

دانشگاه صنعتی امیرکبیر (پلی‌تکنیک تهران)

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پایان‌نامه کارشناسی ارشد

در رشته مهندسی کامپیوتر گرایش هوش مصنوعی

عنوان پایان‌نامه:

آشکارسازی چهره با الهام از مغز

نگارش:

کورش مشگی

استاد راهنما:

دکتر سعید شیری

پاییز ۱۳۸۹

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ
اللّٰهُمَّ اكْفُنْهُ مِنَ الْجَنَّةِ
وَلَا تُنَزِّلْهُ إِلَى الْجَهَنَّمِ
وَلَا يَرْجِعْهُ إِلَى دُنْدُونِ
شَرِّهِ إِنَّمَا أَنْهَاكُنَا
عَنِ الدُّنْدُونِ بِأَنَّهُمْ
كُفَّارٌ لَا يَعْلَمُونَ

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،

وبه او که در دلم جایش بود و از رگ گردنه نزدیکتر

سپاسگزاری

با سپاس از جناب آقای دکتر سعید شیری که با حمایت بی‌دریغ، پشتیبانی علمی، راهنمایی‌های گرانقدر و دلسوزی پدرانه 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 بپره مند نمودند.

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

تصویب نامه

چکیده

یکی از عملکردهای زیربنایی سیستم بینایی، کدگزاری سنگ بنای تصاویر طبیعی یعنی لبه‌ها، بافت‌ها و شکل‌هاست که در عملیاتی مانند تشخیص چهره و فهم صحنه‌های طبیعی از آن استفاده می‌شود. برای این منظور باید یک بازنمایی مجرد تشکیل داد که از نمونه‌های خاصی از ورودی‌های بینایی تعمیم یابد. از آنچه درباره بخش اولیه بینایی در قشر خاکستری مغز انسان دانسته شده، برمی‌آید که این سیستم ترکیبی از ویژگی‌های تصویر را بکار می‌گیرد 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 می‌شود، استفاده می‌کند. در مدل پیشنهادی این پژوهش تابع توزیعی که با تصاویر ورودی بیشترین همخوانی را دارد پیدا شده و توسط میزان فعالیت نورونهایی که برای کدگزاری آن در نظر گرفته شده اند بازنمایی می‌شود. این مدل با تصاویر طبیعی آموزش می‌بیند و با یادگیری یک مجموعه فشرده از توزیعهایی که معمولاً در تصاویر به آنها برخورد می‌کنیم، قابلیت تعمیم می‌یابد. نورونهای مدل خصوصیت‌های زیادی را از خود نشان می‌دهد که شباهت زیادی به سلولهای بخش بینایی قشر خاکستری مغز دارد و با برقراری این تناظر می‌توان بسیاری از ویژگی‌های سلولهای پیچیده از جمله پدیده‌های غیرخطی مشاهده شده در آنرا توجیه کرد. این مدل از استراتژی‌های کدگزاری در قشر اولیه بینایی انسان و لایه‌های بالاتر آن الهام گرفته است و بازنمایی حاصل از آن قادر است که تصاویری که دارای توزیعی مشابه هستند را با هم خوش 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 چهره‌ها را انجام می‌دهد. مدل حاصل بنابر آزمایشاتی که شرح داده خواهد شد، در برابر تغییر نورپردازی و حالت چهره مقاوم است و قابلیت آشکارسازی و مکانیابی چهره‌ها را در تصاویر دارد. از مهمترین ویژگی‌های این روش مقاومت آن در برابر پس زمینه‌های شلوغ و یافتن چند چهره موجود در تصویر حتی با وجود پوشانده شدن بخشی از آنها می‌باشد.

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

پیشگفتار

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 رساند. 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 باشур 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 کامپیوت و آن در این دهه بدست می آید. این مسئله در وهله اول بسیار بعيد به نظری 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 گزاری 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 پنهانی باند ارتباطی به طور نمایی رشد می کنند. تا همین اواخر اینکار با تسريع نرخ کلک انجام می شد که در طول حیات خودمان نیز شاهد نمو آن از کیلوهرتز تا گیگاهرتز بوده ایم. اما نرخ کلک مدتی است به سطح ثابتی رسیده است و اکنون پیشرفتها 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 دستگاههایی بر مبنای اصول پردازش عصبی ساخت. این دستگاهها همانند مغز به جای منطق قطعی بر مبنای منطق اختلالاتی 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 می‌پردازیم. نوامبر قبل، یکی از محققان IBM به نام دارمندرا مودا^۵ در یک کنفرانس ابررایانش اعلام کرد که تیم او برنامه‌ای نوشته است که مغز یک گربه را شبیه سازی می‌کند. این اخبار باعث شگفتی بسیاری شد زیرا او از مغز موش فراتر رفته بود و گوی سبقت را در این مهم از سایر گروه‌ها ربوده بود. برای انجام این کار، مودا جایزه گوردن بل بنیاد ACM را که به دستاوردهای شگفت انگیز در زمینه برنامه‌های پردازش با کارایی بالا اعطای می‌شود 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 پژوهش مغز آبی^۸ است که در سال ۲۰۰۹ به صراحة گفته بود: «ساخت مغز یک انسان غیر ممکن نیست و ما می‌توانیم در ۱۰ سال آینده آنرا انجام دهیم.» مارکرام در نامه سرگشاده‌ای به سرپرست فنی IBM برنارد مایرسون، مودا را به فریب گستردۀ 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 رایانه‌ای حرکت می‌کردند 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 کرده‌اند. هر دو مدل از آنچه در واقعیت اتفاق می‌افتد بسیار بسیار کندتر هستند. مدل مودا تنها از یک سوما^۹ – بدن سلول حاوی هسته سلول- و نیزه‌های ساده سازی شده پاسخ نورونی تشکیل شده است. از طرف دیگر مدل مارکرام بازسازی دقیقی از نورونهای ساده نظری کانالهای یونی را در بر می‌گیرد. سیناپسها و اتصالات بین نورونها طیف کاملی از مکانیسمهای دروازه بندی و ارتباطاتی نظری کانالهای یونی را در بر می‌گیرد. 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 می‌سازد که از یک سو مدل‌های ساده سازی شده نورونها برای اجرای سریعتر شبیه سازیها استفاده می‌شود و از سوی دیگر به جزئیات بیولوژیکی نورون برای شناخت آنها توجه می‌شود. در واقع فیزیکدانان و مهندسان با نگاه به نورون به دنبال سادگی هستند در حالی که بیولوژیستها با نگاه به همان نورون به دنبال پیچیدگی می‌گردند. مشکل مدل‌های ساده شده اینست که آنها گاهی اصل را نیز هنگام حذف حواشی دور می‌ریزند. 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 می‌رود؟ 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 شبیه سازیهای گسترده‌ای که توسط هر در 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 خروجی های حرکتی است. همچنین هردوی آنها قادر ساختار بخش‌های مغز - مثلاً مخچه که حرکات را هماهنگ می کند، آمیگدالا^{۱۱} که وضعیتهای عاطفی را پدید می آورد، و نخاع که سازمان عضلانی را تحت نظر دارد - می باشند. با این حال ما از مدل مودا چگونگی برنامه نویسی معماری های موازی در مقیاس بالا را می آموزیم تا بتوانیم شبیه سازیهایی که مقیاسی به بزرگی تعداد زیاد نورونها و سیناپسها در مغز واقعی دارند را اجرا کنیم. از مدل مارکرام نیز چگونگی تعابیه سطوح مختلف جزئیات درون مدل را فرا می گیریم. مودا در مقاله اش پیش بینی کرده است که بزرگترین ابرایانه می تواند عناصر ابتدایی مغز بشر را در سال ۲۰۱۹ بصورن بلادرنگ شبیه سازی کند، بنابراین او و مارکرام بالأخره در یک چیز تفاهم دارند. البته این شبیه سازیها در بهترین حالت می توانند مشابه مغز یک کودک یا یک دیوانه^{۱۲} باشد؛ مغز چیزی بسیار بیش از حاصل جمع قسمتهای مختلفش است.

طمئنأً ساختن یک گربه یا انسان لزومی ندارد و مطلوب نیست چرا که همین الان هم انسانها و گربه هایی با کارکرد کامل وجود دارد! این فناوری می تواند کاربردهای دیگری داشته باشد. در سال ۲۰۰۵ سیمون هایکین^{۱۳}، مدیر آزمایشگاه سیستمهای شناختی در دانشگاه مک مستر^{۱۴} مقاله ای تأثیرگذار بنام «رادیوی باشعور: ارتباطات بیسیم قدرت گرفته از مغز^{۱۵}» نوشت که بستر را برای نسل جدید شبکه های بیسیم فراهم می کرد. این شبکه ها از اصول محاسباتی مغز برای پیش بینی مدل استفاده از طیف الکترومغناطیسی بهره می گیرد و در استفاده از پهنهای باند از روش‌های فعلی کاراتر است. چنین سیستمی چیزی دور از دسترس نیست. برنامه هایی برای مناقصه پروژه پیاده سازی این سیستمهای مخابراتی هوشمند در آمریکا تصویب شده است.

بزودی راههای مشابه بیشتری برای توسعه سایر امکانات فعلی بوجود خواهد آمد که می توان 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 جاده ها، الگوهای هوا، شیوع بیماریها، و ... - را روان خواهند کرد و سیستم از این اطلاعات برای بهینه سازی هدف خود - کاهش مصرف برق، زمان سفر، و ... - با منظم کردن جریان منابع سود می جوید. قسمتهایی از این سیستمهایی همینک نیز موجودند ولی همچنان یک سیستم عصبی مرکزی برای تجمعی این سیل اطلاعات و اتخاذ تدبیر لازم به وجود نیامده است. در روزی نه چندان دور، این امر محقق می شود و به تدریج با تقلید از نحوه کار مغزهای ما، دنیای اطراف ما هوشمندانه تر و کاراتر می شود. همانطور که ای زیرساختها تکامل می یابند، 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]

1. فهرست عناوين
2. فهرست شكلها

فهرست عناوین:

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

۳۳	تخمین درستنمای 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	قوانين یادگیری	۱-۳-۲
۴۰	قانون یادگیری هب	۱-۲-۳-۲
۴۱	قوانين 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	V1 لایه	۳-۴-۲
۶۳	V2 لایه	۴-۴-۲
۶۵	V4 لایه	۵-۴-۲
۶۶	IT لایه	۶-۴-۲
۶۸	ویژگیهای آماری تصاویر طبیعی	۵-۲
۶۸	مقدمه	۱-۵-۲
۷۰	آمارگان تک پیکسلی	۲-۵-۲

۷۰	قاعده مندیهای 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 مندیهای 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 مندیهای 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	کدگزاری بهینه	۲-۱-۳
۸۹	مدلهای پایین به بالا	۳-۱-۳
۹۰	VisNet مدل	۱-۳-۱-۳
۹۲	مدل نقشه های قشری مغز	۲-۳-۱-۳
۹۳	HMax مدل	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	ساختر سلسله مراتبی	۳-۴-۱-۳
۱۴۳	مروری بر روش‌های آشکارسازی چهره	
۱۴۳	مقدمه	۱-۲-۳
۱۴۶	روشهای مبتنی بر دانش	۲-۲-۳
۱۴۸	روشهای مبتنی بر ویژگی	۳-۲-۳
۱۴۹	ویژگی‌های 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	کلیشه‌های از پیش تعریف شده	۱-۴-۲-۳
۱۵۸	کلیشه‌های انعطاف پذیر	۲-۴-۲-۳
۱۵۹	روشهای مبتنی بر ظاهر	۵-۲-۳
۱۶۰	چهره‌های یکه	۱-۵-۲-۳
۱۶۱	روشهای مبتنی بر توزیع	۲-۵-۲-۳
۱۶۳	شبکه‌های 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	کدگزاری جمعیتی	۳-۳-۴
۲۰۳	تعیین 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 شباخت 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	ن-۱
۲۶۷	مخفف ها	ن-۲

فهرست شکلها:

عنوان	شماره
مفهوم ساختار توسط وایستگیها پدید می آید	۱-۱
تصویر پنجره ها	۱-۲
پنجره های سینوسی	۲-۲
سیگنالهای اصلی	۳-۲
مخلوط مشاهده شده از سیگنالهای اصلی شکل ۳-۲	۴-۲
تخمینی از سیگنالهای اصلی	۵-۲
توابع پایه ICA از تصاویر طبیعی	۶-۲
توزیع توام مؤلفه های λ_1 (محور افقی) و λ_2 (محور عمودی) با توزیع یکنواخت	۷-۲
توزیع توام مخلوطهای مشاهده شده λ_1 (محور افقی) و λ_2 (محور عمودی)	۸-۲
توزیع چند متغیره دو متغیر مستقل گوسی	۹-۲
تابع توزیع توزیع لابلائین	۱۰-۲
نمایشی از روش تعقیب افکنش و جالب بودن افکنشهای غیرگوسی	۱۱-۲
برش عرضی چشم	۱۲-۲
پدیده شکست نور در چشم(چپ)، پدیده تطابق(راست)	۱۳-۲
جایگاه ماکولا، صفحه نوری و فوویا روی شبکیه	۱۴-۲
میدان گیرندهای چشم	۱۵-۲
ساختار کلی LGN	۱۶-۲
مکان عقده های سر راه سیگنالهای بینایی	۱۷-۲
نواحی دخیل قشر مخ در ادراف بینایی انسان	۱۸-۲
نواحی قشری موجود در مسیرهای بینایی و ارتباطات آنها	۱۹-۲
اتصالات نواحی بینایی قشر مخ به یکدیگر	۲۰-۲
نواحی قشری موجود در مسیرهای بینایی و کانالهای موازی آنها	۲۱-۲
مکان نواحی دخیل قشر مخ در ادراف بینایی انسان	۲۲-۲

۵۶	قشر ابتدایی بینایی	۲۳-۲
۵۶	دو مسیر پردازش داده ها در قشر بینایی	۲۴-۲
۵۷	مسیر بطئی در قشر بینایی	۲۵-۲
۵۸	: نواحی بینایی قشر مخ میمون ماکاکو	۲۶-۲
۵۹	ساختر ناحیه دریافت برای چند نورون در ناحیه V1	۲۷-۲
۶۰	ساختر 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	ساختر 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 f_1/f است	۴۰-۲
۷۴	فیلترهای واپتینینگ 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 فرآکامل پراکنده با واحدهای نویز دار که به تنظیم فرکانس مکانی زیادی دست می ابد	۴۷-۲
۸۳	وابستگی پاسخ مؤلفه های مستقل به زمینه در تصویر	۴۸-۲

۸۴	توزیع احتمال توان پاسخ فیلترهای گبر به مجموعه تصاویر از کلاس های مختلف.	۴۹-۲
۹۱	مدل VisNet	۱-۳
۹۱	یادگیری بر اساس تبدیلات پیوسته	۲-۳
۹۳	مدل تقشه 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 جایگایی 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	تابع 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	واریانس های ضرایب ICA از یک محتوا تا محتوا 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 چند مقیاسه موجک برای تصویر انسیتین	۲۱-۳
۱۱۸	مدل ترکیب متغیرهای گوسی برای توجیه فتار وابستگی واریانس	۲۲-۳
۱۲۰	توزیع توان 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 توان شرطی برای تکه های تصویری متفاوت و نویز سفید	۲۵-۳

۱۲۵	مقایسه آمارگان ضرایبی 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 ۴ن رابطه بین مدلها توزیع	۳۰-۳
۱۳۰	خلاصه ای از مدلها گوسی	۳۱-۳
۱۳۲	یک بازنمایی گرافیکی از مدل سلسله مرانبی برای وایستگی 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 کدگزاری 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 هم راستا	۴۲-۳
۱۵۸	یک کلیشه 15×16 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 متعدد F	۴۶-۳
۱۶۴	نمونه های هر کلاس چهره برای SOM کوھون 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}) و تخمینهای پیشخور آنها($\hat{\hat{y}}$)	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 کدگزاری 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 مولدهای 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 یکه	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

۲۳۷

نتیجه هر چهار مدل برای آزمایش دقت آشکارسازی

۱۰-۵

فصل نخست

[مقدمه]

۱. بینایی رایانه‌ای و مسئله آشکارسازی چهره
۲. اهداف و محدوده پژوهش
۳. ساختار گزارش

۱-۱- بینایی رایانه‌ای و مسئله آشکارسازی چهره

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

مسئلی که برای روانشناسان و عصب شناسان در طراحی سیستمهای مربوط به چهره انسان مطرح است، با بر یک مقاله موروی 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؟ وجود تمایز چهره‌ها در مغز به چه خصوصیاتی از آنها باز می‌گردد؟ نقش تحلیل فرکанс مکانی در این عملیات مغز چیست؟ آیا این مسئله نسبت به نقطه دید نامتفاوت 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 یافته‌های عصب شناسی و دانسته‌های جامعه علمی از مغز مدل کنند. این روشها یا مستقیماً به مدلسازی پردازش مغز با استفاده از ترکیب واحدی‌های پردازشی آن یعنی نورونها 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 یافته‌های عصب شناسی بستجند، یا با الهام گرفتن از انواع پردازش‌های درون مغز، مدل‌هایی ساخته اند که همان رفتار یا تقریبی از آنرا با رویکردی دیگر (مثلًا با استفاده 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 می‌دهد. 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 رایانه‌ای استفاده می‌شود 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 گستردۀ ای در مورد سیستم بینایی انسان، ساختار مغز انسان خصوصاً بخش بینایی آن، نحوه پرداش سیستم بینایی مغز و مدل‌های آن، روش‌های تشخیص چهره کنونی و نقاط قوت و ضعف آنها، و

¹⁶subjective test

ردگیری ایده‌های استفاده شده در ادبیات پردازش چهره تا پیدا شدن اصول حاکم بر آن نیاز می‌شود. علاوه بر این نکات باید به نقش خلاقیت، استفاده از شهود، رها کردن مسائل پیچیده و یافتن اصول زیربنایی، و بررسی هم و قویهای مشکوک و پدیده‌های خارج از انتظار (که این پایان نامه به آنها بسیار بها می‌دهد) 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 می‌پردازیم در حالی که بینایی ما تنها از ورودیهای ساده که همان پرتوهای نور تابیده شده روی شبکیه هستند تغذیه می‌شود. اطلاعاتی حیاتی که تصمیمات اساسی ما را هدایت می‌کند – مکان یک میوه روی درخت، سرعت یک اتومبیل نزدیک شونده، چهره یک فرد آشنا و ... – تماماً در این سیگنال ورودی مخفی شده‌اند. سوال اینجاست که درک جامع و پایدار ما از اشیا، سطوح، بافتها، سرعت، فاصله و ... چگونه در قشر خاکستری مغز ما شکل می‌گیرد؟ به بیانی دقیقتر، چه خرده اطلاعاتی توسط فعالیت هر نورون خاص منتقل می‌شود؟ چگونه یک جمعیت نورونی در کنار هم می‌توانند یک ساختار تصویر را بازنمایی کنند؟ چه اصل محاسباتی در مدارهای نورونی پیاده سازی می‌شوند؟

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 می‌سازد 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 شبکیه و عقده زانویی خارجی (نخستین مراحل مسیر بینایی، ر.ک. فصل ۲)، گامهای محاسباتی ابتدایی تا حد زیادی شناخته شده است (هر چند که پدیده‌های زیادی نیز وجود دارد که توجیهی برای آنها یافت نشده است). نورونهای این نواحی را می‌توان با توجه به پاسخ آنها به حرکه‌های ساده مشخص کرد و مدل‌هایی برای توجیه بسیاری از خصوصیات‌های پاسخها پدید آمده است (برای مثال ر.ک. به(4)Carandini 2004)) با اینحال در مراحل بعدی مسیر بینایی، در نواحی بینایی قشر خاکستری مخ، فعالیتهای نورونی نشان دهنده وجههای بسیار پیچیده تر و مجردتری از ورودی بینایی است. پاسخها گزینندگی بیشتری می‌یابند؛ یعنی سلولها به اغلب تصاویری که در آزمایش موجودند پاسخ ضعیفی می‌دهند و پاسخ به حرکه‌های ساده‌تر 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 پنجره‌ها، میله‌ها و نقطه‌ها) 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 کدگزاری می‌کنند و اگر حدسهای 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، پیش‌بینی هایی 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 دهه های اخیر بسیار مورد مطالعه قرار گرفته است و خصوصیتهای پیچیده ای 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 پدیده هایی پیچیده

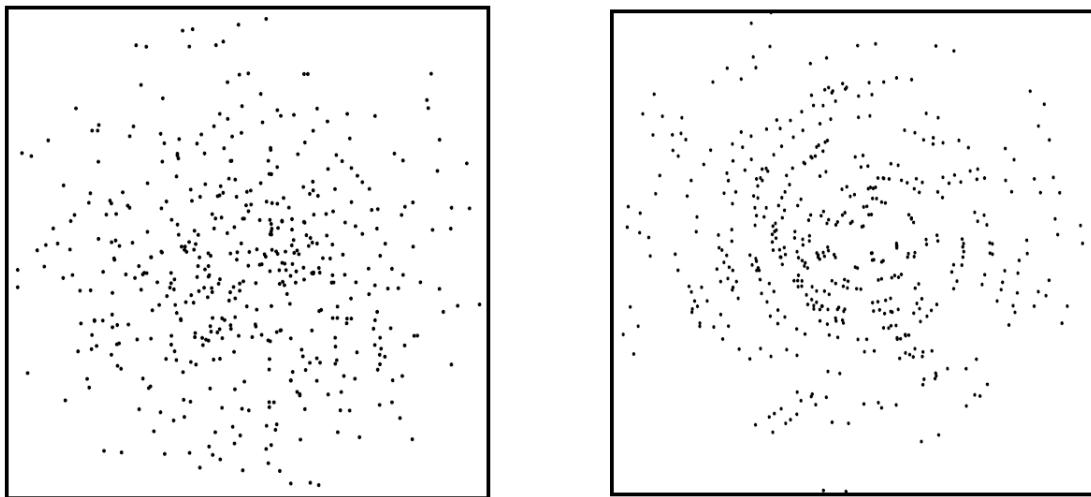
هوراس بارلو^{۱۷} معتقد است که افزونگی دربردارنده دانش است (Barlow 1989). با توجه به این نظریه، افزونگی آماری حاوی اطلاعاتی درباره ساختار محیط است. اطلاعات مهمی که یک سیستم ادراکی باید آنرا بیابد «هم و قویهای مشکوکی» است که قاعده مندیهای آماری جدیدی در ورودیهای حسی محیط است، که با آنچه سیستم با آن سازگار شده است متفاوت است. به عنوان مثال میله ها و لبه ها مکانهایی در ورودیهای تصویری هستند که در آن بین مقیاسهای مختلف مکانی هم راستایی فاز وجود دارد و می توان آنرا یک «هم و قوی مشکوک» نامید (Barlow 1994). مکانیسمهای یادگیری ای که افزونگی که در ورودی محتمل است را کدگذاری می‌کند و آنرا از خروجی حذف می‌کند، به سیستم امکان تشخیص مطمئن تر این قاعده مندیهای 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 مندیهای حرکتی حسی است. حرکتی 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 مجموعه ای از نقاط اولیه حول مرکز توزیع تولید شده است. این وابستگی ساده بین جفت نقطه ها این ظاهر دارای ساختار را بوجود آورده است.

مطابق نظریه بارلو، آنچه کشف آن برای سیستم حائز اهمیت است، قاعده مندیهای جدیدی در داده های حسی است که با محیطی که سیستم با آن سازگار شده است، متفاوت باشد. بارلو ای وابستگیهای جدید را «هم و قویهای مشکوک»^{۱۸} نامیده است. به عنوان مثال میله ها و لبه ها مکانهایی در ورودیهای تصویری هستند که در آن بین مقیاسهای مختلف مکانی هم راستایی فاز وجود دارد و می توان آنرا یک هم و قوی مشکوک نامید (Barlow 1994).

¹⁷Horace Barlow

¹⁸suspicious coincidents



شکل ۱-۱: مفهوم ساختار توسط وابستگیها پدید می‌آید (چپ) مجموعه‌ای از نقاط از یک توزیع گوسی (راست) نیمی از نقطه‌ها از توزیع گوسی انتخاب شده اند و نیم دیگر از دوران ۵ درجه‌ای حول مرکز توزیع پدید آمده‌اند.

