self supervised learning
مسیرهای قشری مخ	cortical pathways
پیشخور	feed forward
پسخور	feed back
محرک	stimuli
میدان گیرندگی	receptive field
تداعی موقتی	temporal association
برانگیخته کننده	excitatory
فروشاننده	suppressive
پنجره های سینوسی	sinusoidal gratings
حول مرکز	center surrond
موقتی مکانی	spatio-temporal
مسیر پشتی (خلفی)	dorsal pathway
مسیر شکمی (بطنی)	ventral pathway
حافظه انجمنی	associative memory
راستای مرجح	preferred orientation
تصویربرداری تشدید مغناطیسی	magnetic resonance imaging
ناحیه دوک مانند چهره	fusiform face area
بین سلولی	intracellular
آسیب (به بافت موجود زنده)	lesion
بخش پس سری مغز	occipital lobe
بخش آهیانه ای مغز	parietal lobe
بخش گیجگاهی مغز	temporal lobe
بخش پیشانی مغز	frontal lobe
فوویا	fovea

رابط	conjunctive
شبکیه	retina
قرنیه	cornea
مایع زلالیه	aqueous humor
مردمک	pupil
عنابیه	iris
مایع زجاجیه	vitreous humor
دورنمای بهینه سازی	optimization landscape
تفریق شناختی	cognitive subtraction
عقدۀ زانویی خارجی	lateral geniculate nucleus
کولیکولوس فوقانی	superior colliculus
صلبیه	sclera
عضلات چشمی	ocular muscle
مشیمیه	choroid
رنگدانه	pigment
شکست نور	refraction
رباطهای معلق	suspensory ligaments
ماهیچه مزگانی	ciliary muscle
تطابق (چشم)	accomodation
صفحه بینایی	optic disk
نقطه کور	blind spot
ماکولا (لکه)	macula
تیزبینی	visual acuity
سلولهای عقدۀ ای شبکیه	retina ganglion cells
کیاسمای بینایی	optic chiasm
مجرای بینایی	optic tractus
ساقه مغز	brain stem
هوشیاری	vigilance
مربوط ه توجه	attentiveness
هسته فوق کیاسمایی	superchiasmatic nucleus
مسیر شبکیه ای	retinotectal pathway
ساختار مشبک	reticular formation
کولیکولوس تحتانی	inferior colliculus
هسته بالشتک	pulvinar
پرنکتوم	pretectum
ناحیه انجمنی	associative area
حالت دوگانه (دو مسیر بینایی)	dichotomy
مسیر کنش	action pathway
حبابها و بین حبابها	blobs and interblobs
نواحی فوق مخطط	extrastriate areas
قشر گیجگاهی زیرین	infrottemporal cortex

نواحی فوق گیجگاهی چند سنسوری	superior temporal polysensory
گزینندگی	selectivity
تکنیک ضبط تک نرون	single cell recording technique
بخش پیش پیشانی قشر مخ	prefrontal cortex
نقاط دور از مرکز شبکیه	retinal eccentricity
مدلهای مبتنی بر انرژی	energy-based models
مدلهای وابستگی موقتی	temporal coherence models
تحریک کردن	innervation
سلولهای بزرگ تک قطبی	large monopolar cells
کانون توجه	focus-of-attention
تحریک	stimulate
فرونشانی اطرافیان	surround suppression
بازداری فعالیت به ازای راستاهای متقاطع	cross-orientation inhibition
اشباع پاسخ کنتراست	contrast response saturation
فرونشانی ماسک	mask supression
نرمال کردن با تقسیم	divisive normalization
پردازش تصویر و هوش مصنوعی	
عصب شناس	neuroscientist
مکانیسم دروازه بندی	gating mechanism
مخچه	cerebellum
سازمان عضلانی	musculature
خروجی حرکتی	motor output
ورودی حسی	sensory input
بیهوشی	epilepsy
ریتم خواب	sleep rythme
باشعور، شناختی	cognitive
ادراکی	perceptual
یادگیری خود نظارتی	self supervised learning
مسیرهای قشری مخ	cortical pathways
پیشخور	feed forward
پسخور	feed back
محرک	stimuli
میدان گیرندگی	receptive field
تداعی موقتی	temporal association
برانگیخته کننده	excitatory
فرونشاننده	suppressive
پنجره های سینوسی	sinusoidal gratings
حول مرکز	center surrond
موقتی مکانی	spatio-temporal
مسیر پشتی (خلفی)	dorsal pathway
مسیر شکمی (بطنی)	ventral pathway