مکانیسمهای یادگیری که افزونگی مورد انتظار در ورودی را کدگزاری می‌کنند و آنرا از خروجی حذف می‌کنند به سیستم امکان کشف مطمئن‌تر قاعده مندیهای جدید را می‌دهد. یاد گرفتن چنین تبدیلی معادل مدلسازی دانش پیشین وابستگیهای آماری ورودیهای (Barlow 1989) است. کدهای مستقل برای کدگزاری 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 می‌شود و هم وقوعیهای مشکوک که یک فاکتور علی جدید را نشان می‌دهند را با اطمینان بیشتری می‌توان پیدا کرد.

بارلو یک اصل سازماندهی برای یادگیری بدون ناظر بر پایه نظریه اطلاعات پیشنهاد کرده است. اطلاعاتی که توسط هر پاسخ بدست می‌آید 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 کرده‌اند:

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

اطلاعات با احتمال نسبت معکوس دارد و می‌توان 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 می‌باشد:

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

آنتروپی در توزیع یکنواخت به بیشترین مقدار خود می‌رسد و در توزیعهایی با کورتوسیس بالا^{۱۹} (دارای قله‌های تیز) کمترین مقدار خود را می‌یابد. آنتروپی توأم بین دو متغیر اینگونه محاسبه می‌شود:

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

که در آن $I(x_1, x_2)$ اطلاعات متقابل بین x_1 و x_2 است که از رابطه $-1 \leq I(x_1, x_2) \leq 1$ با استفاده از توزیع احتمال توأم $p(x_1, x_2)$ بدست می‌آید. همانطور که اشاره شد، بارلو کدگزاری کمینه آنتروپی را به دید یک استراتژی بازنمایی عمومی می‌نگریست. کدهای با آنتروپی کمینه و کورتوسیس بالا دارای اطلاعات متقابل اند که بین عناصر خود هستند. دلیل این امر اینست که آنتروپی توأم یک کد چندبعدی به صورت مجموع تک تک آنتروپی‌ها منهای اطلاعات متقابل بین عناصر آن تعریف می‌شود (رابطه ۳-۱). از آنجایی که آنتروپی توأم کد، با کمینه کردن جمع تک تک آنتروپیها ثابت باقی می‌ماند، اطلاعات متقابل نیز کمینه می‌شود. راه دیگری برای توجیه این قضیه جابجا کردن افزونگی بین عناصر به افزونگی بین توزیعهای تک عناصر است. توزیعهای تک تک عناصر با آنتروپی کمینه از این لحاظ افزونه است که آنها اغلب مقادیر مشابهی را اختیار می‌کنند.

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 Heeger (1992) حساب تغییرات واریانس سیگنالهای ورودی می‌جویند (Barlow 1989). کنترل بهره کنترast^{۲۰} که در V1 نیز شرح داده شده است) در دریافت کننده‌ها دریافت می‌شوند ناکارا هستند. اتیک مزایای تکاملی کدگزاری مؤثر مانند کنار آمدن با گلوگاههای اطلاعاتی را که به دلیل محدودیت پهنای باند و بازه دینامیکی بوجود می‌آیند شرح داده است (Atick 1992). وی اصل کارایی بازنمایی اطلاعات را به عنوان اساس طراحی کدگزاری حسگرها مطرح می‌کند و شواهدی طبیعی مانند شبکیه مگس^{۲۱} گوشت^{۲۲} و پستانداران برای آن می‌آورد. سلولهای بزرگ تک قطبی^{۲۳} در چشم مرکب مگس گوشت، ناکارایی را به کمک استفاده نامساوی از سطوح پاسخ اعصاب از بین می‌برد. مؤثرترین بهره پاسخ زمانی رخ می‌دهد که توزیع احتمال خروجیها برای تمام حالت‌های خروجی ثابت باشد. راه حل اینست که بهره تابع انتقال با تابع توزیع تجمعی ورودی منطبق باشد. لافلین

اتیک و ردلیچ^{۲۴} (Atick and Redlich 1992) چنین سیستمهای از بین بردن همبستگی را به عنوان استراتژی کدگذاری عمومی برای سیستمهای تصویری معرفی می‌کنند. روش آتیک و ردلیچ از دیدگاه کدگزاری مؤثر به کاهش افزونگی پرداخته است. آنها دریافته اند که محركهای طبیعی بسیار افزونه اند و در نتیجه نمونه‌های سیگنالهایی که توسط آرایه‌ای از دریافت کننده‌ها دریافت می‌شوند ناکارا هستند. اتیک مزایای تکاملی کدگزاری مؤثر مانند کنار آمدن با گلوگاههای اطلاعاتی را که به دلیل محدودیت پهنای باند و بازه دینامیکی بوجود می‌آیند شرح داده است (Atick 1992). وی اصل کارایی بازنمایی اطلاعات را به عنوان اساس طراحی کدگزاری حسگرها مطرح می‌کند و شواهدی طبیعی مانند شبکیه مگس^{۲۱} گوشت^{۲۲} و پستانداران برای آن می‌آورد. سلولهای بزرگ تک قطبی^{۲۳} در چشم مرکب مگس گوشت، ناکارایی را به کمک استفاده نامساوی از سطوح پاسخ اعصاب از بین می‌برد. مؤثرترین بهره پاسخ زمانی رخ می‌دهد که توزیع احتمال خروجیها برای تمام حالت‌های خروجی 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 تجمعی احتمال کنتراست را در محیط مگس اندازه گرفت و تطابق زیادی بین بهره نورونهای LMC وتابع توزیع تجمعی احتمال یافت.

اتیک نتایج مشابهی درباره شبکیه تابع انتقال مدولاسیون^{۲۴} شبکیه پستانداران را بدست آورده است. چگالی تجمعی طیف دامنه صحنه های طبیعی تقریباً I/f است (Field 1987). تابع MTF یک کدگزاری 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 تحریبی در I/f یک خروجی نسبتاً هموار برای فرکانس‌های کمتر از ۳cpd^{۲۵} بدست می دهد. اتیک چنین توابع انتقالی را فیلترهای واپتینینگ^{۲۶} می خواند زیرا توزیع پاسخ خروجی را برای تمام فرکانس‌ها برابر می سازد.

این تحلیلها ابزاری بدست دادند که به وسیله آن بتوان کارایی کدگزاری 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 کدگزاری 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 کدگزاری در فرمولاسیون بارلو است که همبستگیهایی که در ورودی انتظار آن می رود را کدگزاری 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 1996) و تحلیل ویژگیهای محلی^{۲۷} (Penev and Atick 1991) بازنماییهای چهره ای داده گرا بر مبنای 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 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 نورونهای 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.

²⁴modulation transfer function (MTF)

²⁵Whitening filters

²⁶local 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 مولّدی است که احتمالاً داده‌های مشاهده شده را پدید آورده‌اند. پارامترهای این مدل‌های مولّد 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 مانده اند از اهمیت بسزایی برخوردار است ولی تاکنون مدل‌های اندکی درباره پردازش در نواحی بینایی پیشرفت‌ههای تر مانند V4 یا V2 دست به پیش‌بینی زده اند یا به توجیه کارکردی پدیده‌های غیرخطی V1 مبادرت ورزیده اند. مدل‌های فعلی چنین توصیف مترآکمی از ساختارهای مرتبه‌های بالا بدست نمی‌دهد و نیاز مبرمی به مدل‌های قویتری که بتواند وابستگی‌های مشاهد شده را در بر بگیرد، قاعده مندیهای 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 شبکیه) 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 تا کنون 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.

۱-۲-۱- اهداف و محدوده پژوهش

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 پیکسلهای است 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:

- از چشم تا انتهای عصب بینایی که وارد قسمت V1 می شود؛
- ناحیه V1 که شامل دو لایه سلولهای ساده و پیچیده می شود،
- ناحیه V2 که شامل اتصالات جانبی آن نیز می شود؛
- 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 شبکیه 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 کهاین 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 پرسپترونی، 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 کدگزاری بهینه در کارایی کد حاصل از قسمتهای اولیه بینایی قشر خاکستری مخ انسان است. این کدگزاری بر اساس وابستگی کواریانس فیلترهای خطی حاصله از تصاویر طبیعی است و نشان داده می شود که نه تنها اهداف کدگزاری 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 مختصراً در مورد اصول ریاضی مورد استفاده، به بررسی روش‌های استنتاج در مدل‌های مولّد پرداخته می شود. پس از آن روش تحلیل مؤلفه‌های مستقل که الهام بخش قسمت بزرگی از کارهای این پژوهه است به طور کامل مورد بررسی قرار می گیرد و روند آن بررسی می شود تا راه حل استفاده شده در پژوهه 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 شبکه‌های 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 مطرحه 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 کدگزاری 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 مغزه استند (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 تکه‌های 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 پژوهه پرداخته می شود و ایده‌های استفاده شده در بخش‌های مختلف مدل یکجا بیان می شود و راه حل عبور از بن‌بستهای 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.

فصل دوم

[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

۲-۱-۲- پایه های ریاضی

۲-۱-۱- حوزه فرکانس

۲-۱-۱-۱- 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 محرك 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 سلوهای V1 دارای خاصیت گرینندگی^{۲۹} نسبت به فرکانسهای مکانی هستند (سلوهای ساده به فاز حساس هستند در حالی که سلوهای پیچیده اینگونه نیستند). بزرگی پاسخ یک سلو 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 سلوها از دید فرکانسهای بهینه خود، یک دنباله را تشکیل می دهند که کرانهای آن از فرکانس مکانی کمی مانند ۰.۲cpd^{۳۰} تا فرکانس های 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).

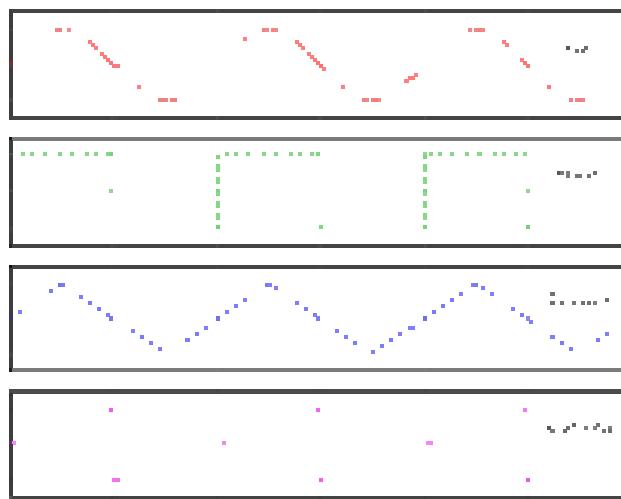
۲-۱-۱-۲- پنجره های سینوسی

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 نورسنچ در راستای عمود بر راستای پنجره بدست می آید. محور افقی فاصله ای که نورسنچ جابجا شده است را نشان می دهد. به عنوان مثال پنجره با موج مربعی دارای خطوط پایین است که با میله های سیاه متنظر 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 حاصل جمع 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 روشناییها بجز میله های لبه تیز داشته باشند. اگر نمودار پنجره ها سینوسی باشد، پنجره مانند میله های سیاه و سفید تاری به نظر می آید و پنجره سینوسی نامیده می شود. پنجره سینوسی در اپتیک

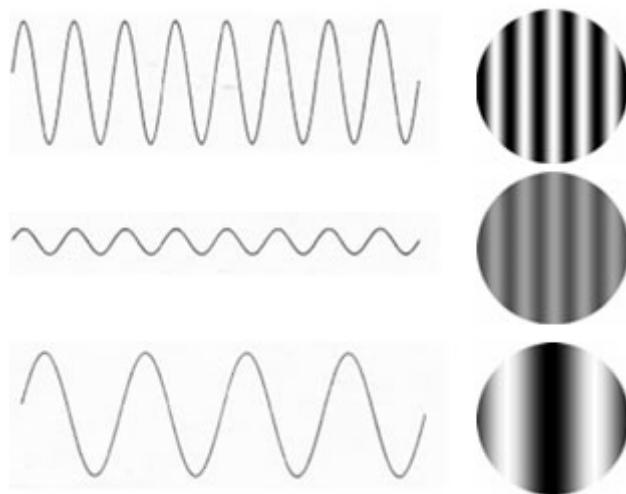
³¹duty cycle

³²pitch

³³Michelson distance

برای تعیین قابلیت انتقال لنزها کاربرد زیادی دارد. یک لنز تصویری از پنجره سینوسی می‌سازد که همچنان سینوسی است ولی با توجه به فرکانس مکانی و احتمالاً تغییر فاز، کنتراست آن کاهش می‌یابد. شاخه‌ای از ریاضیات که با این قسمت از علم اپتیک سر و کار دارد تحلیل فوریه است. پنجره‌های سینوسی در پژوهش روی ادراک تصویری نیز بطور گسترده‌ای مورد استفاده قرار می‌گیرند. کمپل و رابسون^{۳۴} استفاده از پنجره سینوسی را باب کردند. استدلال آنها این بود که بینایی انسان یک تحلیل فوریه روی تصاویر شبکیه انجام می‌دهد (Campbell and Robson 1968).

در مطالعه ادراک تصویری، غالباً پنجره‌های سینوسی (تصاویری شامل نوارهای سفید و سیاه با یک الگوی خاص، شکل ۲-۲) برای بررسی قابلیتهای سیستم بینایی استفاده می‌شود. در این محرک فرکانس مکانی به صورت تعداد چرخه‌ها در واحد درجه (از زاویه بینایی) بیان می‌شود.



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

۲-۱-۲- استنتاج در مدل‌های مولّد

یکی از روش‌های یادگیری بدون ناظر سعی می‌کند یک بازنمایی از داده‌ها ارائه دهد که در آن توزیع احتمال زیرین داده‌ها را آشکار کند. در این روش، یک مدل از پیش در نظر گرفته می‌شود که شکل کلی توزیع احتمال داده‌ها را محدود می‌کند. سپس پارامترهای این مدل با بیشینه کردن درستنمایی مدل که داده‌های مشاهده شده را تولید کرده است بدست می‌آید. پارامترهای مدل را می‌توان به عنوان وزنهای شبکه نیز در نظر گرفت. مثلاً یک مدل مخلوط گوسی (ر.ک. فصل ۳) فرض می‌کند که هر نقطه داده از ترکیبی از مؤلفه ϕ بوجود آمده است که هریک از این مؤلفه‌ها توزیعی گوسی دارای میانگین μ_i واریانس σ_i^2 و احتمال اولیه π_i می‌باشد. در این چارچوب شبکه به صورت یک مدل مولّد احتمالاتی داده‌ها دیده می‌شود. فرض می‌کنیم $(x_1, \dots, x_n) = \vec{x}$ نمایانگر داده‌های مشاهده شده باشد که نمونه‌ها در آن مستقلند. احتمال داده‌ها با داشتن این مدل از رابطه زیر بدست می‌آید:

³⁴Campbel & Robson

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

احتمال داده ها بر حسب احتمال اولیه هر یک از زیرمدلهای $(\phi_i | \Phi)$ و احتمال ثانویه داده ها با فرض داشتن زیرمدل $p(\vec{x} | \phi_i)$ تعریف می شود که ϕ_i به صورت (u_i, σ_i, π_i) تعریف می شود. پارامترهای هر یک از زیرمدلهای (u_i, σ_i, π_i) ، با شبیه نوری 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 وزنهای شبکه در نظر گرفته می شود. چنین مدلها می توانند قرار می گیرند که در آن مدل طوری انتخاب می شود که داده های مشاهده شده به احتمال زیادتری توسط آن تولید شده باشند. مدلها بیشینه سازی درستنمایی حالت خاصی از مدل استنتاج بیزی است. احتمال رخداد مدل با فرض داشتن داده ها از رابطه زیر بدست می آید:

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

تابع هزینه بیشینه درستنمایی، مقدار $p(\Phi | \vec{x})$ را بیشینه می کند، که اگر فرض کنیم تابع احتمال اولیه مدل $p(\Phi)$ یکنواخت است آنگاه $p(\Phi | \vec{x})$ نیز بیشینه می شود چرا که $p(\vec{x})$ 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 احتمالهای اولیه زیر مدلها $p(\phi_i)$ به عنوان جمع وزنار گوسی به مرودی، به یک عملگر نرمال سازی بیشینه گیری نرم^{۳۵} داده می شوند. این احتمالهای اولیه نسبتهاي ترکيب، π_i ، را بدست می دهند. شباهتهايي بين قانونهاي بروزرسانی که از اين مشتق جزئی چنین توابع هدفي بدست می آيد و ساير قوانين يادگيری بدون ناظر، مانند يادگيری هب^{۳۶} (که در قسمت ۲-۳-۲ شرح داده شده است) وجود دارد. به عنوان مثال قانون بروزرسانی يادگيری رقابتی بیشینه درستنمایی از يك مؤلفه هب نرمال شده و يك بخش تنزل وزن بدست می آيد.

یک محدودیت مدلها مولد اینست که بجز برای مدلهاي بسيار ساده، هر الگو می تواند از مقداردهی های زيادي به پارامترها بدست بيايد که تنظيم پارامترها برای اين تعداد نمايی حالتها برای بيشينه کردن احتمال الگوهای مشاهده شده چندان راحت نيست. ماشين هلمهولتز^{۳۷} (Dayan, Hinton et al. 1995) با بيشينه کردن کران پايان برای احتمال مشاهدات که راحت محاسبه می شود، راه حلی برای غلبه بر اين انفجار ترکيباتي ارائه کرده است. اين روش می تواند به صورت يك يادگيری سلسه مراتبي خود نظارتی در نظر گرفته شود که مسیرهای قشری پیشخور و پسخور مغز را مرتبط سازد. اتصالات «شناسايی» پايان به بالا وروديهها را به بازنمايي هاي در لايده هاي مخفی بعدی تبديل می کنند و اتصالات «مولد» بازنمايي يك لايه را برای بازنمايي لايه بالا بازسازی می کنند. اين شبکه يك مدل معکوس («شناسايی») برای تخمين توزيع درست احتمال ثانویه داده های ورودی استفاده می کند.

³⁵soft-maximum normalization

³⁶Hebbian learning

³⁷the 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 لایه‌های مخفی را می فرستد، سپس تفاضل بین بردار ورودی و بردار بازسازی شده از روی بازنمایی لایه‌های مخفی را ارسال می کند. کمینه کردن طول توصیف، شبکه را مجبور به یادگیری یک بازنمایی مقرر می کند که قاعده مندیهای ضمنی داده را تسخیر کرده است. یک تابع هزینه C به صورت مجموع تعداد کل بیت‌های لازم برای توصیف تمام وضعیتهای مخفی در تمام لایه‌های مخفی، α ، به علاوه هزینه توصیف اطلاعات باقیمانده در بردار ورودی d با داشتن وضعیتهای مخفی تعریف می شود.

$$C(\alpha, d) = C(\alpha)C(d | \alpha) \quad (4-2)$$

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

$$E(C(\alpha, d)) = \sum_{\alpha} Q(\alpha | d)C(\alpha, d) \quad (5-2)$$

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 نااریب 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 قربت زیادی به روش‌های بیشینه 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 کدگزاری مؤثر دسته بندی خواهند شد. روش دیگری برای نیل به کدگزاری 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 کدگزار می‌توان افزونگی را کمینه کرد، همان طوری که در طول توصیف کمینه اینگونه بود، و یا می‌توان بدون فشرده سازی بازنمایی با بردن به فضایی با ابعاد بالاتر و نمایش به صورت کد پراکنده، افزونگی را تقلیل داد.

۱-۲-۳- تحلیل مؤلفه اصلی

تحلیل مؤلفه اصلی (PCA) یک مجموعه متعامد از محورها در راستاهای بیشترین کواریانس‌های داده‌ها می‌یابد. ماتریس X را مجموعه داده‌ای در نظر می‌گیریم که در آن هر ستون یک مشاهده و هر سطر یک اندازه گیری با میانگین صفر است. محورهای مؤلفه‌های اصلی عبارتند از بردارهای یکه ماتریس کواریانس اندازه گیریها، $\frac{1}{N}XX^T$ ، که در آن N تعداد مشاهدات را نشان می‌دهد. مقدارهای ویژه متناظر آنها نسبت تنوع در داده‌هایی را نشان می‌دهد که هر بردار یکه بدان مربوط است. نخستین مؤلفه اصلی به راستای بیشترین تنوع اشاره می‌کند، دومین بردار یکه به راستای بیشترین تنوع عمود بر اولی اشاره می‌کند و الی آخر. داده‌ها با تصویر کردن بردار هر داده روی این محورهای جدید، به صورت جدیدی نمایش داده می‌شوند. فرض می‌کنیم P ماتریس حاوی بردارهای یکه مؤلفه اصلی در ستون خودش باشد. بازنمایی PCA برای هر مشاهده در سطرهای A با این رابطه بدست می‌آید:

$$A = X^T P \quad (6-2)$$

بردارهای یکه درون P می‌تواند به صورت یک مجموع وزن روی داده‌ها، X دیده شود که خروجی‌ها ضرایب درون ماتریس A را می‌سازند. از آنجا که بردارهای یکه مؤلفه اصلی متعامدند، در واقع بردارهای پایه مجموعه داده نیز به شمار می‌روند. این امر می‌تواند 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 $X^T AP = A$. بنابراین داده‌های اصلی می‌توانند با استفاده از بردارهای یکه P (اینبار به عنوان بردار پایه) از روی ضرایب A ساخته شوند. یک بازنمایی با ابعاد کمتر را می‌توان با انتخاب مجموعه ای از مؤلفه‌های اصلی که بیشترین مقادیر یکه را دارا هستند بدست آورد. نشان داده شده است که برای هر تعداد بُعد، بازنمایی مؤلفه‌های اصلی خطای میانگین مربعی بازسازی را کمینه می‌کند. از آنجایی که بردارهای یکه در فضای کواریانس به جهات عمود برهم اشاره می‌کنند، بازنمایی مؤلفه‌های اصلی ناهمبسته است یعنی اینکه ضرایب یکی از محورها را نمی‌توان به صورت خطی از روی ضرایب محور دیگری پیش بینی کرد. راه دیگری برای نگرش به بازنمایی مؤلفه اصلی از طریق مدل‌های تخمین چگالی است. روش PCA داده‌ها را به صورت یک مدل گوسی چند متغیره مدل‌سازی می‌کند که ماتریس کواریانس آن الزاماً قطری است. می‌توان نشان داد که یک مدل مولد که درستنمایی داده‌ایی با توزیع گوسی و کواریانس قطری را بیشینه می‌کند، در واقع خطای میانگین مربعی داده‌های تولید شده را کمینه می‌کند.

۱-۲-۴- تحلیل مؤلفه مستقل

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

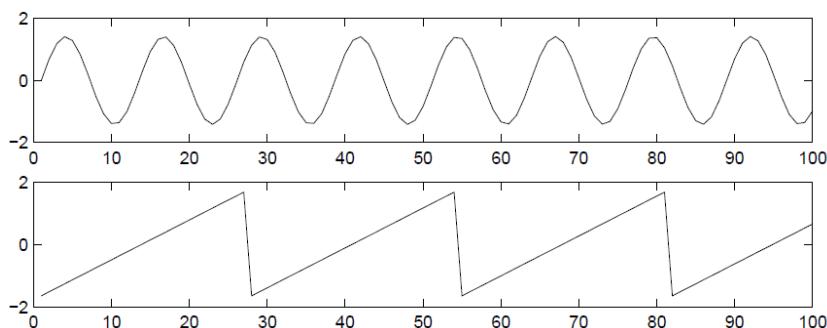
خطی معروف شامل تحلیل مؤلفه اصلی، تحلیل فاکتور، و پیگیری افکنش است. تحلیل مؤلفه مستقل (ICA) روشی است که هدف آن پیدا کردن ترکیبی خطی از داده های غیرگوسی است بطوری که مؤلفه ها از هم کاملاً (و یا حتی الامکان) استقلال آماری داشته باشند. چنین بازنمایی به نظر می رسد که ساختار اصلی داده در بسیاری از کاربردها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

۱-۴-۱-۲- 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 تعییه 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:

$$\begin{aligned}x_1(t) &= a_{11}s_1 + a_{12}s_2 \\x_2(t) &= a_{21}s_1 + a_{22}s_2\end{aligned}\quad (7-2)$$

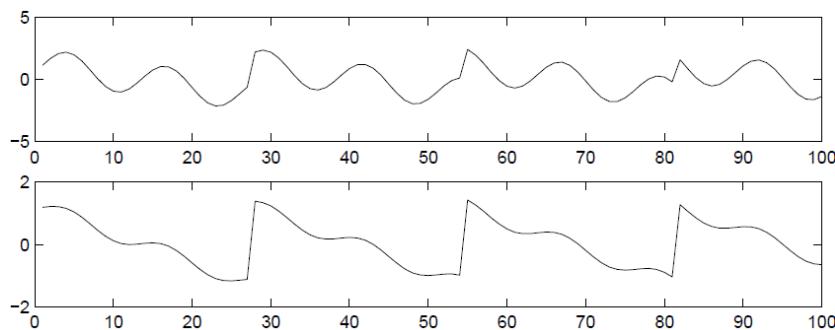
که در آن a_{11} , a_{12} و a_{21} , a_{22} پارامترهایی هستند که به فاصله میکروفون از گویندگان وابسته هستند. اگر بتوان تنها با استفاده از دو سیگنال ضبط شده $x_1(t)$ و $x_2(t)$ دو سیگنال اصلی گفتار $s_1(t)$ و $s_2(t)$ را تخمین زد، روش مورد استفاده، بسیار پرکاربرد خواهد بود. این مسئله به مسئله مهمانی کوکتیل^{۴۱} 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

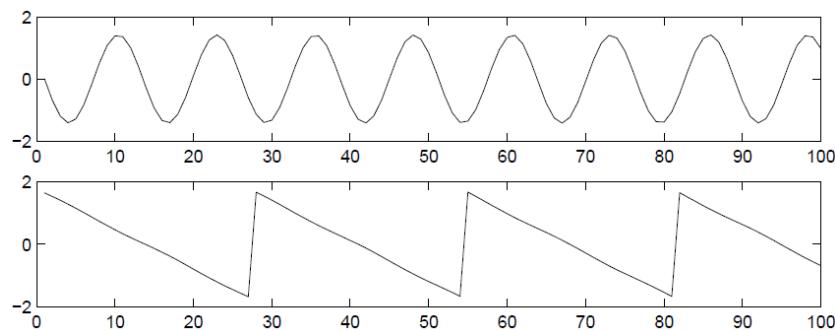
برای درک بهتر، شکل موجهای موجود در شکل ۳-۲ و شکل ۴-۲ را در نظر بگیرید. البته این سیگنالهای گفتار واقعی نیستند ولی برای روشن کردن موضوع کافی هستند. سیگنالهای گفتار اصلی ممکن چیزی مشابه موجهای شکل ۳-۲ باشند و سیگنالهای مخلوط ممکن است شبیه موجهای شکل ۴-۲ باشند. مسئله عبارتست از بازیابی داده های شکل ۳-۲ 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 ۳-۲: مخلوط مشاهده شده از سیگنالهای اصلی شکل ۲

یک روش حل این مسئله استفاده از خواص آماری سیگنال (t) برای تخمین a_{ii} است. به صور غیرمنتظره ای تنها فرض استقلال آماری $s_1(t)$ و $s_2(t)$ در هر لحظه t برای حل مسئله کافیست. آین فرض در خیلی از موارد چندان هم غیر واقعی نیست و در عمل لزومی هم ندارد که کاملاً برقرار باشد. تکنیک ICA می‌تواند برای تخمین a_{ij} بر اساس اطلاعات استقلال آنها به کار رود که این امکان را می‌دهد که دو سیگنال گفتار اصلی تفکیک شود. شکل ۲-۵ دو سیگنال که توسط روش ICA بدست آمده اند را نشان می‌دهد. همانطور که مشخص است این سیگنالها بسیار به سیگنالهای اصلی منبع مشابهند (علامت آنها بر عکس است که تفاوت چندانی ندارد).

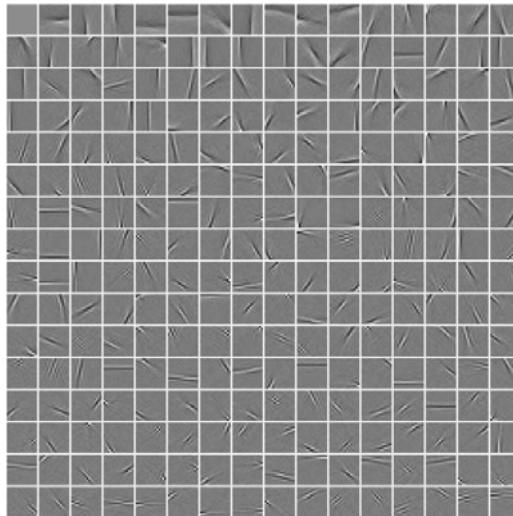


شکل ۴-۲: تخمینی از سیگنالهای اصلی که از روی سیگنالهای مخلوط مشاهده شده در شکل ۲-۴ بدست آمده است. سیگنالهای اصلی فارغ از علامت آنها بسیار دقیق تخمین زده شده اند.

تحلیل مؤلفه مستقل در اصل برای مسائلی طراحی شده است که بسیار شبیه مسئله مهمانی کوکتیل هستند. از کاربردهای گسترده آن می‌توان به تفکیک سیگنال EEG اشاره کرد. فعالیتهای مغزی که ضبط الکتریکی آنها توسط دستگاه الکتروانسفالوگرام انجام می‌شود، از ضبط پتانسیلهای الکتریکی بخش‌های بسیار متفاوتی از مغز متشكل شده است که توسط الکترودها روی جمجمه جمع آوری می‌شوند. این پتانسیلهای الکتریکی احتمالاً از اختلاط چندین مورد فعالیت بخش‌های متفاوت مغز تولید شده اند. این وضعیت بسیار مشابه مسئله مهمانی کوکتیل است: می‌خواهیم مؤلفه‌های اصلی فعالیت مغز را بیابیم ولی آنچه جمع آوری می‌شود تنها مخلوطی از این مؤلفه‌هاست. اطلاعات جالبی از اعمال روش ICA بر روی این سیگنالهای جمع آوری شده بدست می‌آید.

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 عموماً بر مبنای تبدیلات خطی (گسسته) است. تبدیلات خطی استانداردی 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 نکه 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 16×16 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.

۲-۴-۲- تعریف تحلیل مؤلفه مستقل

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

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n, \forall j \quad (8-2)$$

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 می‌توان با کم کردن میانگین از آن به مرکز برد که اینکار آنرا دارای میانگین صفر خواهد کرد.

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

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

گاهی به ستونهای ماتریس A نیاز داریم و آنها را اینگونه \vec{a}_i نشان می‌دهیم. بنابراین مدل را می‌توان اینگونه نیز نوشت:

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

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

نقشه 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 توزیعهایی غیر گوسی داشته باشند. با اینحال در این مدل ابتدایی فرضی نمی‌کنیم که این توزیعها را می‌دانیم (اگر این توزیعها معلوم باشند مسئله به طور چشمگیری ساده خواهد شد). برای سادگی همچنین فرض می‌کنیم که ماتریس مجھول تلفیق یک ماتریس مربعی باشد (که البته این فرض گاهی می‌تواند در نظر گرفته نشود). پس از تخمین ماتریس A می‌توان معکوس آنرا، که به آن W می‌گویند، بدست آورد و مؤلفه های مستقل را بسادگی از رابطه زیر حساب کرد:

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

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 می‌گیرد. 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.

۱-۲-۳-۴-۵- ابعامات تحلیل مؤلفه مستقل

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

۱- نمی توان واریانس (انرژی) مؤلفه های مستقل را تعیین کرد. علت این امر اینست که هر دوی A و \vec{s} مجهولند و ضریب اسکالر هر یک از منابع s_i همیشه می تواند با تقسیم شدن ستون متناظر \vec{a}_i از A بر همان اسکالر خنثی شود، رابطه ۱۰-۲ 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 ($1 = \{s_i^2\}_{ICA}$). آنگاه در A ماتریس A طوری سازگار می شود که این محدودیت را جبران کند. باید توجه داشت که همچنان ابعام در زمینه علامت را باقی می گذارد: می توان بدون تاثیر خاصی در مدل مؤلفه مستقل را در ۱- ضرب کرد. خوبشخانه این محدودیت در اغلب کاربردها چندان مهم نیست.

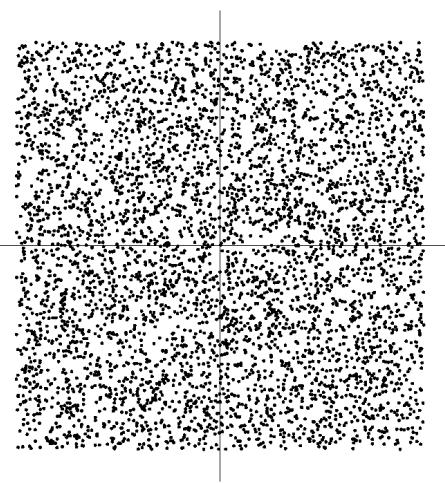
۲- ترتیب مؤلفه های مستقل مشخص نیست. علت آن بازهم به مجهول بودن هر دوی A و \vec{s} باز می گردد. ترتیب جملات حاصلجمع در رابطه ۱۱-۲ دلخواه است و هر مؤلفه مستقلی می تواند اول باشد. به طور رسمی یک ماتریس جایگشت P و معکوس آن اینگونه در مدل گنجانده می شود $P\vec{s} = AP^{-1}P\vec{x}$. عناصر \vec{s} مؤلفه های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

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

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 ۱۵-۲ برابر حاصلضرب چگالی جانبی آنهاست. این چگالی توأم در شکل ۷-۲ با نشان دادن نقاط داده ای که بطور اتفاقی از توزیع آورده شده اند، نشان داده شده است.

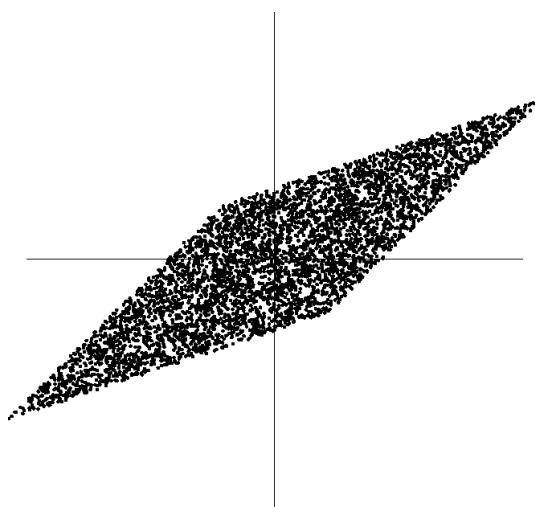


شکل ۲-۷: توزیع توام مؤلفه های s_1 (محور افقی) و s_2 (محور عمودی) با توزیع یکنواخت

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

$$A_0 = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \quad (13-2)$$

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، مثل x_2 را از روی مقدار دیگر 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 دانستن مقدار s_1 اصلاً کمکی به حدس زدن مقدار s_2 نمی کند.



شکل ۲-۸: توزیع توام مخلوطهای مشاهده شده s_1 (محور افقی) و s_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.

۴-۴-۱-۲- 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 نداهد و بالعکس. در بالا هم همانطور که اشاره شد این شرط برای y_1 و y_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 $p(y_1)$ تابع توزیع احتمال حاشیه ای y_1 باشد (یعنی زمانی که y_2 تنها در نظر گرفته شود):

$$p_1(y_1) = \int p(y_1, y_2) dy_2 \quad (14-2)$$

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:

$$p(y_1, y_2) = p_1(y_1)p_2(y_2) \quad (15-2)$$

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:

$$E\{h_1(y_1)h_2(y_2)\} = E\{h_1(y_1)\}E\{h_2(y_2)\} \quad (16-2)$$

۴-۵-۱-۲- 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:

$$E\{y_1y_2\} - E\{y_1\}E\{y_2\} = 0 \quad (17-2)$$

اگر متغیرها مستقل باشند، ناهمبسته هم هستند که در رابطه ۱۶-۲ اگر $y_1 = h(y_1)$ و $y_2 = h(y_2)$ 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 $(1, 0)$ و $(0, -1)$ و $(-1, 0)$ ، دارای احتمال 0.25 برابر است. طبق رابطه ۱۶-۲ ناهمبسته هستند. از سوی دیگر $E\{y_1^2 y_2^2\} = 0 \neq 0.25 = E\{y_1^2\}E\{y_2^2\}$ است، بنابراین شرایط رابطه ۱۶-۲ برقرار نیست و نمی توان آنها را مستقل دانست.

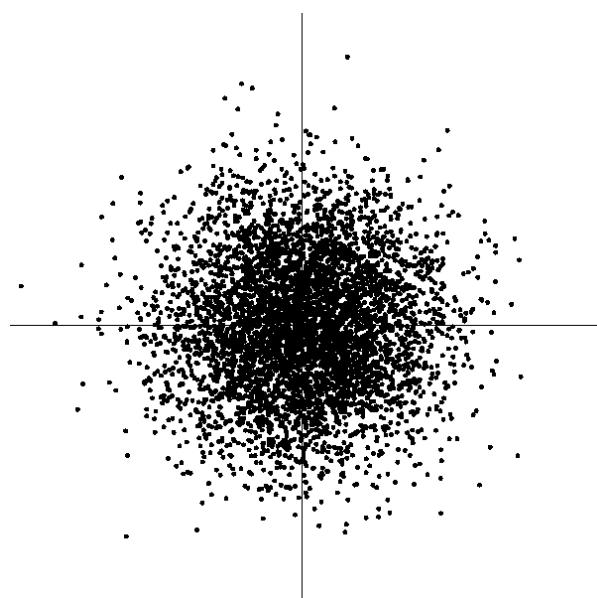
از آنجایی که استقلال به طور ضمنی ناهمبستگی را در بر دارد، بسیاری از روش‌های ICA فرایند تخمین خود را طوری تحدید می کنند که در هر مرحله تخمین مؤلفه های مستقل ناهمبسته باشند. اینکار تعداد پارامترهای آزاد را کاهش می دهد و مسئله را ساده سازی می کند.

۴-۱-۲- چرا متغیرهای گوسی ممنوع هستند؟

محدودیت اساسی روش ICA اینست که مؤلفه های مستقل ICA باید غیر گوسی باشند تا روش کار بکند. برای آنکه روش 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:

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

این توزیع در شکل ۹-۲ به تصویر کشیده شده است. این تصویر نشان می دهد کهتابع 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 ۹-۲: توزیع چند متغیره دو متغیر مستقل گوسی

به عبارت دقیقتر، می‌توان 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 می‌توان تا تبدیل متعامد تخمین زد. یعنی برای مؤلفه‌های مستقل گوسی، ماتریس A قابل تشخیص نیست. البته اگر فقط یکی از مؤلفه‌ها گوسی باشد باز هم می‌توان مدل ICA را تخمین زد.

۷-۴-۱-۲- مؤلفه‌های غیرگوسی مستقلند

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

قضیه حد مرکزی، یک نتیجه کلاسیک از نظریه احتمالات، بیان می‌دارد که توزیع جمع متغیرهای تصادفی مستقل تحت شرایط خاصی به سمت توزیع گوسی می‌میل می‌کند. بنابراین جمع دو متغیر تصادفی مستقل معمولاً نسبت به دو تک تک این متغیرها، دارای توزیعی نزدیکتر به گوسی می‌باشد.

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

برای اینکه مشخص باشد که چگونه این به اصل ابتدایی تخمین ICA منجر می‌شود، تغییر کوچکی در متغیرها می‌دهیم و $\vec{z} = A^T \vec{w}$ را اینگونه تعریف می‌کنیم. در نتیجه داریم $\vec{w}^T \vec{x} = \vec{w}^T A \vec{s} = \vec{z}^T \vec{s}$. بنابراین لازم ترکیبی خطی از s_i با وزنهایی که z_i مشخص می‌کند 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 s_i ‌های \vec{z} غیرصفر است (و فرض اینست که s_i ‌ها توزیع یکسانی داشته باشند).

بنابراین \vec{w} برداری در نظر گرفته می‌شود 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 غیرصفر دارد. این بدین معناست که $\vec{w}^T \vec{x} = \vec{z}^T \vec{s}$ برابر یکی از مؤلفه‌های مستقل است!

بنابراین بیشینه کردن غیرگوسی بودن $\bar{W}^T \bar{W}$ یکی از مؤلفه های مستقل را به ما می دهد در واقع نمودار بهینه سازی غیرگوسی بودن فضای n بعدی \bar{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 جتسجو کرد. این کار متناظر با متعامد کردن^{۴۳} در فضایی است که به طور مناسبی تبدیل یافته است (تبدیل واپتینینگ^{۴۴}). باید توجه داشت که این روش مطرح شده یک هیوریستیک نیست و اثبات آن موجود است و در ادامه خواهد آمد.

۸-۱-۲-۴- معیارهای غیرگوسی بودن

برای استفاده از غیرگوسی بودن در تخمین ICA، باید معیاری کمی از غیرگوسی بودن متغیری تصادفی مانند $\|z\|$ داشته باشیم. برای سادگی فرض می کنیم که z دارای میانگین صفر است و واریانس آن واحد است. یکی از کاربردهای پیش پردازشها ای نظیر به مرکز آوردن، واپتینینگ، و تبدیلات روی سریهای زمانی برای تامین پیش فرضهای این ساده سازیهای است.

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

$$\tilde{\kappa}(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (19-2)$$

از آنجایی که فرض می کنیم که z 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 z ، گشتاور مرتبه چهارم برابر $(E\{y^2\})^2 - 3$ 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 پلتی 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 بزرگ و در مقادیر بزرگ متغیر نیز مقدار زیادی دارند ولی در مقادیر میانی کوچک هستند. یک مثال رایج برای آن توزیع لپلاس 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

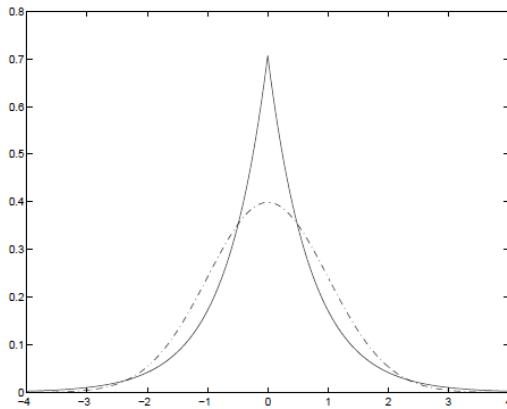
⁴⁸ leptokurtic

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 ۲۰-۲ نشان داده شده است. از سوی دیگر، متغیرهای تصادفی زیرگوسی معمولاً دارای توزیع مسطوحی 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 ۱۲-۲ است.



شکل ۲۰-۲: تابع توزیع توزیع لابلائسین که یک توزیع زیرگوسی رایج است و برای مقاسیه توزیع گوسی نیز به طور خط چین در آن رسم شده است.

معمولًا 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 محدودی از متغیرهای تصادفی غیرگوسی دارای کورتوسیس صفر هستند ولی می توان از آنها چشم پوشی کرد.

کورتوسیس یا مقدار مطلق آن به طور گستردگی 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 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 $A\vec{x} = \vec{y}$ توجه کرد. فرض کنیم

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

یکبار دیگر از $\vec{A}^T \vec{w} = \vec{z}^T \vec{s} = z_1 s_1 + z_2 s_2$ بهره گرفته می شود. خواهیم داشت

بر اساس خصوصیات کورتوفسیس داریم $(\tilde{K}(s_1) + \tilde{K}(s_2)) = z_1^4 \tilde{K}(s_1) + z_2^4 \tilde{K}(s_2)$. 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 \vec{z} منتهی می شود که $E\{y^2\} = z_1^2 + z_2^2 = 1$. از نظر هندسی این بدین معناست که بردار \vec{z} به دایره ای با شعاع واحد روی صفحه ای دو بعدی محدود است. حال مسئله بهینه سازی عبارتست از یافتن بیشینه های تابع $|z_1^4 \tilde{K}(s_1) + z_2^4 \tilde{K}(s_2)|$ روی دایره واحد. برای سادگی می توان فرض کرد که کورتوفسیس ها علامت مشابه داشته باشند که در این صورت می توان از علامت قدر مطلق صرف نظر 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 \vec{z} برابر صفر و مابقی غیر صفر باشند. به دلیل مقید بودن به دایره واحد، عناصر غیر صفر باید ۱ یا -۱ باشند. ولی این نقاط دقیقاً آنها بیست که \vec{z} برابر یکی از مؤلفه های مستقل s_i است، و مسئله حل می شود.

در عمل ما از یک بردار وزن \vec{w} شروع می کنیم، راستایی را که در آن $\vec{w}^T \vec{x} = y$ قویترین رشد را می کند (اگر کورتوفسیس مثبت باشد) و یا قویترین کاهش را دارد (اگر کورتوفسیس منفی باشد) را بر اساس نمونه های موجود $\vec{x}(1), \dots, \vec{x}(T)$ از بردار مخلوط \vec{x} محاسبه می کنیم، و از یک روش شبیب نوری یا توسعه ای از آن برای پیدا کردن بردار جدید \vec{w} استفاده می کنیم. مثالی که زده شد می تواند به هر چند بعدی که دلخواه است تعمیم داده شود و نشان داده شود که کورتوفسیس می تواند به عنوان یک معیار برای بهینه سازی مسئله ICA استفاده شود. بهر حال کورتوفسیس 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 براساس تنها چند اندازه گیری از دنباله توزیع تخمین زده شود که می توانند مشاهداتی دارای خطأ و یا بی ربط باشند. به عبارت دیگر کورتوفسیس معیار مقاومی برای غیر گوسی بودن نیست. بنابراین گاهی اوقات معیارهای دیگری برای محاسبه غیر گوسی بودن مناسبترند. در ادامه معیار نگانتروپی^{۵۳} معرفی می شود که خواصی تقریباً متضاد با خواص کورتوفسیس دارد و راههایی برای تخمین آن ارائه می شود.

دومین معیار مهم غیر گوسی 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

⁵¹optimization landscape

⁵²outliers

⁵³negentropy

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

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

تعريف 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 \vec{y} با توزیع $(\vec{y})^f$ به صورت زیر تعریف می شود:

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

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 آنtronپی 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. آنtronپی برای توزیعهایی که در مقادیر خاصی متمرکزند اندک است و چنین توزیعهایی یا کاملاً خوش 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 مشبت باشد، می توان از یک نسخه کمی تغییر کرده آنtronپی تفاضلی به نام نگانtronپی استفاده کرد. نگانtronپی J به صورت زیر تعریف می شود:

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

که در آن \vec{y}_{gauss} یک متغیر تصادفی گوسی با کواریانسی برابر با کواریانسی \vec{y} است. به دلیل خواصی که اشاره شد، نگانtronپی همیشه مشبت است و زمانی صفر است که \vec{y} دارای توزیع گوسی باشد. نگانtronپی ویژگی جالب دیگری نیز دارد که در مقابل تبدیلات خطی قابل بازگشت^{۵۴}، نامتغیر باقی می ماند.

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

تخمین زدن نگانtronپی چنان که اشاره شد، دشوار است و در عمل باید از برخی از تقریبها استفاده کرد. در اینجا تنها به تقریبها بسته می شود که خاصیتهای مناسبی دارند و در تولید یک روش کارا برای ICA استفاده می شوند. روشهای کلاسیک برای محاسبه نگانtronپی 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 \bar{z} فرض می شود که دارای میانگین صفر و واریانس واحد است. با اینحال اعتبار چنین تقریبهايی ممکن است محدود باشد و این روش در قیاس با کورتوسیس از پایداری کمتری برخوردار باشد. برای اجتناب از مشکلاتی که این تقریبهاي نگانتروپی ممکن است ایجاد کند، تقریبهاي دیگری نیز تهیه شده است که بر مبنای اصل آنتروپی بیشینه استوار شده است. به طور کلی این تقریب بدست می آيد:

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

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 عیار 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 $y = G(u)$ این رابطه مبتنی بر کورتوسیس دقیقاً بدست می آید.

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

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

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

که $1 \leq a_1 \leq 2$ انتخاب مناسبی است. بنابراین تقریبهاي از نگانتروپی بدست آمد که مصالحه خوبی را بین خصوصیات دو معیار کلاسیک غیر گوسی بودن یعنی کورتوسیس و نگانتروپی 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 غیر گوسی بودن ختم می شود.