حافظه انجمنی	associative memory
راستای مرجح	preferred orientation
تصویربرداری تشدید مغناطیسی	magnetic resonance imaging
ناحیه دوک مانند چهره	fusiform face area
مدلهای تبدیل شونده سه بعدی	3D morphable model
روشهای مبتنی بر ظاهر	appearance-based methods
کلیشه اولیه ارتجاعی	a priori elastic model
مدل فعال شکل	active shape model
مسئله انتساب	assignment problem
حبابها و رگه ها	blobs and streaks
خلوص رنگی	chrominance
بستن (عملگر مورفولوژی)	closing
درشت به ظریف	coarse-to-fine
مسئله مهمانی کوکتیل	cocktail party problem
ثبات رنگ	color constancy
نقشه آشکارسازی	conspicuity map
ستاره ای	constellation
شبکه های عصبی پیچشی	convolutional neural network
انحناء	curvature
کلیشه های انعطاف پذیر	deformable templates
آشکارگر	detector
میزان عدم تطابق	disparity
نقشه عدم توافق	disparity map
متمایز	distinct
تطابق گراف دسته انعطاف پذیر	elastic buch graph matching
لبه	edge
چهره یکه	eigenface
تعقیب تکاملی	evolutionary pursuit
تصدیق هویت چهره	face authentication
تشخیص چهره	face detection
مکانیابی چهره	face localization
شناسایی چهره	face recognition
شناسایی/تعیین هویت چهره	face recognition/identification
ردگیری چهره	face tracking
تشخیص حالت چهره	facial expression recognition
تشخیص ویژگیهای چهره	facial feature detection
اشتباهاً منفی	false negative
اشتباهاً مثبت	false positive
روشهای نامتغیر با ویژگی	feature invariant approaches
تمام رخ	frontal face
خاکستری فام	grayscale

یادگیری هبی	Hebbian learning
الگوریتم اکتشافی	heuristic
هموارسازی هیستوگرام	histogram equalization
مدل مخفی مارکوف	hidden Markov model
یکپارچه	holistic
تکه تصویر	image patch
ناحیه تصویر	image region
تفکیک پذیری تصویر	image resolution
مقادیر شدت روشنایی	intensity values
روشهای مبتنی بر دانش	knowledge-based methods
یکنواخت سازی هیستوگرام محلی	local histogram equalization
سطح روشنایی	luminance
پیکسلهای حاوی بیشترین اطلاعات	most informative pixels
حرکت	motion
تابع انتقال مدولاسیون	modulation transfer function
انسداد	occlusion
مدل جستجوی نقطه ای	point distribution model
تبدیل رادون	radon transform
بازسازی	reconstruct
ناحیه مورد علاقه	region of interest
مقاوم	robust
مقیاس	scale
خودسازماندهی	self organization
شکل	shape
شمایل	silhouette
شبکه تنک غربالها	sparse network of winnows
فیلترهای سوبل	Sobel filters
متضاد مکانی	spatially opponent
پایداری اکید حسی	strict sense stationary
آزمون انسان محور	subjective test
ماشینهای بردار پشتیبان	support vector machine
روشهای تطبیق کلیشه	template matching methods
تبدیل ردگیری	trace transform
تابع انتقال	transfer function
تصدیق	verification
نما	view
تقطیع ویتربی	Viterbi segmentation
محلی بودن	locality
مدلهای تبدیل شونده سه بعدی	3D morphable model
روشهای مبتنی بر ظاهر	appearance-based methods
کلیشه اولیه ارتجاعی	apriori elastic model

مدل فعال شکل	active shape model
مسئله انتساب	assignment problem
حبابها و رگه ها	blobs and streaks
خلوص رنگی	chrominance
بستن (عملگر مورفولوژی)	closing
درشت به ظریف	coarse-to-fine
مسئله مهمانی کوکتیل	cocktail party problem
ثبات رنگ	color constancy
نقشه آشکارسازی	conspicuity map
ستاره ای	constellation
شبکه های عصبی پیچشی	convolutional neural network
انحناء	curvature
کلیشه های انعطاف پذیر	deformable templates
آشکارگر	detector
میزان عدم تطابق	disparity
نقشه عدم توافق	disparity map
متمایز	distinct
تطابق گراف دسته انعطاف پذیر	elastic buch graph matching
لبه	edge
چهره یکه	eigenface
تعقیب تکاملی	evolutionary pursuit
تصدیق هویت چهره	face authentication
تشخیص چهره	face detection
مکانیابی چهره	face localization
شناسایی چهره	face recognition
شناسایی/تعیین هویت چهره	face recognition/identification
ردگیری چهره	face tracking
تشخیص حالت چهره	facial expression recognition
تشخیص ویژگیهای چهره	facial feature detection
اشتباهاً منفی	false negative
اشتباهاً مثبت	false positive
روشهای نامتغیر با ویژگی	feature invariant approaches
تمام رخ	frontal face
خاکستری فام	grayscale
یادگیری هبی	Hebbian learning
الگوریتم اکتشافی	heuristic
هموارسازی هیستوگرام	histogram equalization
مدل مخفی مارکوف	hidden Markov model
یکپارچه	holistic
تکه تصویر	image patch
ناحیه تصویر	image region

تفکیک پذیری تصویر	image resolution
مقادیر شدت روشنایی	intensity values
روشهای مبتنی بر دانش	knowledge-based methods
یکنواخت سازی هیستوگرام محلی	local histogram equalization
سطح روشنایی	luminance
پیکسلهای حاوی بیشترین اطلاعات	most informative pixels
حرکت	motion
تابع انتقال مدولاسیون	modulation transfer function
انسداد	occlusion
مدل جستجوی نقطه ای	point distribution model
تبدیل رادون	radon transform
بازسازی	reconstruct
ناحیه مورد علاقه	region of interest
مقاوم	robust
مقیاس	scale
خودسازماندهی	self organization
شکل	shape
شمایل	silhouette
شبکه تنک غربالها	sparse network of winnows
فیلترهای سوبل	Sobel filters
متضاد مکانی	spatially opponent
پایداری اکید حسی	strict sense stationary
آزمون انسان محور	subjective test
ماشینهای بردار پشتیبان	support vector machine
روشهای تطبیق کلیشه	template matching methods
تبدیل ردگیری	trace transform
تابع انتقال	transfer function
تصدیق	verification
نما	view
تقطیع ویتربی	Viterbi segmentation
محلی بودن	locality
مدلهای تبدیل شونده سه بعدی	3D morphable model
روشهای مبتنی بر ظاهر	appearance-based methods
کلیشه اولیه ارتجاعی	a priori elastic model
علم آمار	
افزونگی	redundancy
هم وقوعی های مشکوک	suspicious coincidences
قاعده مندیهی آماری	statistical regularities
تحلیل مولفه اصلی	principle component analysis
همبستگی	correlation
تحلیل ویژگی محلی	local feature analysis

تحلیل مولفه مستقل	independent component analysis
آنتروپی توأم	joint entropy
اطلاعات متقابل	mutual information
مدل مولد	generative model
حفظ اطلاعات	information preservation
درست‌نمایی	likelihood
توزیع احتمال	probability distribution
احتمال اولیه	prior probability
احتمال ثانویه	posterior probability
شیب نوردی صعودی	gradient ascend
فاکتور مقیاس گذاری	scaling factor
مدل گوسی توأم	mixture of Gaussians
قانون بروزرسانی	update rule
تنزل وزن	weight decay
انفجار ترکیبیاتی	combinatorial explosion
اتفاقی	stochastic
نااریب	unbiased
کد پراکنده	sparse code
امید ریاضی، مقدار مورد انتظار	expected value
کدگذاری مؤثر	efficient coding
بهره	gain
بردار یکه	eigenvector
مقدار یکه	eigenvalue
بردارهای پایه	basis vectors
جمع وزندار	weighted sum
اصل حفظ بیشینه اطلاعات	maximum information preservation principle
نامتغیر بودن	invariance
ناهمبسته سازی	decorrelation
طیف درشتی	amplitude spectrum
طیف توان	power spectrum
خود-همبستگی	autocorrelation
جابجایی	translation
آنتروپی جانبی	marginal entropy
تحلیل عامل	factor analysis
پیگیری افکنش	projection pursuit
متغیر نهفته	latent variable
ماتریس تلفیق	mixing matrix
قضیه حد مرکزی	central limit theorem
همپوشانی	overlap
موجک	wavelet
متعامد یکه	orthonormal