با بهره گیری از مفهوم آنتروپی تفاضلی، می‌توان اطلاعات متقابل I بین m متغیر تصادفی اسکالر y_1, \dots, y_m که است را اینگونه تعریف کرد:

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

اطلاعات متقابل یک معیار طبیعی برای تعیین میزان وابستگی بین متغیرهای تصادفی است و در واقع معادل افتراق کولبک-لیبلر⁵⁵ بین توزیع تؤام $(\vec{y})f$ و حاصلضرب توزیعهای حاشیه ای است که معیاری بدیهی برای محاسبه استقلال آماری می‌باشد. 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 می‌گیرد و برخلاف PCA و روش‌های مشابه تنها به کواریانس بسنده نمی‌کند. اطلاعات متقابل را با توجه به تفسیر آنتروپی می‌توان طول کد تفسیر کرد. عبارت $H(\vec{y})$ طول کد را هنگامی که i -ها جدا جدا کد می‌شوند نشان می‌دهد و $H(\vec{y})$ طول کد را وقتی که \vec{y} به صورت یک بردار تصادفی کد می‌شود (یعنی تمام مؤلفه‌ها توسط یک کد، کدگزاری شوند) بدست می‌دهد. بنابراین اطلاعات متقابل نشان می‌دهد که چقدر کاهش طول کد در اثر کد کردن تمام بردار بجای کد کردن مجزای مؤلفه‌ها بدست می‌آید. کلاً با استفاده از کدگزاری تمام بردار کدهای بهتری بدست می‌آید. با اینحال اگر i -ها مستقل باشند هیچ اطلاعاتی از یکدیگر بدست نمی‌دهند و کد کردن جداجدای مؤلفه‌ها از نظر طول فرقی با کد کردن یکجای آنها ندارد. یک خاصیت مهم اطلاعات متقابل اینست که برای هر تبدیل بازگشت پذیر $\vec{y} = W\vec{x}$ داریم:

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

حال اگر 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:

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

و این نکته بدین معناست که W باید ثابت باشد. افزون بر این برای y_i که دارای واریانس واحد است، آنتروپی و نگانتروپی 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 نگانتروپی را نشان می‌دهد.

از آنجایی که اطلاعات متقابل یک معیار طبیعی نظریه اطلاعات برای استقلال متغیرهای تصادفی است، می‌توان از آن به عنوان معیاری برای یافتن تبدیل ICA استفاده کرد. در این روش که جایگزینی برای روش تقریب مدل است، W یک بردار تصادفی \vec{x} به صورت یک تبدیل برگشت پذیر مانند معادله $\vec{W}\vec{x} = \vec{s}$ تعریف می‌شود که در آن ماتریس W طوری تعریف می‌شود که اطلاعات متقابل مؤلفه‌های تبدیل شده \vec{s} 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 نگانتروپی بیشینه می‌شود. و یا به عبارت دیگر معادل یافتن زیرفضایی یک بُعدی است که افکنشها در آن بیشترین نگانتروپی را داشته باشند. رابطه ۳۳-۲ 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 ICA را ساده می‌کند که این ساده شدن از رابطه ۳۱-۲ تا رابطه ۳۳-۲ مشهودست. بنابراین دیده می‌شود که فرمولاسیون ICA به صورت کمینه کردن اطلاعات متقابل اثبات دیگری برایده ذهنی یافتن غیرگوسی ترین جهت‌هاست.

۱-۲-۴-۱۰- تخمین درستنمایی پیشینه

یک روش بسیار متداول برای تخمین مدل ICA تخمین درستنمایی بیشینه است که قربت 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 $(\vec{W}_1, \dots, \vec{W}_n)$ نشان داده شود، درستنمایی لگاریتمی شکل زیر را می‌گیرد (Pham and Garat 1992):

$$L = \sum_{t=1}^T \sum_{i=1}^n \log f_i(\vec{w}_i^T \vec{x}(t)) + T \log |\det W| \quad (34-2)$$

که در آن i -نیتای توزیع s_i است (که در اینجا دانسته فرض شده است). عبارت $\log |\det \mathbf{W}|$ 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 \mathbf{W} ، توزیع $\vec{y} = \mathbf{W}\vec{x}$ برآورده است با $p_x(\mathbf{W}\vec{x})|\det \mathbf{W}|$

یکتابع 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 می شود که \vec{x} ورودی شبکه عصبی است و خروجی آن به شکل $(\vec{w}_i^T \vec{x}) \phi_i$ است که ϕ_i یک تابع اسکالار غیرخطی است و \vec{w}_i بردار وزنهای نورونهاست. برای بیشینه کردن آنتروپی خروجی داریم:

$$L_2 = H(\phi_1(\vec{w}_i^T \vec{x}), \dots, \phi_n(\vec{w}_i^T \vec{x})) \quad (\textcircled{3} \Delta - 2)$$

اگر ϕ خوب انتخاب شود این چارچوب می‌تواند ICA را نیز تخمین بزند. در واقع پژوهشگران (Cardoso 1989) نتایج شگفت‌انگیزی را منتشر کرده‌اند 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 که در شبکه عصبی استفاده می‌شوند برابر تابع توزیع تجمعی توزیعهای f_i باشند، یعنی $f'_i(\cdot) = \phi'_j(\cdot)$.

⁵⁶ Infomax principle

⁵⁷ network entropy maximization principle

برای دیدن ارتباط بین درستنماهی و اطلاعات متقابل، امید ریاضی درستنماهی لگاریتمی گرفته می شود:

$$\frac{1}{T} E\{L\} = \sum_{i=1}^n \log f_i(\vec{w}_i^T \vec{x}) + \log |\det W| \quad (36-2)$$

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

در هر حال تفاوت کوچکی وجود دارد که ممکن است در عمل خیلی به چشم بیاید. مشکل تخمین درستنماهی بیشینه اینست که توزیعهای f_i باید درست تخمین زده شده باشد، البته دقت بالای آنها مهم نیست بلکه کافیست که زیرگوسی Cardoso 1989; Lee, Mumford et al. 1998; Hyvärinen and Oja 2000. در بسیاری از حالات دانش پیشین کافی از مؤلفه های مستقل موجود است و نیازی به تخمین حالت ذاتی آنها از داده ها وجود ندارد. در هر حال اگر اطلاعاتی که در مورد ذات مؤلفه های مستقل در دست است نادرست باشد، تخمین ML نتایج کاملاً غلطی بدست می دهد. بالعکس اگر معیار غیرگوسی بودن مناسبی استفاده شود این مشکل معمولاً رخ نمی دهد.

۲-۱-۴-۱۱-۲- ارتباط ICA و تعقیب افکنش

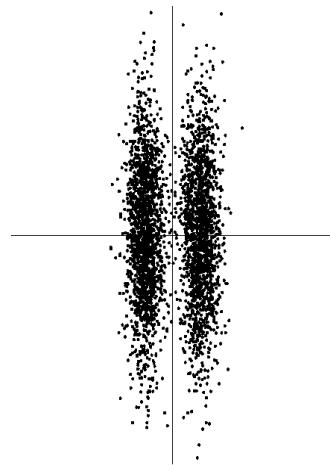
تعقیب افکنش (Friedman and Tukey 1971; Huber 1985; Friedman 1987) تکنیکی آماری است که افکنشهای «جالب» برای داده های چندبعدی را می یابد. چنین افکنشهایی می توانند برای مصور سازی بهینه داده ها، تخمین توزیع و رگرسیون استفاده شود. در نسخه یک بعدی تعقیب افکنش، این روش به دنبال جهتیابی می گردد که افکنش روی آنها دارای توزیع جالبی باشد، یعنی ساختار خاصی را نشان دهد. بحث شده است (Huber 1985) که جالبترین جهتیابی آنها بی هستند که کمترین خاصیت گوسی را از خود نشان می دهند که دقیقاً همان نکته ایست که برای تخمین ICA استفاده می شود.

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

بنابراین در حالت کلی ICA را می توان نسخه ای از تعقیب افکنش در نظر گرفت. تمام معیارهای غیرگوسی بودن و الگوریتمهای ICA مربوطه را می توان به نوعی راهکارهای تعقیب افکنش دانست. البته تعقیب افکنش این اجزه را می دهد که وقتی مؤلفه های مستقل s_i کمتری از متغیرهای اصلی x_i وجود دارد نیز کار را ادامه داد. برای این منظور این روش فرض می کند که ابعادی از فضا که توسط مؤلفه های مستقل پوشش داده نشده اند پر از نویز گوسی هستند. بدین طریق با محاسبه

⁵⁸projection Pursuit

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



شکل ۱۱-۲: نمایشی از روش تعقیب افکنش و جالب بودن افکنشهای غیرگوسی. داده‌هایی که در شک دیده می‌شود آشکارا به دو خوش تقسیم می‌شود. با اینحال مؤلفه اصلی یعنی راستای واریانس بیشینه عمودی است که تمایزی بین خوش‌ها قائل نمی‌شود. از سوی دیگر جهت شدیداً غیرگوسی تعقیب افکنش افقی است که بهترین جداکنندگی را برای خوش‌ها ایجاد می‌کند.

۱۲-۴-۱-۲- نتیجه گیری

روش ICA یک تکنیک همه منظوره آماری است که در آن داده‌های تصادفی مشاهده شده با یک تبدیل خطی به مؤلفه‌هایی تبدیل می‌شوند که تا حد امکان مستقلند و توزیع جالبی دارند. می‌توان ICA را به عنوان یک مدل تخمین متغیرهای نهفته فرمول کرد. راهکار شهودی استفاده از بیشترین غیرگوسی بودن برای تولین توابع‌های هدف گوناگونی استفاده شود که بهینه سازی آنها به تخمین مدل ICA می‌انجامد. از سوی دیگر با مدد از روش‌های کلاسیکتری مانند مانند تخمین درستنمایی بیشینه یا کمینه کردن اطلاعات متقابل می‌توان 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 می‌برد. کاربردهای متعددی برای ICA موجود است که از آن قبیل پردازش صوت، پردازش سیگنالهای زیستی، مخابرات، اقتصاد و پردازش تصویر را می‌توان نام برد. در این پژوهش از روند مطالعات، ایده‌های مطرح شده، و راهکارهای حل مسائل ICA در زمینه‌های دیگر برای تکوین مدلی برای پردازش تصویر استفاده شده است که به آنها مجدداً اشاره خواهد شد.

۲-۲-۱- روش مطالعات عصب شناسی

۲-۲-۲- مطالعات عملکرد گروهی نورونها

از دهه ۱۹۸۰ میلادی، پزشکان از اسکترهای MRI^{۵۹} برای تشخیص بیماریهای بافت نرم استفاده می‌کردند. این ماشینها با قرار دادن بخشی از بدن درون یک میدان مغناطیسی قوی و اسکن آن کار می‌کنند. سپس میدانهای ضعیفتری با زاویه‌های خاصی نسبت به میدان اصلی به بدن اعمال می‌شوند که باعث می‌شوند که هسته هیدروژن موجود در بافت‌های بدن (که عمدها در مولکولهای آب و چربی یافت می‌شوند) 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. داده‌های جمع آوری شده توسط اسکترهای MRI در تصاویری سه بعدی گرد هم آوره می‌شوند که به پزشکان امکان تشخیص ناهنجاریهای بسیاری را بدون هیچ عملیات تخریبی می‌دهند.

دانشمندان بسیار زود متوجه شدند که فناوری MRI می‌تواند راهی را نیز برای تعیین فعالیت ارگانهای خاص فراهم کند. ایده اصلی اینکار اسکن سریعتر ولی کم دقیق‌تر است که عکس‌هایی بسیار فوری از کارکرد پویای بدن بدست می‌دهد. این متداول‌وزیر به عنوان MRI کارکردی^{۶۰} یا اختصاراً fMRI شناخته می‌شود.

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 می‌کند 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 می‌توان 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 می‌توان انجام داد یافتن ارتباط بخش‌هایی از مغز که فعالیت بیشتری دارند با نوع حرکه‌ایی است که به فرد نشان داده می‌شود. به عنوان مثال در پژوهشی که در سال ۱۹۹۹ توسط کانویشر^{۶۱} و همکاران توسط fMRI صورت گرفت مشخص شد که بخشی موسوم به^{۶۲} FFA در مغز انسان به چهره‌های انسانی (وتا حدی کمتر به سر انسان) واکنش نشان می‌دهد 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 می‌شود و مبتنی بر این ایده است که دو محرك که کمی با هم متفاوتند (معمولًاً یکی از آنها دارا یک ویژگی افزونه نسبت به دیگری است) به فرد تحت نظر نشان داده می‌شود و بخش‌های جدیدی از مغز نیز علاوه بر بخش‌های قبلی بکار می‌افتد که تاثیر آن ویژگی را نشان می‌دهند.

⁵⁹ magnetic resonance imaging

⁶⁰functional MRI

⁶¹Kanwisher et al.

⁶²fusiform face area

⁶³cognitive subtraction

این روش به دلیل ماهیت ماقروسکوپیک 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 می‌پردازند و کارکرد آنرا به ازای محرکهای مختلف می‌سنجدند که به طور کلی آنها را ضبط تک نورون^{۶۴} می‌نمند. به عنوان مثال برای این بررسی این نورونها باید در موجود زنده (غلب گربه یا میمونهای ماکاکو^{۶۵}) یک پروب را دقیقاً در سلول مورد نظر قرار داد. برای اینکار از میکروسکوپهای لیزری استفاده می‌شود تا از ورود پروب به همان نورون اطمینان حاصل شود. سپس دامنه فعالیت الکتریکی نورون به ازای محرکهای گوناگون روی الکتروسکوپ نمایش داده می‌شود و تابعی که نورون از آن تعییت می‌کند کشف می‌شود. محرکهای استفاده شده معمولاً شامل پنجرهای ساده برای بررسی عملکرد سلولهای ساده و خاصیت پاسخها، گوشه‌ها و زاویه‌ها، شکلهای ساده و گهگاه شکلهای رسته‌هایی خاص از اشیا است.

۲-۲-۳- مطالعات آسیب‌شناسی

بعضی از این مطالعات شامل ایجاد آسیبهای عمدی^{۶۶} در هر بخش‌هایی از مغز می‌میمون ها مثلاً در بخش‌های مختلف مسیر بینایی آنهاست. بقیه هم شامل مطالعه بر روی انسان است که آسیب‌های مغزی در انسان را که در بخش خاصی از مغز بوده است نشان می‌دهد. برای مثال دو نمونه از کشفياتی که از این طریق حاصل شده است در پایین آورده می‌شود:

نورونهایی در ناحیه IT یافت می‌شود که اختصاصاً به تصاویر چهره انسانی پاسخ می‌دهند. 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 مديدة به دنبال علت بیماری پروسپاکتوزیا^{۶۸} بودند که در آن بیمار با وجود اینکه دارای بینایی سالمی در بقیه قسمتهاست، در تشخیص چهره دچار مشکل می‌شود. این عارضه در پی آسیبهایی به بخش فوق مخطط قشر خاکستری بینایی بوجود می‌آید که تا حد زیادی بیانگر اینست که نورونهای IT در آن دخیل هستند.

ویژوال اگنوزیا^{۶۹} که معمولاً در پی آسیبهای دو طرفه پس سری-گیجگاهی بوجود می‌آید، عارضه ایست که فرد را در تمایز شکل اشیا ناتوان می‌سازد. حتی گاهی بیمار قدرت تمیز خوبی برای رنگها و بافتها دارد. مبتلایات به ویژوال اگنوزیا در تشخیص شکل هندسی اشیا، حروف الفبا و نقاشی‌های سیاه و سفید ضعیف عمل می‌کنند ولی در حرکت دادن دستشان به سمت یک شی، چرخاندن مچشان برای وارد کردن یک شی به یک محفظه، یا قرار دادن درست انگشتانشان برای بلند

⁶⁴single cell recording

⁶⁵Macaque Monkeys

⁶⁶selective lesions

⁶⁷intracellualar

⁶⁸prosopagnosia

⁶⁹Visual agnosia

کردن یک شئ 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. بر عکس، در بیماری اپتیکال آتاکسیا⁷⁰ افرادی که بخش آهیانه ایمفر 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.

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

مدلهای اولیه نورونهای قشر خاکستری مخ بر مبنای آزمایشاتی بنا شد که سعی داشتند خصوصیات پاسخ مشاهده شده به مجموعه محدودی از محركها - میله ها، پنجره ها، و شکلهای ساده - را به نورونهای فعال شده نسبت دهند و هرجا 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 محرك برای تعیین توابع پاسخ نورونها استفاده می شد (De Boer and Kuyper 1968; Jones and Palmer 1987; Van Steveninck and Bialek 1988; Schwartz, Pillow et al. 2006 و نوع غیرخطی بودن چند فرض می کردند تا پاسخ نورون را تعریف کنند. این مدلها میدانهای گیرندهای⁷¹ 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 نورونهای مراتب بالاتر کردند (Rust, Schwartz et al. 2005; Touryan, Felsen et al. 2005). بدلیل محدودیتهای محاسباتی این روش تنها وقتی مناسب است که نورونها دارای خصوصیات پاسخ نسبتاً ساده ای باشند. به علاوه توزیع محرك باید تصادفی باشد و از خصوصیات خاصی تبعیت کند (مثلاً باید دارای تقارن کروی باشد) (Paninski 2003). این مسئله از آنروی اهمیت دارد که استفاده از این روش را هنگام مطالعه پاسخها به تصاویر طبیعی مسدود می کند (یا دست کم بایس زیادی را در آن ایجاد می کند). این پاسخها بسیار غیرگویی هستند و آمارگان آنها چندان شناخته شده نیست (Rust and Movshon 2005). از سوی دیگر محرك 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 توانند به اندازه کافی از نورونهای 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 افرون بر اینها انتخاب یک مجموعه خاص از این محركها تفسیری از کدگزاری نورونی را به دنبال خواهد داشت که اجباراً مناسب نیست (مثلاً نسبت دادن تحلیل فوریه، لبه یابی، کدگزاری شکل به بخشها ای از سیستم بینایی). اگر نواحی بینایی خصوصیات مجردتری را از تصویر کدگزاری می کنند، فعالیت نورونی برای همه محركهایی که یک خاصیت خاص را دارند بدون تغییر خواهد بود. آزمودن با یک محرك نمی تواند این نامتفاوتی را آشکار کند و یا خاصیت مجرد کدگزاری شده را کشف کند. در همان حال، تعامل بین نورونها می تواند الگوهای پاسخ پیچیده ای تولید کند که با ضبط محدود تک نورونها بسادگی قابل شرح نباشد.

⁷⁰optical ataxia

⁷¹receptive field

در تحلیل های آماری از تصاویر، معمولاً مدلی در نظر گرفته می شود که نحوه 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 مستقل⁷² از تصاویر طبیعی را نام برد (Bell and Sejnowski 1997). مؤلفه های مستقل برایک 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 شدمدها به عنوان مسئول محاسبات پردازشی 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 بخشهايی از مغز که ناشناخته باقی مانده اند از اهمیت بسزایی برخوردار است ولی تاکنون مدلهاي 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 کدگزاری نورونی ورودیهای تصویری پیچیده لازم هستند. این پژوهش در راستای این نیاز قرار گرفته است و از مدلهاي نظری برای پیش بینی ویژگیهای تک تک سلولها و نیز جمعیت نورونها استفاده می کند.

۳-۲- شبکه های عصبی برای بازنمایی

۳-۱- قوانین یادگیری

۱-۲-۳- 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 زیر 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 (w_1, \dots, w_n) را به سمت نخستین مؤلفه اصلی داده ها \tilde{x} سوق می دهد. در این الگوریتم ساده یادگیری، اندازه W محدود نشده است. اوجا⁷³ این الگوریتم را طوری تغییر داد که اندازه W پس از هر گام نرماییزه 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 N 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 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. لینکسر⁷⁴ 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 شانون⁷⁵

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

مقدار اطلاعاتی را که دانستن خروجی y درباره ورودی x منتقل می کنبدست می دهد و معادل اطلاعات متقابل بین آنهاست، $I(x, y)$. برای یک خروجی تنها با توزیع گوسی، رابطه ۴۰-۲ با بیشینه کردن واریانس خروجی بیشینه می شود[124]. بیشینه کردن واریانس خروجی با قید 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 ۴۰-۲ 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).

۲-۳-۲- 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 نولید 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 کندي دارد مگر در موز اشیاء، الگوریتمهای یادگیری که صریحاً وابستگیهای آماری ورودی را کدگزاری می کنند می کوشند که این پایداریها⁷⁶ 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 کدگزاری 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 بیشینه کردن تفاوت بین P و Q کشف کند. عامل P توسط پاسخ هه مجموعه داده با این تنظیمات فعلی وزنها تعریف می شود و به طور صریح می تواند با نمونه گیری تمام⁷⁷ حالت ممکن Q واحد ورودی محاسبه شود. قانون یادگیری GMAX محدود به یادگیری یک واحد تک خروجی و واحدهای باینری احتمالاتی است.

پکر⁷⁸ 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 که چگونه سیگنالهای آموزش هدایت شده از درون سیستم ادراکی می

⁷⁶constancy

⁷⁷Becker

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، a و 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 a و b از رابطه زیر بدست می‌آید:

$$I = \frac{1}{2} \log \frac{V(y_a + y_b)}{V(y_a - y_b)} \quad (41-2)$$

در این رابطه V تابع واریانس روی حالت‌های آموزشی است. بیشینه کردن I مربع تفاضل بین خروجی دو مژول را بسته به اینکه چقدر دو مژول با تغییر ورودی تغییر کرده‌اند 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 مژول را به عنوان تابع هزینه استفاده کرد (مخرج کسر ۴۱-۲). استفاده از تابع هزینه کمترین مربع تفاضل به تنها یک هر دو مژول را مجبور می‌کند که یک خروجی ثابت را بسانند 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 مژول به چیزی که در ورودی هر دو مشترک است پاسخ دهند. بکر و هینتون نشان داده‌اند که بیشینه کردن اطلاعات متقابل بین قسمتهای مجاور از لحاظ مکانی تصویر می‌تواند عمق را در نقاط اتفاقی 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 می‌زان عدم تطابقی که اندک اندک تغییر می‌کند تشکیل شده است. پس از آموزش، با وجود اینکه هیچ دانش پیشین در مورد بعد 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.

شراودolf 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 افزوده‌ان 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 هینتون^{۷۹} الگوریتم IMAX را در مسئله یادگیری بازنمایی پارامترهای نمای اشیای ساده، مانند اندازه شی، مکان و مقیاس آن به کار بردند. 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 تراز اشیای دو بعدی ساختگی دچار محدودیت است زیرا محاسبه دترمینان ماتریسهای بد ترکیب^{۸۰} دشوار است.

۲-۳-۲- خواص مغز

یکی از شگفت انگیزترین خصوصیات سازماندهی مغز اینست که بسیاری از سطوح گیرندهای پیرامونی^{۸۱} (مانند مسیر شبکیه چشم، مجرای حلقه‌ی گوش میانی، و سطح پوست) در مرحله‌های متوالی پردازشی مغز، بازنمایی توپوگرافیکی دارند. گروههای سلولی همسایه در شبکیه به گروههای سلولی همسایه در بخش بینایی تalamوس مرتبط است که آن هم به نوبه خود با ناحیه‌های همسایه‌ای در قشر بینایی اتصال دارد. بدین ترتیب یک نقشه نورونی منظم از اطلاعات اخذ شده در گیرندهای سطحی در مراحل متوالی مغز حفظ می شود. چنین نقشه‌های نورونی نه تنها موقعیت گیرندهای را منعکس می کند، بلکه به کمک چگالی اتصالات میزان حساسیت حسگر به حرکتها را نیز مشخص می کند. بطور مثال بخش مرکزی شبکیه، فوویا^{۸۲}، دارای بالاترین چگالی گیرندهای در میان بخش‌های شبکیه است و از اینرو از بیشترین دقق و تیزبینی نیز برخوردار است. متناظر با این نکته بخشی از قشر خاکستری مخ که به اطلاعات رسیده از فوویا اختصاص دارد از ناحیه‌ای که بخش جانبی شبکیه را بازنمایی می کند بزرگتر است که در ناحیه متناظر با بخش جانبی چگالی گیرندهای 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 خارجی^{۸۳} (LGN) که در تalamوس قرار دارد به لکه نوری دیده شده در بخش خاصی از میدان دید پاسخ می دهد. آکسونهای چند نورون همسایه تalamوس روی سلولهای بخش اولیه بینایی قشر خاکستری مخ همگرا می شوند که در آن هر نورون تنها زمانی آتش می کند که چینش 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 پاسخهایست 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 کنند(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).

۴-۲- 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 روباتیک و ... را به خود جلب کرده است. انسان در تشخیص چهره ها، اشیا و صدایها 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.

۴-۱-۱- فرایند بینایی

۴-۱-۱- چشم

همانطور که در برش عرضی چشم دیده می شود (شکل ۱۲-۲) 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 عنبه^{۸۹})، عدسی و در آخر مایع زجاجیه^{۹۰} (مایع غلیظتر از مایع زلالیه) می باشد-(McGill-University 2006).

⁸⁴conjunctive

⁸⁵retina

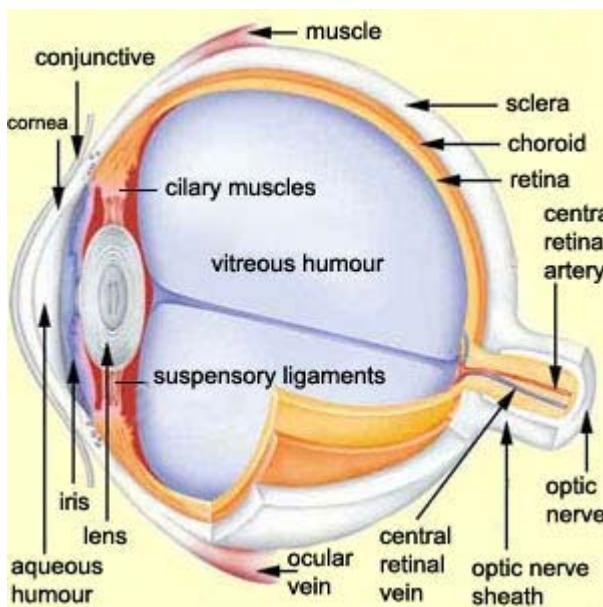
⁸⁶cornea

⁸⁷aqueous humor

⁸⁸pupil

⁸⁹iris

⁹⁰vitreous humor

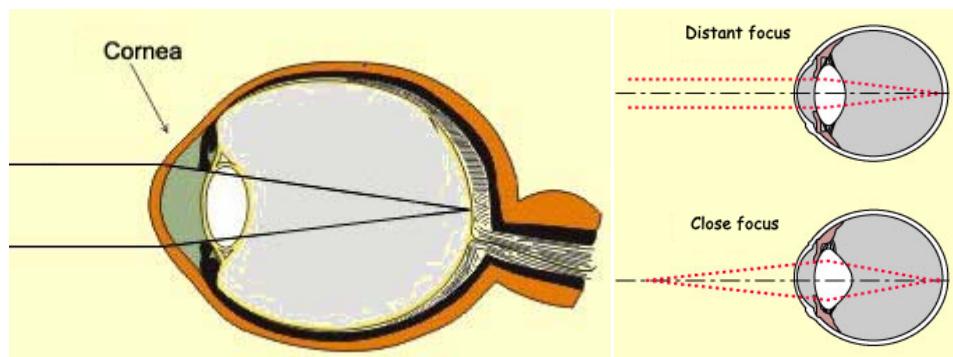


شکل ۱۲-۲: برش عرضی چشم

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 پر عروق 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 رسد. 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 کند. اما بقیه پرتوها از همین فاصله دور به بخش های منحنی شکل قرنیه برخورد می کند. این پرتوها انحراف درونی پیدا کرده و دقیقاً در نقطه مرکزی روی شبکیه می افتد 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 فالتری را در تمرکز تصویر روی شبکیه ایفا می کند. هر چقدر پرتوهای نوری رسیده به چشم از اشیای نزدیکتر باشد بیشتر از هم دور می شوند، بنابراین باید بیشتر شکست پیدا کنند تا بتوانند روی پرده شبکیه به هم

^{۹۱} sclera^{۹۲} ocular muscle^{۹۳} choroid^{۹۴} pigment^{۹۵} 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) برگرد. 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 ۱۳-۲).



شکل ۱۳-۲: 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)، عملأً خالی از رگ های خونی می باشد. در مرکز لکه، یک گودی کوچک نقطه سیاهی به قطر ۲ میلیمتر

⁹⁶suspensory ligaments

⁹⁷ciliary muscle

⁹⁸accommodation

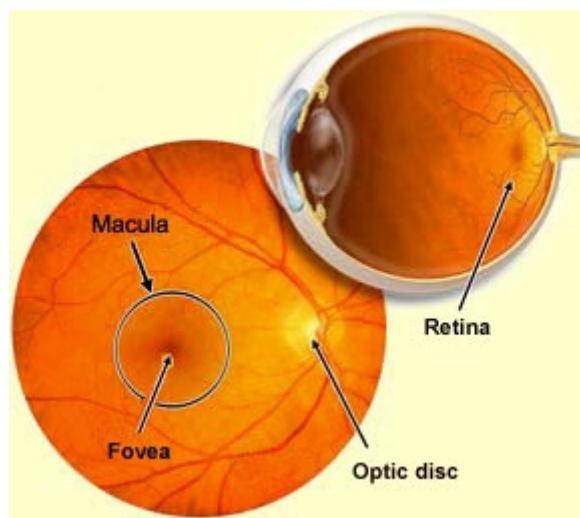
⁹⁹ophthalmoscope

¹⁰⁰optic disk

¹⁰¹blind spot

¹⁰²macula

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 دقیق تجربه می شود (شکل ۱۴-۲).



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

۱۴-۲-۱-۴-۲- قسمت انتهایی عصب بینایی

آکسون های سلولهای عقدہ ای شبکیه^{۱۰۴} از میان عصب بینایی، کیاسماهی 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 تalamوس قرار گرفته است. مقصد اصلی برای مجرای بینایی LGN می باشد (شکل ۱۵-۲).

در برش عرضی، LGN های سمت راست و چپ در یک ساختار ۶ لایه از انواع متفاوت سلولی سازماندهی شده اند (شکل ۱۶-۲). در بررسی سه بعدی، این ۶ لایه شبیه دسته ای از دیسکهای متحدمالمرکز 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.

¹⁰³visual acuity

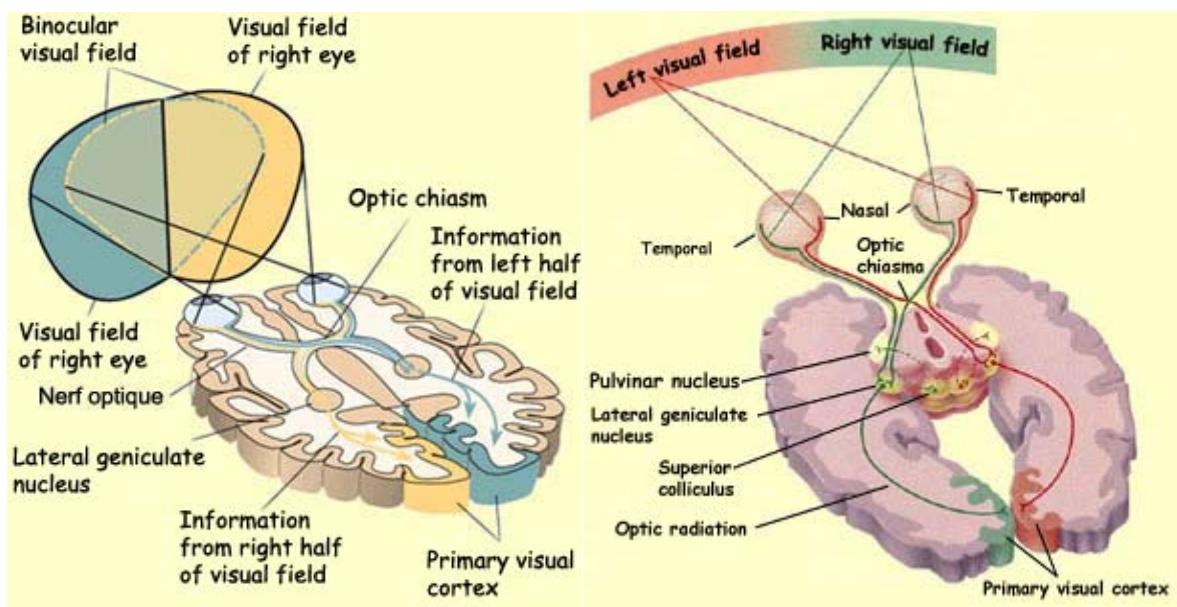
¹⁰⁴retina ganglion cells

¹⁰⁵optic chiasm

¹⁰⁶optic tractus

¹⁰⁷Nasal retina

¹⁰⁸Temporal retina



شکل ۱۵-۲: میدان دید یا میدان گیرنده‌ی چشم ناحیه‌ای از فضا است که وقتی چشم روی نقطه دوری متتمرکز است، روی شبکیه می‌افتد. میدان دید دو چشم تا حد زیادی همپوشانی دارند ولی میدان راست به راست کشیده‌تر است و بالعکس. ناحیه میانی که در آن میدان دید دو چشم همپوشانی دارند، میدان دید دو چشمی خوانده می‌شود. تصویر سمت چپ همچنان نشان می‌دهد که چگونه میدان بینایی چشم راست در قشر بینایی سمت چپ تحلیل می‌شود و بالعکس. تصویر سمت راست نواحی عبور مسیرهای جدأگانه سیگنال بینایی را نشان می‌دهد.

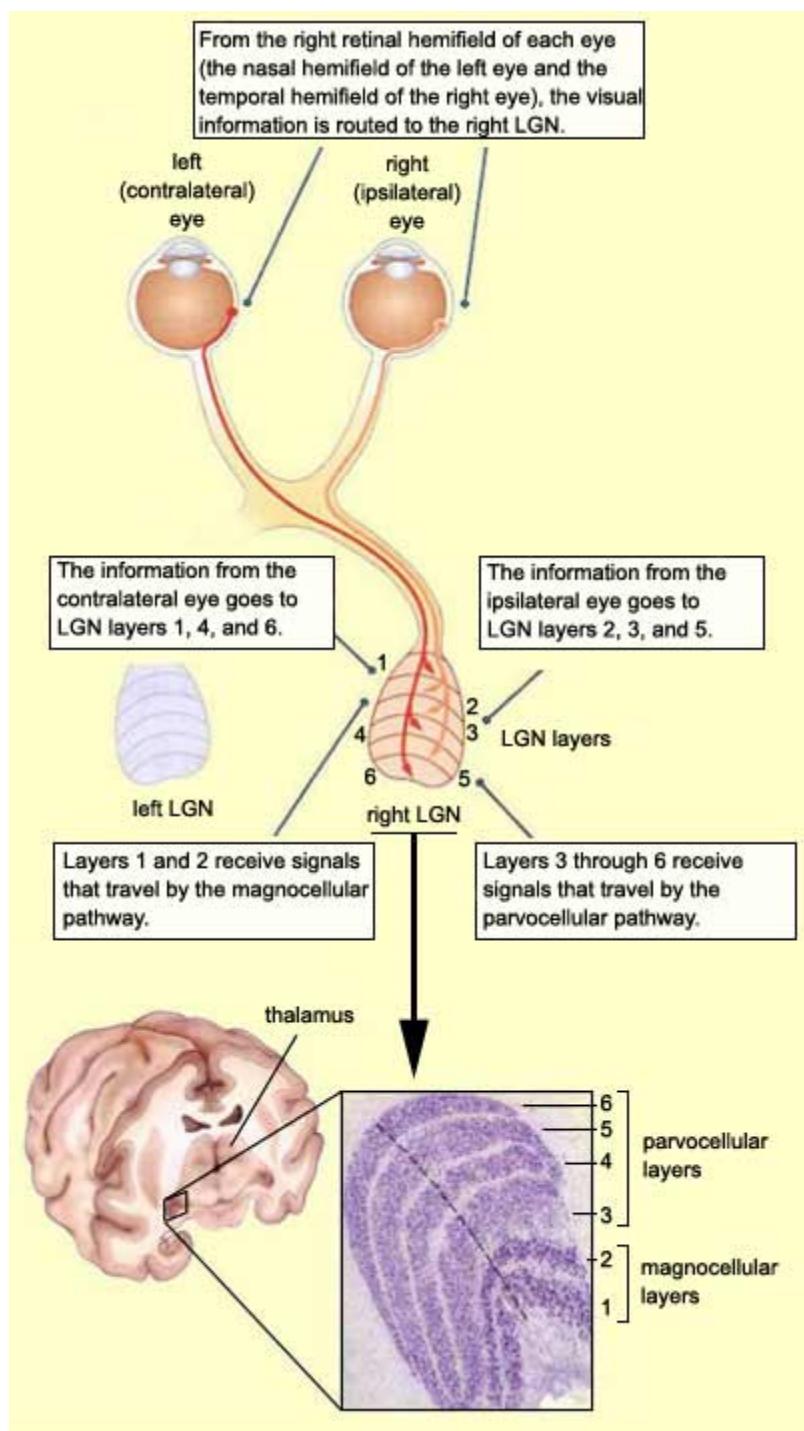
تحقیقات نشان می‌دهد که آکسون‌های سمت چشم راست روی لایه‌های ۲ و ۳ و ۵ در LGN سینپاپس دارند در حالیکه آنهایی که از چشم سمت چپ هستند به لایه‌های ۱ و ۴ و ۶ می‌روند. این بررسی‌ها همچنان نشان می‌دهد که لایه‌های درونی LGN (۱ و ۲) شامل نورون‌های بزرگتری نسبت به لایه‌های خلفی (۳ و ۴ و ۵ و ۶) هستند که به ۲ لایه درونی، لایه‌های درشت سلول^{۱۰۹} (M) گفته می‌شود در حالیکه چهار لایه دیگر، لایه‌های ریزسلول^{۱۱۰} (P) نامیده می‌شود که در واقع سلول‌های عقده‌ای نوع M به لایه‌های درشت سلولی LGN می‌روند و سلول‌های عقده‌ای نوع P به لایه‌های ریزسلول می‌روند.

آکسون‌های برخی از سلول‌های عقده‌ای شبکیه از راه بینایی به ساختارهایی به غیر از LGN (که رابط اصلی بین شبکیه و قشر بینایی است) انشعاب می‌یابد (شکل ۱۷-۲). یکی از این ساختارها، هیپوتalamوس و به ویژه هسته فوق کیاسما^{۱۱۱} آن است که تعداد مشخصی از آکسون‌های شبکیه را دریافت می‌کند. هسته فوق کیاسما بخش اصلی ساعت بیولوژیکی داخلی بدن بشمار می‌آید. سیگنال‌های بینایی که از آکسون‌های شبکیه می‌رسند پیوسته درباره تاریکی و روشنایی محیط اطراف اطلاعاتی را گردآوری می‌کنند، بنابراین این امکان فراهم می‌شود که طیف وسیعی از ریتم‌های بیولوژیکی مانند شامل خواب و بیداری که به چرخه شبانه روز مرتبط است با هم هماهنگ شوند.

¹⁰⁹magnocellular

¹¹⁰parvocellular

¹¹¹superchiasmatic nucleus



شکل ۲-۱۶: ساختار کلی LGN

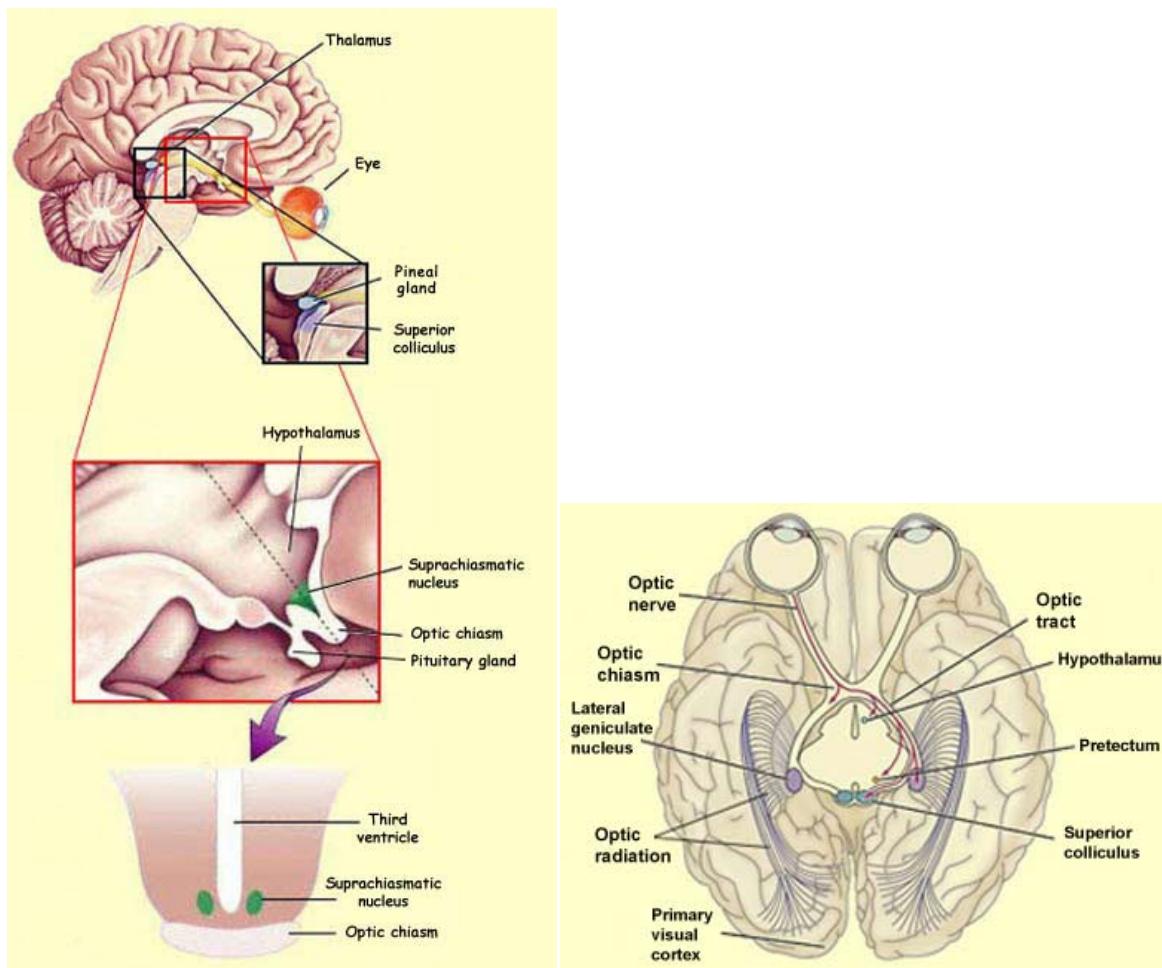
آکسون هایی که از برخی دیگر از سلول های عقده ای شبکیه می آیند به پرتکتوم^{۱۱۲} می روند. پرتکتوم شی از مغز میانی است که میزان بازشدن مردمک و حرکت های خاصی از چشم را کنترل می کند. سرانجام حدود ۱۰٪ از آکسون هایی

¹¹² prepectum

پایان نامه کارشناسی ارشد کورش مشگی - پاییز ۱۳۸۹

دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر

که از شبکیه می آیند به یک بخش ریشه ای در مغز میانی که کالیکولوس فوکانی 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: مکان عقده های سر راه سیگنالهای بینایی

در حقیقت کالیکولوس فوکانی 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 همدمیگر می افتدند، یک پرتوی نوری تابیده شده بر روی شبکیه، تعداد زیادی از نورونهای کالیکولوس فوکانی 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 حرکت می شود که در آغاز در میدان دید محیطی ظاهر می شود. کالیکولوس فوکانی مانند LGN ارتباطاتی از قشر بینایی اولیه دریافت می کند. نورون های کالیکولوس فوکانی به نوبه خود آکسون هایشان به ساختارهای زیر قشری مانند ساختار مشبك^{۱۱۴}، کالیکولوس تحتانی^{۱۱۵} و نخاع می رود. همچنین نورون های کالیکولوس فوکانی دو ساختار دیگر را تحت تاثیر قرار می دهند: LGN و هسته بالشتک^{۱۱۶}.

^{۱۱۳}retinotectal pathway

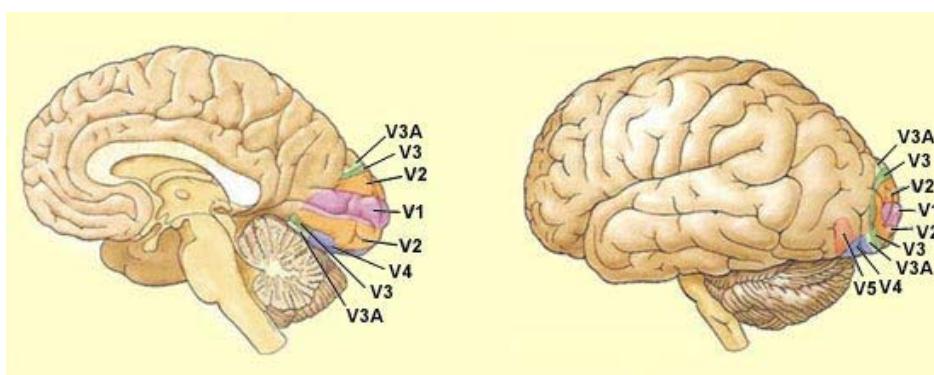
^{۱۱۴} reticular formation

هسته بالشتک، یک هسته در بخش پشتی تalamوس استکه تحت تاثیر مجرای بینایی از طریق LGN می باشد (از طریق آکسون های متوازی). مثل خیلی از هسته های دیگر تalamوس، 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 آوردن 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 یافتد. 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.

۳-۱-۴-۲- 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 V5, V4, V3 (MT), PO و ... احاطه شده اند (شکل ۱۸-۲). به هر صورت از روی همه این پیچیدگی ها می توان یک الگوی کلی استخراج کرد.