فراکامل	overcomplete
تجزیه موجک متعامد یگه	orthonormal wavelet decomposition
همگنی	homogeneity
نوسان	fluctuation
آشوب	turbulance
مدل مخلوط مقیاس گوسی	Gaussian scale mixture model
ایستا	stationary
ماتریس مثبت-معین	positive definite matrix
ماتریس همانی	identity matrix
تجزیه فرکانسی	spectral decomposition
عنصر به عنصر	element-wise
ماتریس مفرد	singular matrix
رتبه ماتریس	matrix rank
تجزیه مقدار مفرد	singular value decomposition
ماتریس شبه معکوس	pseudo-inverse matrix
فروکامل	undercomplete
توزیع یکنواخت	uniform distribution
توزیع لاپلاسی	Laplacian distribution
توزیع گوسی	Gaussian distribution
توزیع نمایی	exponential disribution
توزیع رایلیخ	Rayleigh distribution
طیف بزرگی	ampliude spectrum
فرکانس مکانی	spatial frequency
آمارگان	statistics
تعامل	interaction
تحلیل جدای پذیری خطی	linear discriminant analysis
افتراق	divergence
واریانس-متغیر	heteroscedastic
کمترین طول توصیف	minimum description length
تناوب کاری	duty cycle
دور بر درجه	cycle per degree
تفکیک کورکورانه منابع	blind source seperation
متعامد کردن	orthogonalization
تبدیل وایتینگ	whitening transform
زبرگوسی	subgaussian, platikurtotic
زبرگوسی	supergaussian, leptokurtotic
نابع احتمال) نوک تیز)	spiky, peaky
دنباله های سنگین	heavy tails
اندازه گیریهای دور از مرکز	outliers
نگانتروپی	negentropy
تبدیل قابل بازگشت	invertible transform

اصل اینفومکس	infomax principle
اصل بیشینه سازی آنتروپی شبکه	network entropy maximization principle
لغات عمومی	
پایین به بالا	bottom up
روش قضاوت	arbiterate scheme
بالا به پایین	top-down
ابرایانه	supercomputer

مخفف ها:

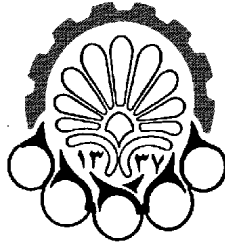
<i>Abbreviation</i>	<i>Full Text</i>
PCA	principle component analysis
LFA	local feature analysis
ICA	independent component analysis
GSM	Gaussian scale mixture
SVD	singular value decomposition
LDA	linear discriminant analysis
cpd	cycle per degree
LGN	lateral geniculate nucleus
IID	independent and identically distributed
EM	expectation maximization
MC	Markove chain Monte Carlo
LMC	large monopolar cells
MTF	modulation transfer function
PDM	point distribution model
ASM	active shape model
FA	factor analysis
SN _o W	sparse network of winnows
HMM	hidden Markov model
EP	evolutionary persuit
EBGM	elastic buch graph matching
MAP	maximum a posteriori

MRI	magnetic resonance imaging
FFA	fusiform face area
SVM	support vector machine
MIP	most informative pixels

Abstract

Encoding the building blocks of a natural image, such as edges, textures and shapes is one of the main functions of a visual system which is the underlying procedure of visual high level tasks such as face detection and natural image understanding. An abstract representation which is generalized over specific samples of input images is the desired output of this encoding. It's now understood that primary visual cortex employs a combination of features extracted from visual data but there is not a prominent theory about forming invariant representation in human brain. Higher-order visual neurons are commonly assumed to use the statistical variations that characterized local image regions. In the proposed model, a search about the most consistent distribution to the input images is conducted and represented using activities of the neurons. This model is feed with natural scenes and human face images and obtains generalization over this training set by learning a compact set of dictionary elements typically found in natural images. The characteristics of model neurons resemble cortical cells, and provide an explanation for nonlinear properties of these cells functionally. Being inspired from coding strategies underlying the processing of primary visual cortex and higher visual areas, the results representation is able to correctly cluster images with similar distribution with is a part of concept formation. Finally a face detection algorithm is proposed which use this coding strategy and other mechanisms inspired from brain neural network. The proposed model is invariant against changes in illumination and facial expression changes and is capable of detection and localization of faces appeared in a still image. The robustness of the algorithm in cluttered backgrounds and its ability to find multiple frontal faces including occluded ones are another features of the proposed model as demonstrated by the experiments.

Keywords: *Natural Image Understanding, Image Distribution Coding, Efficient Coding, Frontal Face Detection, Visual Cortex Top-Down Model*



Amirkabir University of Technology
Computer Engineering and Information Technology Department

Submitted to the Department of
Artificial Intelligence
in partial fulfillment of the requirements for the degree of
Master of Science

Thesis Title:

Brain Inspired Face Detection

By:

Kourosh Meshgi

Supervisor:

Dr. Saeed Shiry Ghidary

October 2010