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

به دنبال مطالعات انقلابی منتشر شده توسط آنگرلیدر^{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

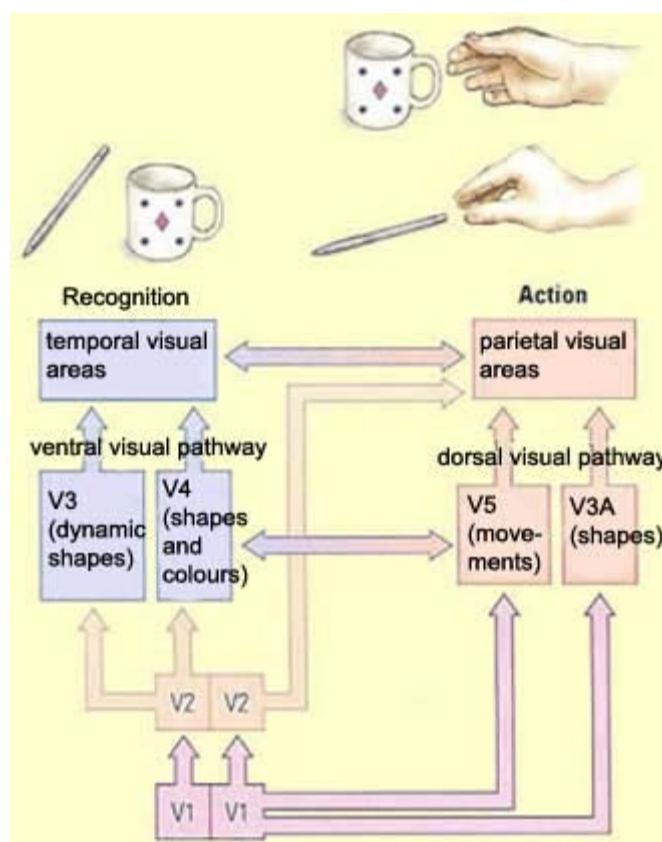
¹¹⁶pulvinar

¹¹⁷attention

¹¹⁸associative areas

¹¹⁹Leslie Ungerleider

تعیین موقعیت آنها در فضای واقع دو سیستم قشری اصلی برای پردازش اطلاعات بینایی وجود دارد: مسیر بینایی بطنی^{۱۲۱} که تا پخش 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 ذاتیانها مانند شکل و رنگمی باشد. عملکرد مسیر بینایی خلفی نیز ظاهرآ کنترل حرکت بینایی روی اشیاء با پردازش خاصیت های خارجیانه است. در خاصیتهای بینظیر اندازه، موقعیت، و راستایشیا در فضای 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 آسیبهای عمده 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 ۱۹-۲: نواحی قشری موجود در مسیرهای بینایی و ارتباطات آنها

امروزه این گمان وجود دارد که عملکرد اصلی مسیر بینایی خلفی، هدایت اعمال بلاذرنگی 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

¹²¹ventral pathway

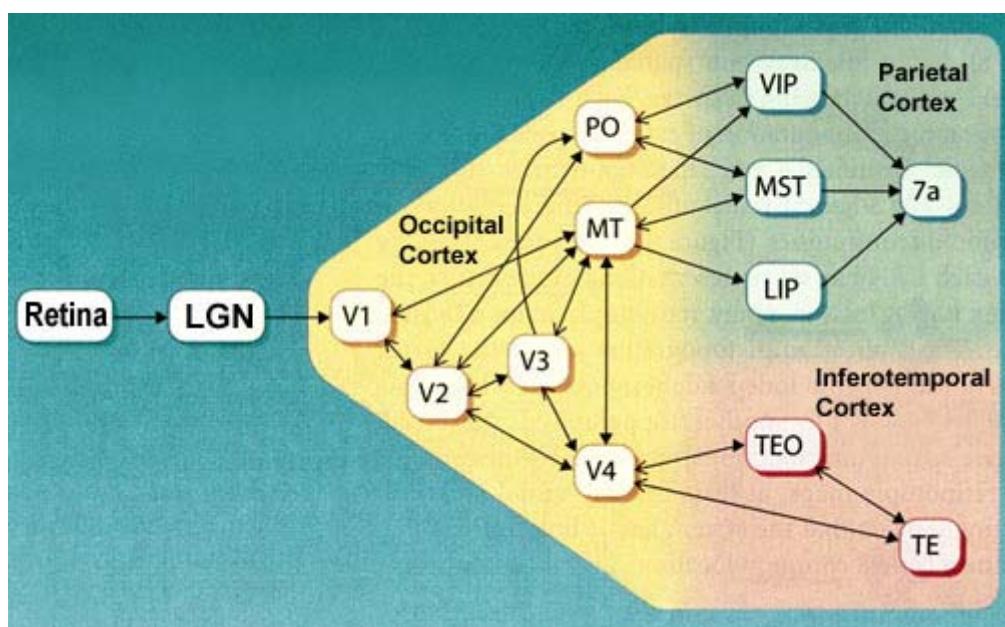
¹²²dorsal pathway

¹²³dichotomy

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 ۱۹-۲). 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 مشابه ای 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 همچنان 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 ۲۰-۲: 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.

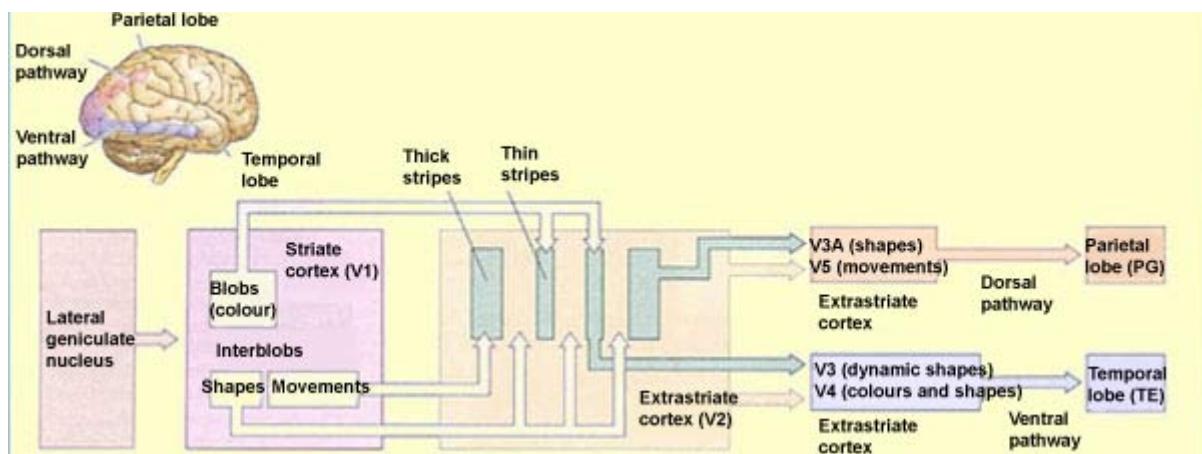
¹²⁴action pathway

¹²⁵medial temporal

¹²⁶medial superior temporal

¹²⁷ventral and lateral inferoparietal

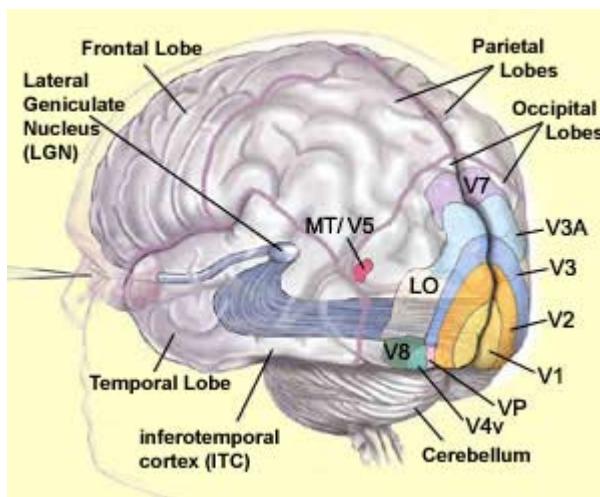
ناحیه V5 یا (MT) عمدتاً به ادراک حرکتی مربوط می شود. این ناحیه تصاویر V2 و V3 را دریافت می کند. همچنین این ناحیه تصاویر لایه IV-B در ناحیه V1 را دریافت می کند، همان لایه ای که به طور قابل توجهی بخشی از کانال درشت سلول است که شامل تحلیل حرکت اشیاء می شود. این کانال همچنین تا ناحیه V2 ادامه می دهد، جایی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



شکل ۲۱-۲: 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 کanal P-IB در گذشته و کanal M در حال حاضر نسبت داده شود.

برای سلول های ناحیه MT به نظر می رسد که حرکت یک شئاز 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 V1 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 محرك خواهند داشت. اما در ناحیه MT به طور عادی سلول های بسیاری که به مسیرهای عمودی پاسخ می دهند، به طور قابل قبولی به دو محرك با زاویه ۴۵ درجه نیز پاسخ دهند و به عبارت دیگر به مسیر ظاهری حرکت پاسخ می دهند.



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

پس از ناحیه MT (شکل ۲۲-۲)، نواحی دیگری مانند ناحیه MST وجود دارد که شامل تحلیل حرکت است. سلول های این ناحیه نه تنها مانند سلول های ناحیه MT به حرکت خطی حساسند، بلکه به حرکت شعاعی (رفت و برگشت به یک نقطه) و حرکت دایره ای (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 می روید در اطراف شما رخ می دهد.

حتی نورون های مشخصی در ناحیه فوق گیجگاهی چند حسی^{۱۲۸} (STP) وجود دارد که به حرکت های بیولوژیکی که برای بقای زندگی افراد لازم است، پاسخ می دهند. از جمله این موارد می توان به شناسایی عضوی از گونه جانوری مشابه از روی طرز راه رفتن او اشاره کرد.

مغز شما ممکن است این حجم زیاد از داده های مرتبط با حرکت را که به وسیله مسیر خلفی جمع آوری شده، برای مقاصد زیادی استفاده کند: برای استخراج اطلاعات مربوط به اشیایی که وقتی حرکت می کنید در میدان بینایی شما هستند، یا به منظور هدایت حرکت های شما، یا جهت هماهنگی حرکت های چشمی شما، یا جهت تشخیص حرکت اشیای اطراف شما که به بقای زندگی شما کمک می کند (مثالاً جهت دقیق حرکت اتومبیلی که به سمت شما می آید).

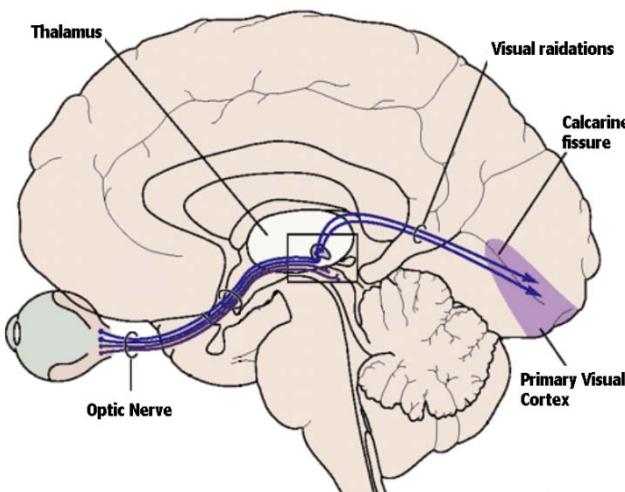
۲-۴-۲- مدل‌های رفتاری بخش بینایی قشر خاکستری مخ

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

¹²⁸superior temporal polysensory

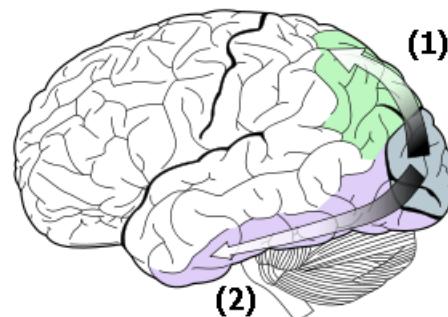
¹²⁹ primary visual cortex

خصوصیات مشابه دارد. این ناحیه معادل ناحیه ۱۷ در تقسیم بندی برادمن^{۱۳۰} می‌باشد. قشر ابتدایی بینایی که به نام V1 نیز شناخته می‌شود خروجی نورون‌های خود را به نواحی مختلفی می‌فرستد.



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

پردازش تصویر ورودی در قشر بینایی پس از ناحیه V1 در دو مسیر کلی به نام‌های مسیر خلفی و مسیر بطئی ادامه می‌یابد (شکل ۲-۲). یکی از نظریات رایج در مورد پردازش در قشر بینایی، جدایی تشخیص شئ از تشخیص مکان اشیا می‌باشد (Mishkin, Ungerleider et al. 1983). مطابق با این نظریه، داده‌های مربوط به تصویر در قشر بینایی مغز در دو مسیر جداگانه پردازش می‌شوند. مسیر بطئی از V1 شروع می‌شود و پس از عبور از V2 و V4 به IT ختم می‌شود. در این مسیر، اطلاعات مربوط به هویت اشیا در تصویر استخراج می‌شوند. در مسیر دیگر که از V1 به قشر آهیانه می‌رود اطلاعات مربوط به مکان اشیا پردازش می‌شود. در این پژوهش تمرکز روی مسیر بطئی است که اشیا موجود در تصویر را تشخیص می‌دهد.

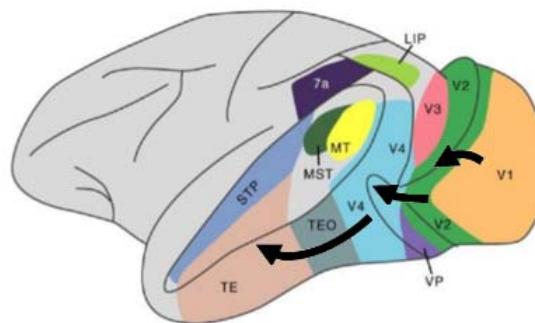


شکل ۲-۲: دو مسیر پردازش داده‌ها در قشر بینایی. داده‌های تصویر در قشر بینایی در دو مسیر بطئی و خلفی پردازش می‌شوند. مسیر خلفی با پیکان شماره (۱) و مسیر بطئی با پیکان شماره (۲) مشخص شده‌اند.

¹³⁰ Brodmann

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 واحدهای تعییر شدند که خصوصیاتی را از تصویر استخراج می کنند. هابل در این مقاله، نورون های V1 را به دو دسته تقسیم کرد: نورون های ساده که خصوصیاتی مانند تعییر 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 ای اتصال این نواحی در شکل ۲-۲ نشان داده شده است. در این سلسله مراتب، هر ناحیه از ناحیه قبلی ورودی دریافت می کند و خروجی خود را به ناحیه بعدیمی فرستد. در این سلسله مراتب، خصوصیاتی که نورون ها در تصویر تشخیص می دهند به تدریج پیچیده تر شده و مقاومت آنها در برابر تعییرات بیشتر می شود.

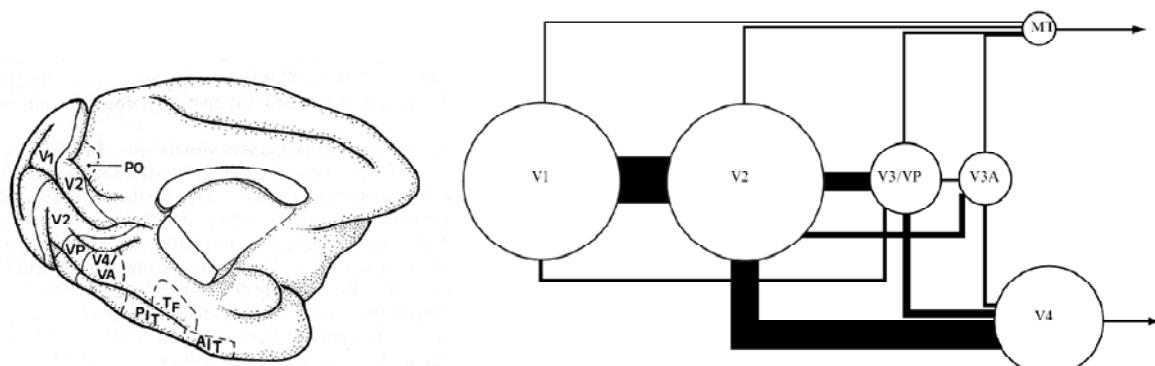


شکل ۲-۲: مسیر بطنی در قشر بینایی. مسیر بطنی شامل نواحی V1، V2، V3، V4 و دو ناحیه TE و TEO در IT می شود. در شکل بالا پیکان های سیاه رنگ جریان داده ها در طول این مسیر را مشخص می کنند.

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 نسبتهای مکانی را کدگزاری 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 کدگزاری می کنند(Felleman and Van Essen 1991; Lennie 1998).

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



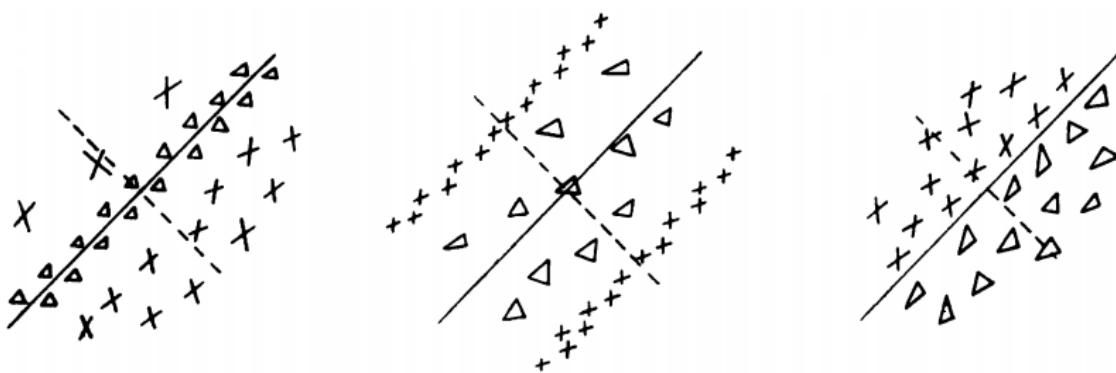
شکل ۲۶-۲: نواحی بینایی قشر مخ میمون ماکاکو و اندازه نسبی در بخش پس سری قشر مخ

برای نشان دادن هرچه بیشتر پیچیدگی این موضوع، باید به این نکته اذعان داشت که تحقیقات جدید حاکی از آن هستند که تفاوت‌های بین نواحی قشر خاکستری مخ در واقع بسیار ظرفی 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 می‌برند (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 می‌شود.

V1 - ۳-۴ - لایه

این ناحیه بزرگترین قسمت از قشر بینایی مغز است که حدود ۲۰ درصد از این قشر را تشکیل می‌دهد. این قسمت در قطعه پس سری قرار دارد و مطابق با ناحیه ۱۷ در تقسیم بندی برادمن می‌باشد. 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 1968 گیرندگی^{۱۳۱} می‌باشد. میدان گیرندگی معادل قسمتی از تصویر ورودی است که تغییر در خصوصیات آن، فعالیت نورون را تحت تاثیر قرار می‌دهد. در ناحیه V1، نورون‌ها دارای ناحیه دریافت کوچکی هستند که اندازه آن در نقطه مرکزی دید حدود 1° (معادل 7×7 پیکسل) می‌باشد. ناحیه دریافت در تصویر محلی می‌باشد و هر نورون قسمت خاصی از تصویر را از طریق ناحیه دریافت خود مشاهده می‌کند. شکل ۲۷-۲ نمونه‌ای از ساختار ناحیه دریافت برای نورون‌های V1 را نشان می‌دهد. در این شکل، مثلث نقاطی در ناحیه دریافت را نشان می‌دهد که تاباندن نور به آنها نورون را از فعالیت باز می‌دارد. ضربدر نقاطی را نشان می‌دهد که تاباندن نور به آنها فعالیت نورون را تشدید می‌کند. همانطور که مشخص است مرز کاملی بین نقاط بازدارنده و نقاط وادرانده وجود دارد. در نورون‌های V1، این مرز در جهت‌های مختلفی مشاهده می‌شود.

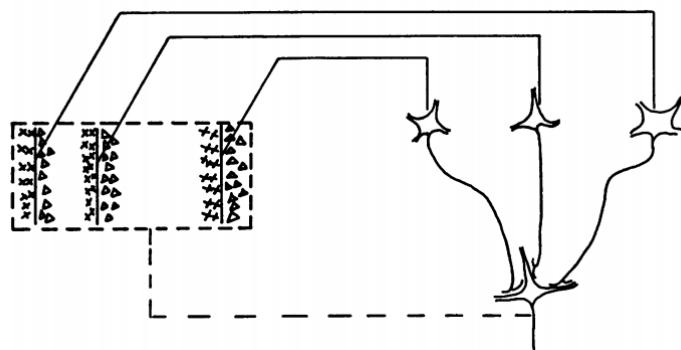


شکل ۲۷-۲: ساختار ناحیه دریافت برای چند نورون در ناحیه V1 (Hubel and Wiesel 1962). مثلث نشان دهنده نقاط بازدارنده و ضربدر نشاندهنده نقاط وادرانده می‌باشد. خط متصل برای هر ناحیه دریافت، جهت آن ناحیه دریافت را مشخص می‌کند. در نورون‌های V1، انواع مختلفی از جهت‌ها بین 0 تا 180 درجه مشاهده می‌شود.

تفسیری که هابل از این نورون‌ها داشت این بود که نورون‌های V1 تغییرات شدت روشنایی را در جهت‌های مختلف تشخیص می‌دهند. می‌توان گفت که این نورون‌ها مانند فیلترهای لبه عمل می‌کنند و لبه‌ها را در جهت‌های مختلف از تصویر استخراج می‌کنند. یکی از وجوده تمایز این نورون‌ها با نورون‌های پیچیده در میزان حساسیت آنها به موقعیت مکانی لبه می‌باشد. نورون‌های ساده به موقعیت مکانی لبه بسیار حساس هستند و با تغییر موقعیت آن، پاسخ این نورون‌ها به سرعت افت می‌کند. از طرف دیگر، نورون‌های پیچیده، لبه در یک جهت خاص را در یک محدوده مکانی که نسبت به نورون‌های ساده گسترده‌تر است تشخیص می‌دهند. این خصوصیات باعث شد که هابل، نظریه پردازش سلسله مراتبی برای این نورون‌ها را ارائه دهد (شکل ۲۸-۲).

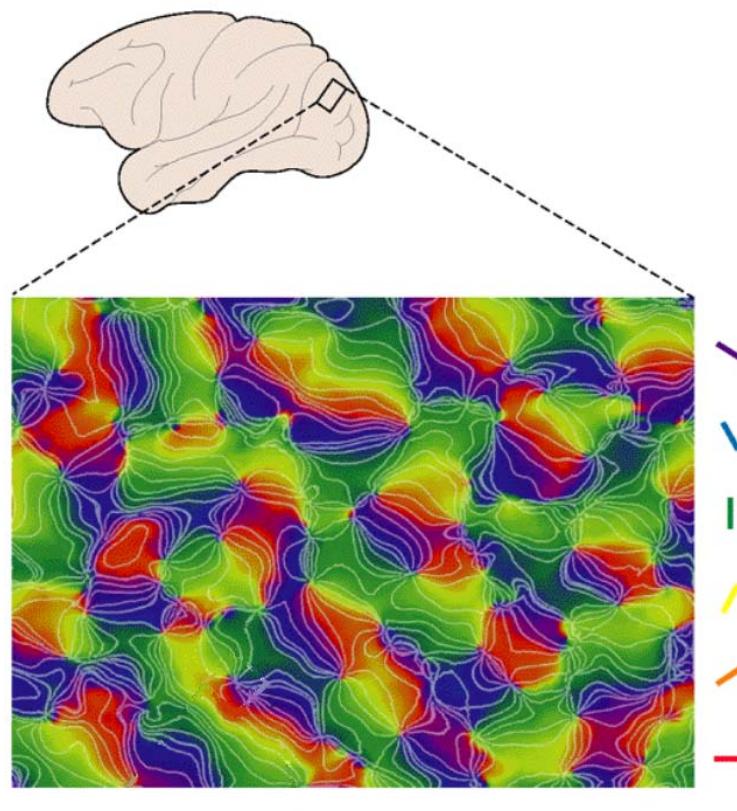
نکته دیگر در مورد نورون‌های V1، ترکیب قرار گرفتن آنها در قشر ابتدایی بینایی با توجه به جهت مورد نظر آنها می‌باشد. گفته شد که نورون‌های V1 انواعی از جهت‌ها را در تصویر ورودی تشخیص می‌دهند. نحوه قرار گرفتن این نورون‌ها در قشر ابتدایی بینایی بدین صورت است که نورون‌های با جهت‌های مشابه در کنار یکدیگر قرار گرفته‌اند. البته بریدگی‌هایی نیز در این فضاییکنواخت مشاهده می‌شود که طیفی از نورون‌ها با جهت‌های مختلف را شامل می‌شود.

¹³¹ receptive field



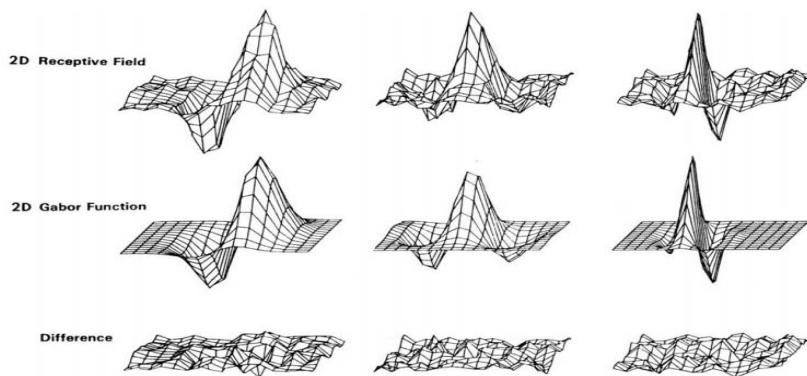
شکل ۲۸-۲: ساختار یک نورون پیچیده ارائه شده توسط هابل. در این شکل، نورون پیچیده در سمت راست از سه نورون ساده در سمت چپ ورودی دریافت می کند. این سه نورون در مکان های مختلف، لبه هاییکسانی را تشخیص می دهند. بدین ترتیب نورون پیچیده قادر است که لبه با جهت خاص را در مکان های مختلف تشخیص دهد.

در شکل ۲۹-۲ نحوه قرار گرفتن نورون های V1 را بر اساس جهت آنها مشهود است. در این شکل، میله های کنار تصویر جهت های متناظر هر رنگ را نشان می دهند. دقیق شود که جهت هایی که نورون های V1 تشخیص می دهند محدود به زوایای ذکر شده نمی باشند. این نورون ها طیف پیوسته ای از جهت ها بین زاویه ۰ تا ۱۸۰ درجه را تشخیص می دهند.



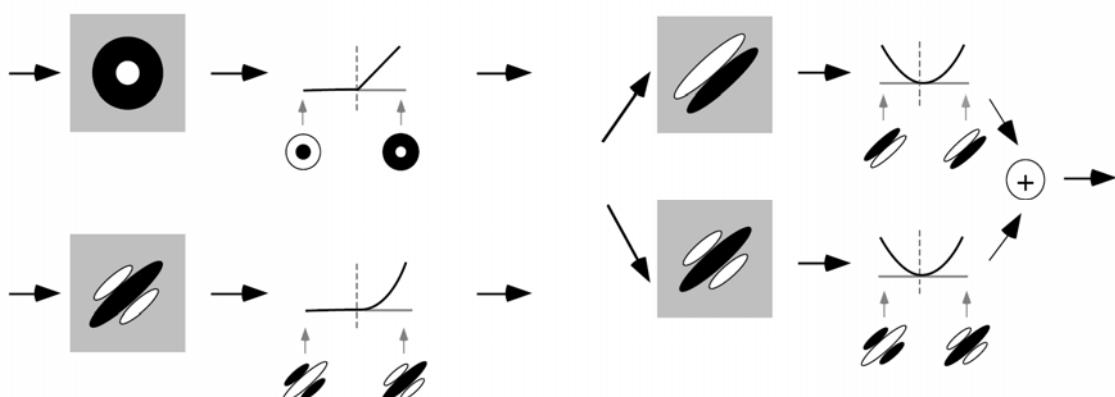
شکل ۲۹-۲: ترکیب قرار گرفتن نورون های V1 بر اساس جهت مورد نظر آنها (Kandel, Schwartz et al. 2001)

امروزه مشخص شده است که سلولهای ساده معمولاً الگوهای محلی راستدار را ترجیح می‌دهند و با یک لایه خطی و بدنبال آن یک لایه غیرخطی بخوبی مدل می‌شوند (شکل ۲-۳۱ پایین-چپ). محرکهایی که سلولهای ساده V1 را تحريك می‌کنند بخوبی توسط تابعهای دوبعدی گابور توصیف می‌شوند (Jones and Palmer 1987) (شکل ۲-۳۰-۲). فرضیه‌ها در مورد سلولهای ساده، به دلیل پاسخ قوی آنها به میله‌ها و پنجره‌ها، آنها را کدگزاری کننده لبه‌ها در نظر می‌گیرند.



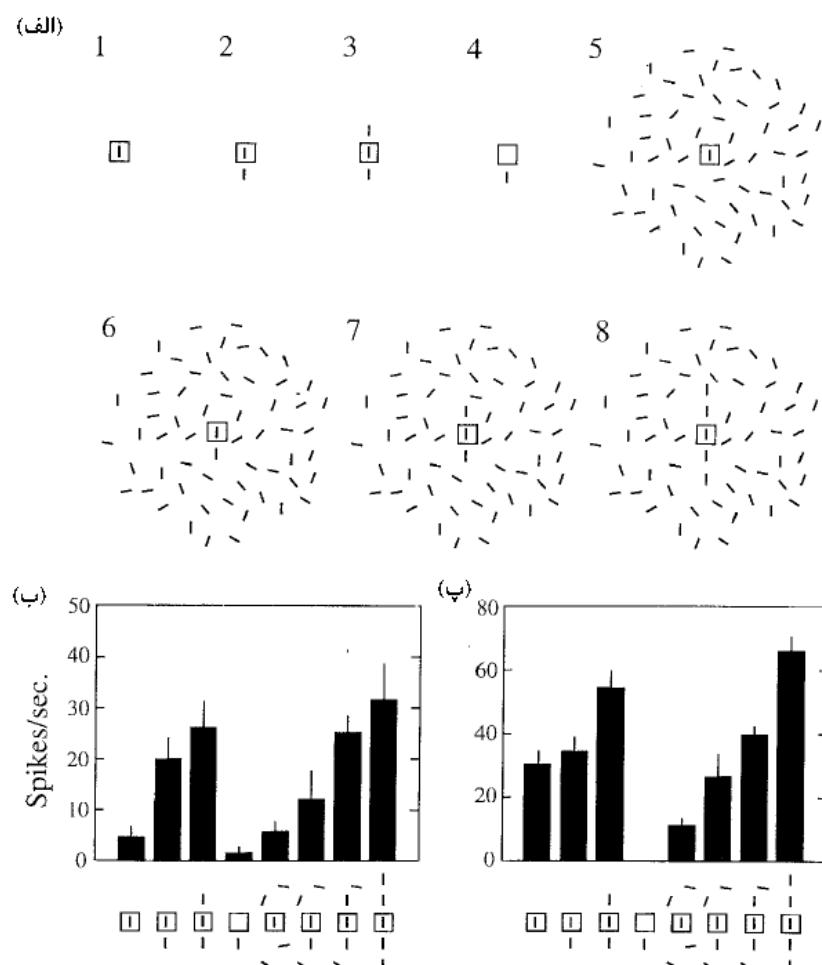
شکل ۲-۳۰-۲: میدان گیرندگی سلولهای ساده V1 به خوبی توسط فیلترهای گابور توصیف می‌شود (Daugman 1989)

سلولهای پیچیده نیز دارای پاسخی قوی به میله‌ها و لبه‌ها هستند ولی به مکان دقیق لبه در میدان گیرندگی خود حساسیت کمتری دارند و معمولاً به لبه‌هایی با قطبیت مخالف نیز پاسخ یکسانی می‌دهند و هنگامی که با پنجره‌های سینوسی تحريك می‌شوند نسبت به فاز آن نامتغیرند. در مدل استاندارد سلولهای پیچیده که به مدل «انرژی» معروف است، دو ویژگی محلی و راستدار (معمولًا با 90° خروج از فاز تابع گابور) با تصویر کانوالو می‌شوند و خروجی آنها مربع شده و جمع می‌شود تا پاسخ نورون را بسازد که شکل ۳۱-۲ آنرا به خوبی نشان می‌دهد (Movshon, Thompson et al. 1978; Adelson and Bergen 1985; Heeger 1992; Heeger, Simoncelli et al. 1996).



شکل ۲-۳۱-۲: 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 گیرنده‌گی گسترش یافته^{۱۳۲} یک نورون می‌باشد. با گسترش یک محرک به خارج از میدان گیرنده‌گی نورون، پاسخ آن نورون تحت تاثیر قرار می‌گیرد. در ناحیه V1 این امر به خاطر پدیده‌ای به نام جمع کردن کانتور^{۱۳۳} مورد توجه قرار گرفته است. نورونهای V1 به قطعات لبه حساس هستند. هنگامی که این لبه‌ها به خارج از میدان گیرنده‌گی نورونهای V1 گسترش می‌یابند پاسخ نورون همچنان تغییر می‌کند و ممکن است تقویت و یا تضعیف شود. این پدیده به وجود اتصالات افقی بین نورونهای V1 نسبت داده شده است زیرا این اتصالات خصوصیاتی از خود نشان داده اند که برای جمع کردن کانتور مناسب به نظر می‌رسد. شکل ۳۲-۲ اثر حضور محرک در میدان گیرنده‌گی گسترش یافته بر پاسخ یک نورون V1 را نشان می‌دهد.

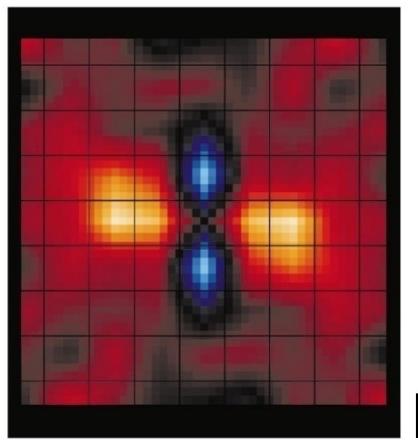


شکل ۳۲-۲: اثر حضور محرک‌ها در ناحیه دریافت گسترش یافته بر پاسخ نورون‌های V1. (الف) ترکیبات مختلف محرک‌ها در تصویر. مربع کوچک در هر شکل ناحیه دریافت اصلی نورون مورد نظر را نشان می‌دهد. (ب) پاسخ نورون به ورودیهای مختلف در (الف). هنگامی که محرک‌ها تشکیل یک خط بدنه‌ند و در راستای یکدیگر باشند، پاسخ نورون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

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 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 تنده ناهموار داشته باشند این اتصالات بازدارنده هستند. شکل ۳۳-۲ اتصالات افقی مشاهده شده در لایه V1 را نشان می دهد.



شکل ۳۳-۲: 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.

۴-۴-۲- لایه V2

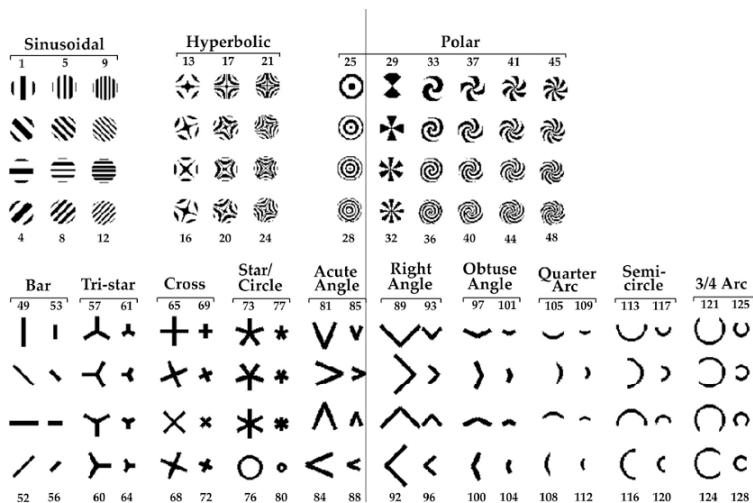
دومین ناحیه از سلسله مراتب پردازش در مسیر بطنی، ناحیه V2 می باشد. این ناحیه از نظر حجم تقریباً برابر با ناحیه V1 است و ورودی خود را از این ناحیه دریافت می کند. علیرغم اهمیت این ناحیه در پردازش بینایی، حجم مطالعاتی که در مورد این ناحیه انجام شده است بسیار کمتر از دیگر نواحی می باشد. با اینحال، در سال 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 میبایند. 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 خصوصیاتی 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

Hegdé and Van Essen 2000; Boynton and Hegdé 2004; Hegdé (and Van Essen 2007).

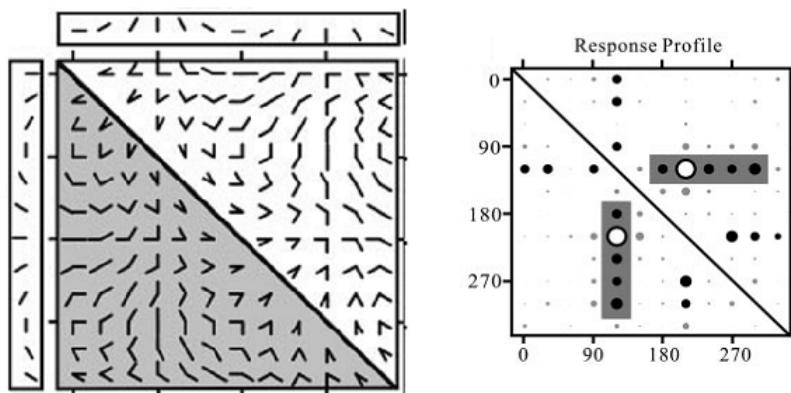
در تایید این ایده، آزمایشاتی انجام شده است که با انتخاب هوشمندانه مجموعه محرک‌ها، توانسته‌اند اطلاعات مناسبی در مورد خصوصیات نورون‌های V2 جمع آوری کنند. به عنوان مثال، در مرجع (Hegdé and Van Essen 2003) مجموعه‌ای از کانتورها و زوایا با جهت‌های مختلف برای تحریک نورون‌های V2 استفاده شده‌اند. این مطالعات با استفاده از تکنیک‌های ضبط تک نورون، توانسته‌اند پاسخ نورون‌های V2 را به مجموعه محرک ذکر شده ضبط کنند. مجموعه محرک در شکل ۳۴-۲ نشان داده است.



شکل ۳۴-۲: مجموعه محرک استفاده شده در (Hegdé and Van Essen 2003) برای آزمایش خصوصیات نورون‌های V2. این مجموعه شامل کانتور با زوایای مختلف و محرک‌های مشبک برای آزمایش تشخیص بافت می‌باشد.

در مجموعه‌ای از آزمایش‌ها که با این مجموعه محرک بر روی نورون‌های ناحیه V2 انجام شد این فرضیه تقویت شد که نورون‌های V2 به محرک‌های کانتور بیشتر از محرک‌های بافت پاسخ می‌دهند (Hegdé and Van Essen 2003; Hegdé and Van Essen 2007). در حقیقت بسیاری از نورون‌های V2 به این مجموعه محرک پاسخ می‌دادند و در میان این محرک‌ها، کانتور و زوایا در تحریک نورون‌های V2 نقش بیشتری داشتند. این مشاهدات مهر تأییدی بر این فرضیه بود که نورون‌های V2 در سلسله مراتب مسیر بطنی، خصوصیاتی مانند زوایا و کانتورها را تشخیص می‌دهند.

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

V4-۴-۵-۲ لایه

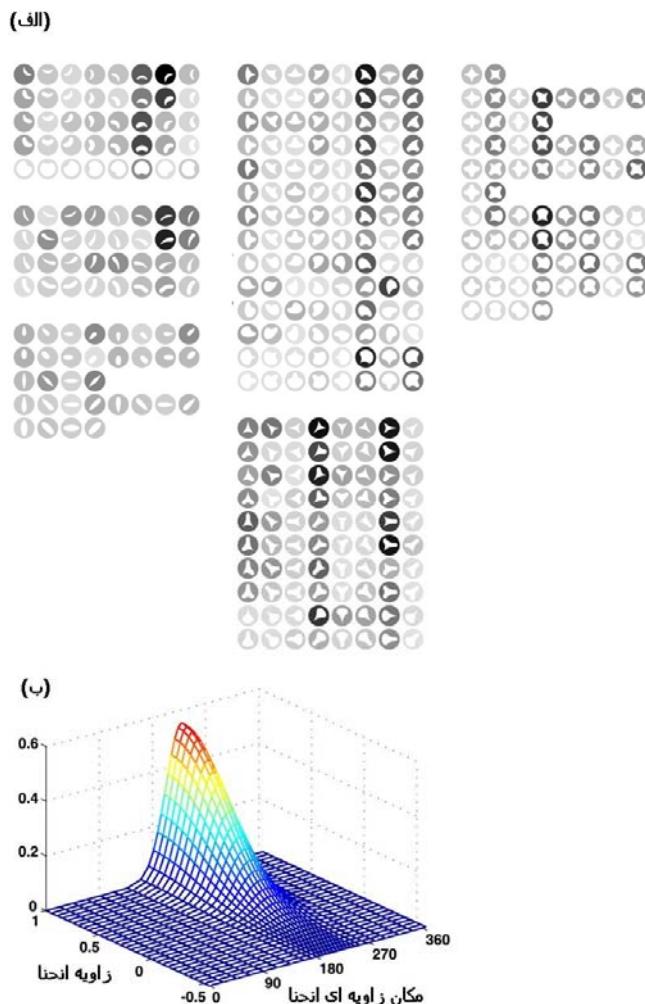
این ناحیه در مجاورت IT قرار دارد و نورون های آن خصوصیات نسبتاً پیچیده ای از تصاویر را تشخیص می دهند. یکی از مهمترین مطالعاتی که در مورد این ناحیه انجام شده است با استفاده از یک مجموعه محرک نشان داده شده در شکل ۲-۳۶ انجام پذیرفته است (Pasupathy and Connor 2001). در این آزمایش، هر محرک شامل یک منحنی بسته 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 احنای کانتور^{۱۳۴} محرک می باشد و دیگری زاویه ای که احنای 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 احنای با زاویه های خاص بود. به عنوان مثال نورونی که در شکل ۲-۳۶ بالا نشان برای نمونه آورده شده به احنای 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، احنای 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 احنای 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 متناسب در یک سلسله مراتب

^{۱۳۴}curvature

سازمان یافته‌اند. این مدل در فصل بعد به تفصیل مورد بررسی قرار می‌گیرد. در این مدل، برای نورون‌های V4 از یک تابع گوسی استفاده شده است که همانند توابع پایه شعاعی (RBF) عمل می‌کند. نتایج آزمایشات، شباهت مدل ذکر شده و نورون‌های V4 را نشان می‌دهند.



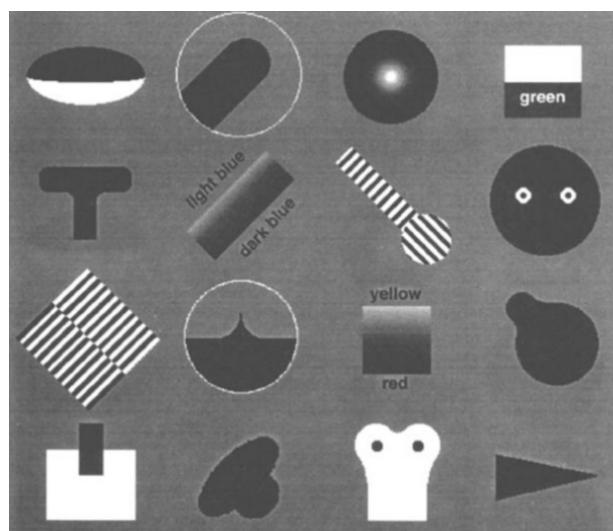
شکل ۳۶-۲: مجموعه محرك استفاده شده در (Pasupathy and Connor 2001) برای آزمایش نورون‌های V4(الف) در این شکل میزان تاریکی پس زمینه نشان دهنده ی شدت پاسخ نورون به آن محرك می باشد. (ب) نموداری که میزان پاسخ نورون شکل الف را بر حسب انحنای زاویه و مکان انحناء نشان می دهد.

۴-۶-۲- لایه IT

ناحیه IT در قطعه مجانبی قشر مغز، آخرین ناحیه‌ای است که داده‌های بینایی را پردازش می‌کند بدون آنکه عوامل سطح بالا در این پردازش تاثیر داشته باشند. داده‌های بینایی پس از این ناحیه به قسمت‌های مختلفی از جمله قشر تداعی کننده در قطعه گیجگاهی و قشر پیش‌پیشانی^{۱۳۵} فرستاده می‌شوند (Miller and Cohen 2001). بنابراین ناحیه IT آخرین

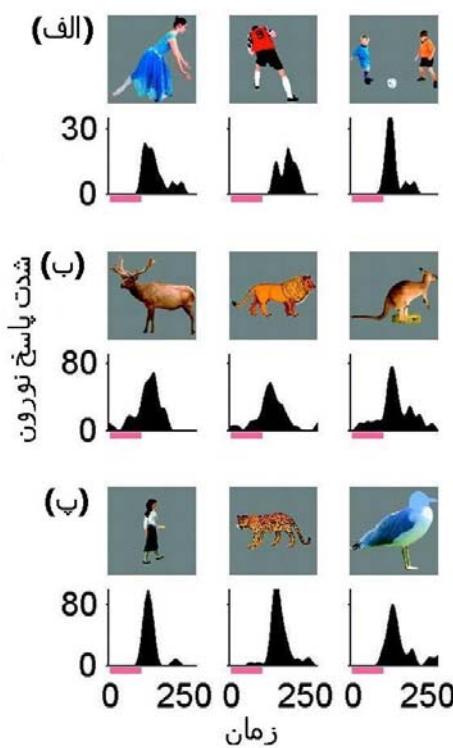
^{۱۳۵}prefrontal cortex

ناحیه‌ای است که داده‌های بینایی را پردازش می‌کند و تنها داده‌های حسی در عملکرد آن موثر هستند. نورون‌های 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 می‌شود. نکته مهم در مطالعه نورون‌های IT استفاده از مجموعه محركی است که بتواند تمام فرضیه‌های احتمالی در مورد خصوصیات این نورون‌ها را شامل شود. در یکی از مطالعات مجموعه بسیار زیادی از اشیا گوناگون برای تحریک نورون‌های IT استفاده شد (Tanaka 1996). سپس هر شکلی که پاسخی قوی از نورون استخراج می‌کرد، در یک فرآیند ساده سازی به اجزای تشکیل دهنده خود تجزیه می‌شد. این ساده سازی تا هنگامی که پاسخ نورون افت قابل ملاحظه‌ای نداشت ادامه می‌یافت. بدین ترتیب، محرك‌هایی برای هر نورون پیدا می‌شد که پاسخ نسبتاً قوی در نورون ایجاد می‌کردند و تا حد امکان ساده بودند. مجموعه ای از این محرك‌ها در شکل ۳۷-۲ ملاحظه می‌شود. این شکل محرك‌هایی را نشان می‌دهد که بیشترین پاسخ را در نورون‌های IT ایجاد کرده‌اند.



شکل ۳۷-۲: مجموعه محرك‌هایی که در نورون‌های IT بیشترین پاسخ را ایجاد کردند (Tanaka 1996).

یکی از خصوصیات جالب در نورون‌های IT، حساسیت آنها به محرك‌های سطح بالا مانند چهره می‌باشد. بدین شکل که نورون‌های مجاور هم به یک شئ سطح بالا پاسخ یکسان می‌دهند. در مطالعه ایجدید مجموعه زیادی از تصاویر از دسته‌های مختلف اشیا به کار برده شدند و پاسخ نورون‌های IT به این محرك‌ها ضبط شد (Kiani, Esteky et al. 2007). نکته‌ای که در این آزمایش مشاهده گردید این بود که نورون‌های IT در دسته‌های مختلفی قرار می‌گیرند و هر دسته به نمونه‌های تصویر از یک دسته خاص از اشیا پاسخ بیشینه می‌دهند. می‌توان گفت که نورون‌های IT اشکال مختلف سطح بالا را به صورت کدهای دسته جمعی از پاسخ نورون‌ها نشان می‌دهند. نمونه‌ای از سه نورون حساس به دسته‌های مختلف در شکل ۳۸-۲ نشان داده شده‌اند.



شکل ۳۸-۲: 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.

۲-۵-۵- ویژگیهای آماری تصاویر طبیعی

۱-۵-۲- مقدمه

تصاویر طبیعی تصادفی نیستند، بلکه قاعده مندیهای 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 مندیهایی 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 مندیهای 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 تصاویری که بر روی شبکیه می افتدند 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 مندیهای 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 تمیز نباشد. اگرچه همچنان راه درازی برای نیل به این مقصود باقی است، ولی دانش انسان درباره تصاویر طبیعی رو به فزونی است.

یک انگیزه مهم برای چنین تحقیقاتی فهم بهتر اینست که سیستم بینایی ما چگونه اطلاعات تصویری را کدگزاری می کند. بر طبق نظریه اطلاعات کدهایی که از نظر آماری کارتر هستند، همانهایی هستند که قاعده مندیهای 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 مندیهای 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 مندیهای 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 Simoncelli and Olshausen 2001; Laughlin and Sejnowski 2003; Olshausen and Field 2005) یافت). ابتدا به قاعده مندیهای آماری تک پیکسلها (مقادیر شدت روشناهی) در تصاویر طبیعی پرداخته می شود. سپس به قاعده مندیهای آماری نواحی تصویر توجه می گردد. این نواحی در اکثر تحلیلهای، 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 بدليل 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).

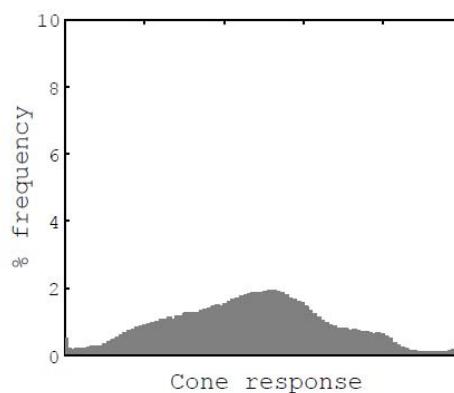
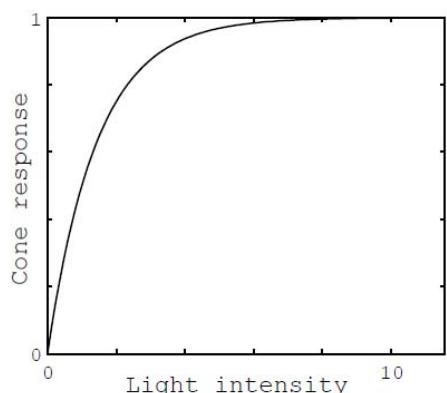
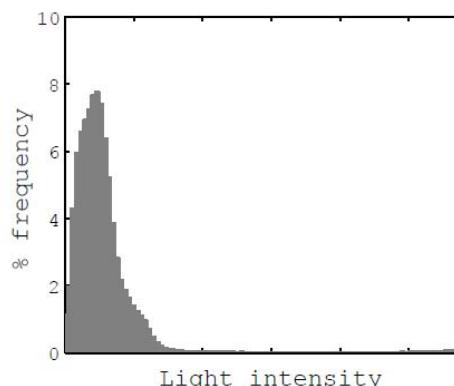
هنگامی که درباره مدلهای نظری صحبت می شود باید دو نوع آنها را از یکدیگر تمیز داد. روش نخست معکوس مدل مولّد داده هاست که در آن مدل احتمالاتی تولید داده ها دانسته فرض می شود و وظیفه سیستم بینایی استنتاج پارامترها و عوامل نهفته ایست که داده ها را تولید میکند (Dayan, Hinton et al. 1995; Olshausen 1996). مزیت 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 مولّد نیست، بلکه برای اراضی قیودی برای بازنمایی مانند ناهمبسته سازی یا انتقال بیشینه اطلاعات طراحی شده اند (Doi and Lewicki 2005).

¹³⁶image 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 ۳۹-۲ بالا-راست). این تحلیل بر مبنای تصاویری Ruderman, Cronin et al. 1998; Van Hateren and Van der Schaaf 1998; Doi, Inui et al. 2003 بدست آمده است که مقدار پیکسل آنها به طور خطی با شدت روشنایی مرتبط است (پایین راست) و شبیه شکل ۳۹-۲ بالا-چپ است. توجه شود که چنین داده هایی به عنوان ورودی برای چشم مناسب هستند در حالی که تصاویری که با دوربینهای معمولی فیلم CCD گرفته می شوند شامل تبدیلات غیر خطی است و برای تحلیل مناسب نیستند.



شکل ۳۹-۲: شدت روشنایی خطی صحنه های 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 ۴۳-۲ 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:

$$x_{nl} = 1 - \exp(-kx_1) \quad (42-2)$$

که در آن x_{nl} شدت روشنایی خطی، x_1 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 ۴۲-۲ در شکل ۳۹-۲ پایین-چپ آمده است که پارامتری آن برای تصویر شکل ۳۹-۲ بالا-چپ تطبیق یافته است. هیستوگرام پاسخ غیرخطی گیرنده های مخروطی حاصل که در شکل ۲-۲-۳۹ پایین-راست آمده است بیانگر اینست که مقادیر پیکسل نسبت به مقادیر خام شدت روشنایی توزیع بسیار یکنواخت تری دارند (تصویر شکل ۳۹-۲ بالا-چپ وقتی با پاسخ گیرنده های مخروطی تبدیل می شوند دارای کنتراست بهبود یافته ای می شوند، شکل ۲-۴۳). شبی مدل تجربی که از معادله ۴۲-۲ 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 ۳۹-۲ متفاوت باشد. یک مدل که خطای تخمین ورودی را با داشتن نویز ذاتی (که به آن نویز کاتال 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).

۲-۵-۳-۲- 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 کدگزاری 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). بنابراین توکیه موثری برای اینکه کدام قید برای سیستم بینایی اولیه مناسب

¹³⁷Luminance

است آورده نشده است، ولی بهر حال 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.

۳-۵-۲- 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.

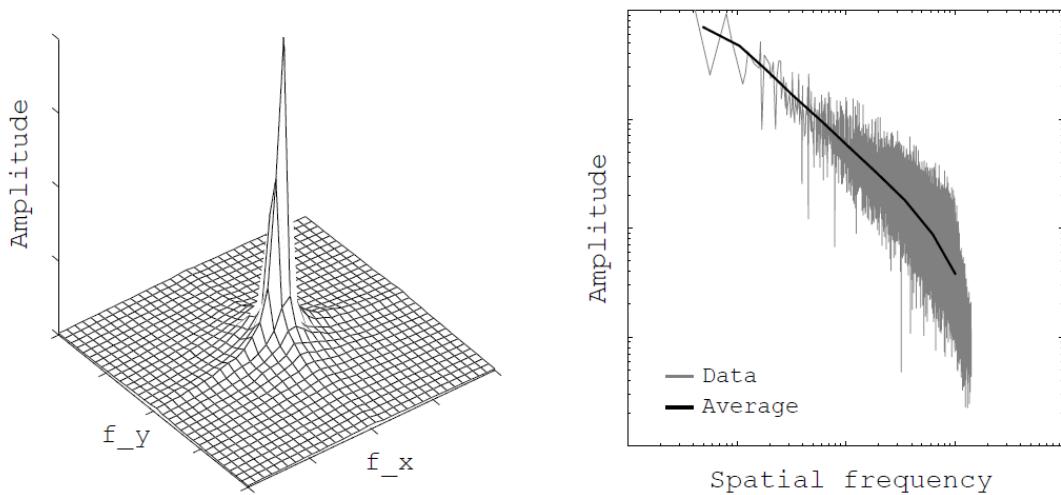
تعريف آمارگان مرتبه λ_m عبارتست از $\left\langle \prod_k^i x_k \right\rangle$ که در آن $\langle \cdot \rangle$ عملگر میانگین گیری کلی روی نمونه ها را نشان می دهد و x_k مقدار پیکسل در مختصات k است. با فرض نامتغیری به ازای جابجایی در تصاویر طبیعی، میانگین کلی می تواند با میانگین گیری روی اندیس مختصات با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

۳-۵-۲-۱- قاعده مندیهای 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 f/I است که f فرکانس مکانی است (Field 1987). در شکل ۴۰-۲ مثالی از طیف بزرگی تصاویر طبیعی آورده شده است. می توان مشاهده کرد که بزرگی در طول راستهای افقی و عمودی اندکی قویتر است که این به دلیل غالب بودن ساختارهای افقی و عمودی در صحنه های طبیعی است (مثلًا درختان و افقها). طیف بزرگی I/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 I/f ، تصاویر رسته های گوناگون طبیعی دارای سبک خود در طیف بزرگی هستند (Torralba and Oliva 2003). به عنوان مثال تصاویر جنگلها دارای مؤلفه های مکانی- فرکانسی بالای زیادی هستند ولی تصاویر ساحل دریا بزرگی بیشتری را در فرکانسها 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 راستاهای یکریخت تر و دارای مؤلفه های مکانی فرکانسی بالایی است، که با راستای مقید نشده دید نسبت به اشیا و جزئیات ریز در صحنه قابل توجیه است. همینطور اگر تصویر از دور گرفته شده باشد، بزرگی در راستای محورهای عمودی و افقی قویتر است و در فرکانسها مکانی پایینتر مرکز است که با غالب بودن

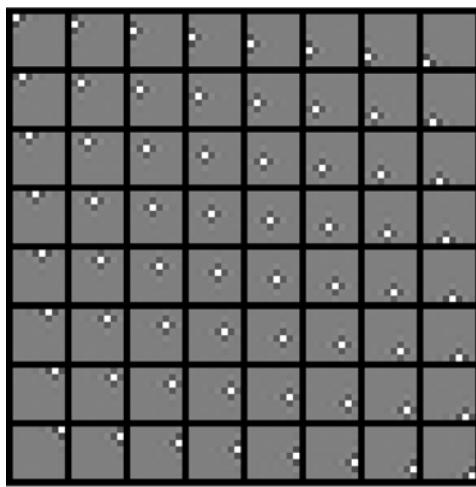
ساخترهای افقی و عمودی در یک نمای دور و افزایش ساختارهای بزرگ و هموار مانند زمینهای، جنگلها و آسمان قابل توجیه است.



شکل ۲-۴۰: طیف بزرگی تصاویر طبیعی متناسب با f_1/f_2 است. داده های تصویری تحلیل شده در شکل ۳۹-۲ نشان داده شده است. (چپ) بزرگی فرکانس مکانی دو بعدی. (راست) بزرگی فرکانس مکانی که از تجمعیت بزرگی دو بعدی روی راستاها بدست آمده است. منحنی خاکستری داده های تمام فرکانسهاست و نسخه هموار شده آن (میانگین آن) با منحنی سیاه مشخص شده است.

۲-۳-۵-۲- ارتباط با خواص پاسخ

هدف سیستم بینایی اولیه، حذف افروزنگی 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 کاراتر باشد (Barlow 1961; Atick 1992). کاهش افروزنگی در قالب آمارگان مرتبه دوم به معنی تبدیلی است که متغیر را دارای واریانس واحد و ناهمبسته با بقیه کند. چنین تبدیلی وايتینینگ نامیده می شود و با یک مجموعه از فیلترهای خطی قابل پیاده سازی است. پس از اعمال وايتینینگ به تکه تصاویر طبیعی یک ساختار فیلتری حول مرکز بدست می دهد که توافق زیادی به میدان گیرندهای سلولهای عقدہ ای شبکیه (RGCs) و LGN دارد (Atick and Redlich 1990; Atick and Redlich 1993; Bell and Sejnowski 1997). تعداد نامحدودی فیلتر وايتینینگ وجود دارد. اگر W_1 یک ماتریس فیلتر وايتینینگ باشد، ماتریس کواریانس خروجی آن باید برابر ماتریس همانی باشد، $W_{1x} \cdot W_{1x}^T = I$ ، که به طور ضمنی اشاره دارد که $W_1 \cdot W_1^T = I$. 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 وايتینینگ است زیر رابطه $W_2 \cdot W_2^T = I$ را ارضا می کند. ماتریس وايتینینگی 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).



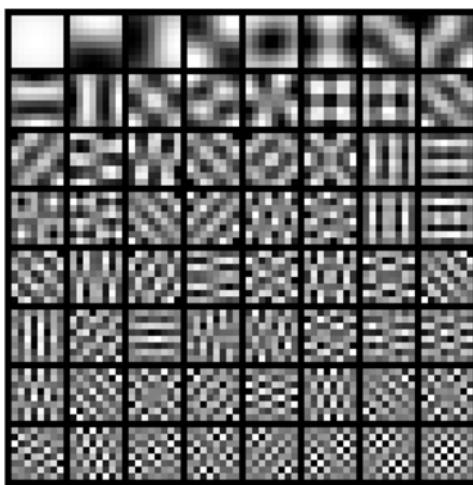
شکل ۲-۴: فیلترهای واپتینینگ برای تصاویر طبیعی سازماندهی میدان گیرندگی حول مرکز را از خود نشان می‌دهد. هر تصویر کوچک با یک فیلتر واپتینینگ برای تکه تصویرهای 8×8 متناظر است. در اینجا تمام مجموعه ۶۴ تایی آن به نمایش در آمده است..

۳-۵-۲- 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 واپتینینگ را نیز در خود محسوب می‌کند. با فرض اینکه واریانس نویز متناسب با میانگین سطح روشنایی باشد، مدل به تطبیق چشگیری با میدانهای گیرندگی RCG‌ها دست می‌یابد که شکل خود را بر اساس میانگین سطح روشنایی سازگار می‌کنند.

واپتینینگ عملیاتی است که قاعده مندیهای آماری مرتبه دوم را از بین می‌برد. به عنوان یک ایده می‌توان از این خاصیت استفاده کرد و به جای حذف این قاعده مندیها از آنها استفاده کرد. اطلاعات رسته که در طیف بزرگی قرار دارد – چنانکه در بالا به آن اشاره شد – نیز می‌تواند 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 می‌دهد و بازنمایی جدید کمترین خطای مربعی را در بازسازی دارد. شکل ۲-۴ یک مجموعه از فیلترهای PCA را نشان می‌دهد که الگوهای فرکانسی را بدون هیچ دخالت محلی نشان می‌دهد و بنابراین هیچ همبستگی آشکاری به سازماندهی میدانهای گیرندگی سیستم بینایی اولیه ندارد. با این حال PCA ممکن است قربتی ساختاری با سیستم بینایی انسان داشته باشد. واریانس داده در زیر فضایی محدود مت Shankل از تعداد اندکی از مؤلفه‌های اصلی متمرکز شده است (مثلاً حدود ۲۰ تا ۳۰٪ مؤلفه‌های اصلی ۹۰٪ واریانس داده‌های تکه‌های تصاویر طبیعی را توصیف می‌کند). 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.



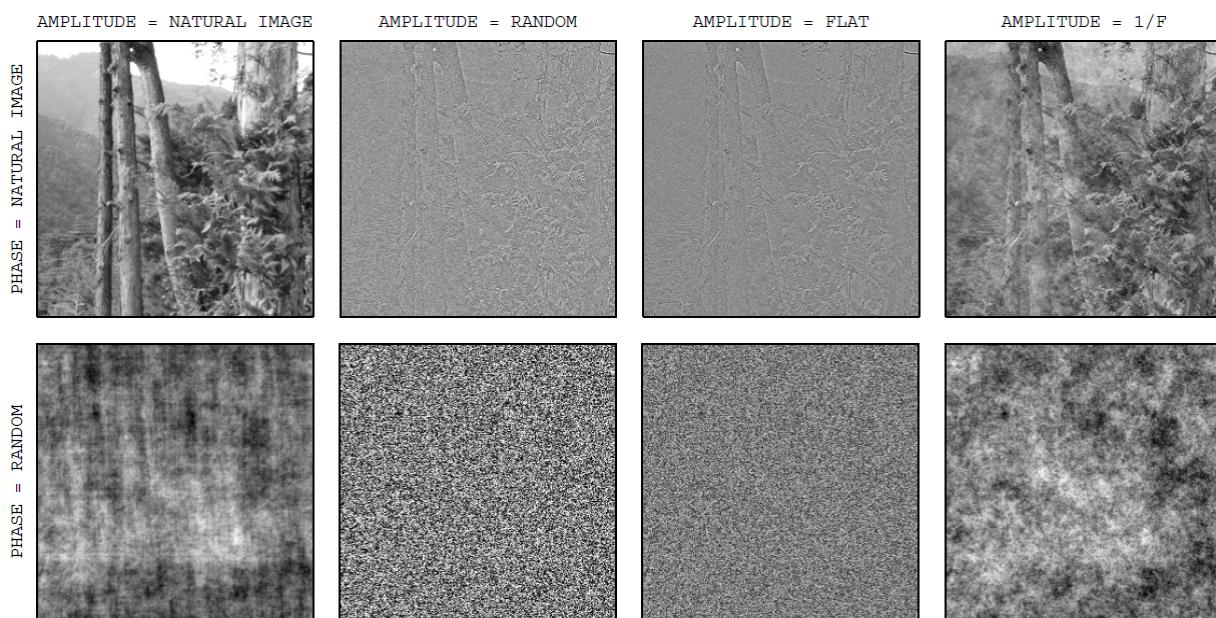
شکل ۲-۴۲: فیلترهای PCA تصاویر طبیعی محلی نیستند. این یک مجموعه از مؤلفه های اصلی برای تکه تصویرهای 8×8 پیکسل تصاویر طبیعی است که به ترتیب نزولی بزرگی از بالا به پایین و چپ به راست مرتب شده است..

۴-۵-۲- آمارگان مرتبه بالاتر

۱-۴-۵-۲- قاعده مندیهای آماری

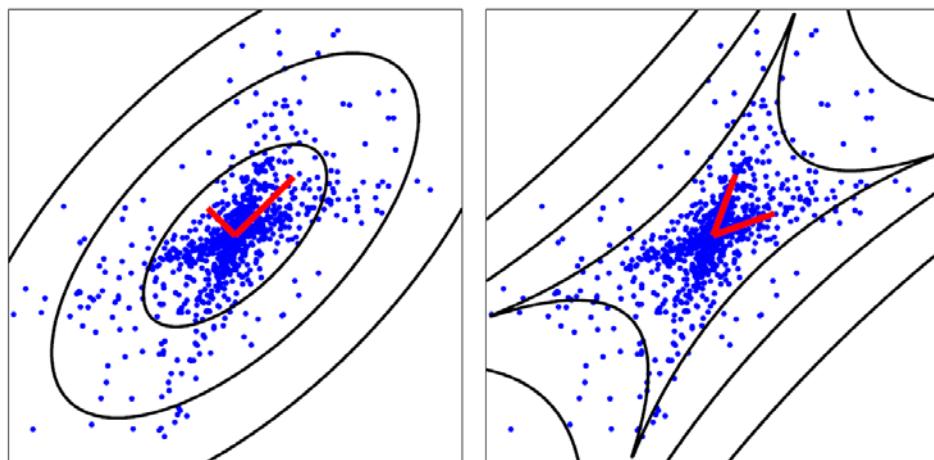
آمارگان مراتب بالاتر تصاویر طبیعی در طیف فاز تبدیل فوریه آنها مقیم هستند که مکمل طیف بزرگی این تصاویر (=آمارگان مرتبه دوم) هستند. آنگونه که اشاره شد، طیف فاز دارای اطلاعاتی ایست که از لحاظ مفهومی مرتبط ترند زیرا اطلاعاتی در مورد لبه ها بدست می دهنند: در مکان لبه ها فازهای فرکانس‌های مختلف همراستا می شوند. این مسئله را می توان با بهم ریختن فاز یک تصویر طبیعی در عین حفظ طیف بزرگی آن مشاهده کرد (شکل ۴۳-۲). 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 بر جسته بین آسمان و کوهها را مثال زد). بهر حال موقعیت آنها (=فاز) هیچ ربطی به تصویر اصلی ندارد و به دلیل فقدان همراستایی فاز، الگوی لبه مانندی مشاهده نمی شود. به طریق مشابه تصاویری با بزرگی های ثابت یکنواخت، ثابت $f/1$ 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 $f/1$ است تصویری بسیار نزدیک به تصویر اصلی می دهد (شکل ۴۳-۲ بالا-راست) که نشان می دهد که اطلاعات طیف بزرگی نقش کمنگتری در ادراک انسان دارد.

طیف بزرگی ماتریس کواریانس داده ها را شکل می دهد و بنابراین می تواند تمام قاعده مندیهای 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 1994; Field 1987). در شکل ۴۴-۲ 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 ۴۴-۲ چپ). برای مقایسه توزیع واقعی داده ها در شکل ۴۴-۲ راست ترسیم شده است که همان چیزی است که باید از داده ها تخمین زده شود.



شکل ۴۳-۲: طیف فاز نسبت به طیف بزرگی در ادراک ما سهم بیشتری دارد. هر تصویر از تلفیق طیفهای فاز و بزرگی ساخته می‌شود. از هر تصویر طبیعی یا تصادفی طیف فاز گرفته شده و در سطر پایین ترسیم می‌شود. از هر تصویر طبیعی یا تصادفی طیف بزرگی با بزرگی $1/f$ یا یکنواخت می‌گیریم. تناظر تصویر {فاز، بزرگی} = {تصویر طبیعی، تصویر طبیعی} مربوط به تصویر طبیعی اصلی است. تصویر {تصادفی، تصادفی} نیز تصویر اصلی تصادفی است و تصویر $\{1/f, 1/f\}$ تصویر همبسته تصادفی است که طیف بزرگی آن $1/f$ است. توجه شود که تصویر {طبیعی، یکنواخت} نسخه واپتینیگ 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 می‌شود: $3 - \tilde{K}(x) = \langle x^4 \rangle / \sigma_x^2$ که کورتوسیس آماره‌ای از مرتبه چهارم است. توزیعی پراکنده تر از گوسی است که کورتوسیس آن مثبت باشد ($0 < \tilde{K}$). جهت‌هایی در فضای تصویر و چود دارد که نگاشت خطی نقاط نمونه روی آنها توزیعی پراکنده دارد (Field 1987; Field 1994). توجه شود که چنین توزیع غیرگوسی اگر داده‌ها گوسی چند متغیره باشند امکان‌پذیر نیست زیرا هر نگاشت خطی از یک گوسی چندمتغیره گوسی است. جالب توجه است که نگاشت خطی روی Daugman 1985; Jones and Palmer 1987) - یک توزیع پراکنده بدست می‌دهد که این ایده را القا می‌کند که سیستم بینایی از قاعده مندیهای مراتب بالاتر استفاده می‌کند.



شکل ۴-۲: آمارگان مرتبه دوم برای مشخص کردن داده های غیرگوسی کافی نیستند. نقطه های آبی در هر دو قسمت از یک توزیع نامتغایر 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 یاتابع پایه توزیع غیرگوسی آنها را نشان می دهد..

۴-۵-۲-۴- ارتباط با خواص پاسخ

مدلهای نظری کدگزاری 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 راستادار میدان گیرندگی سلولهای ساده وجود دارد (Olshausen 1996; Bell and Sejnowski 1997; Hyvärinen, Hoyer et al. 2001

x توسط تجمعی تابعهای پایه \vec{a}_i که با متغیرهای مخفی s_i وزن دهی شده اند بوجود آمده اند:

$$\vec{x} = \sum_{i=1}^m s_i \vec{a}_i = A \vec{s} \quad (43-2)$$

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

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

$$L = \langle p(\vec{x} | A) \rangle \quad (44-2)$$

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

$$p(\vec{x} | A) = \int p(\vec{x} | A, \vec{s}) p(\vec{s}) d\vec{s} \quad (45-2)$$

این مدلها فرض می کند که اولاً متغیرهای مخفی مستقل آماری هستند و ثانیاً هر متغیر مخفی از یک توزیع پراکنده مانند لاپلاسین 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 نورده صعودی برای یافتن بیشینه معادله ۴۶-۲ محاسبه می‌شود. فرض این مدل اینست که هدف محاسباتی سیستم بینایی استنتاج محتملترین بازنمایی محرک (که احتمالاً دارای نویز است) است که به صورت زیر فرموله می‌شود:

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

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

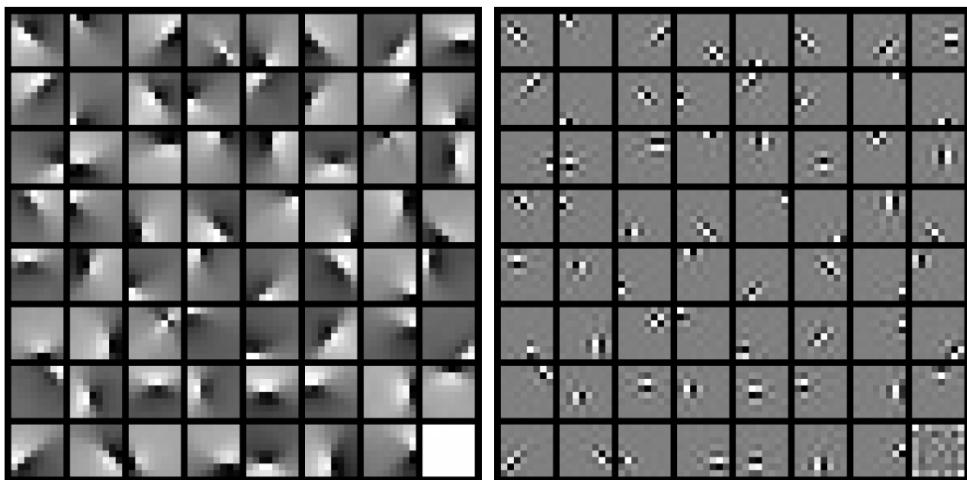
$$\hat{s} = A^{-1}\vec{x} = W\vec{x} \quad (48-2)$$

که \hat{s} تبدیل شده خطی x است (پس $P(s|x,A)=\delta(s-A^{-1}x)$). از سوی دیگر در کدگزاری پراکنده، نگاشت به این حالت خاص محدود نیستو باید معادله ۴۷-۲ که به شکل زیر نوشته می‌شود، 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 سطح خطای بازسازی را تعریف می‌کند. این بدین معنی است که \hat{s} با کمینه کردن توأم خطای مربعی بازسازی (جمله اول) و درستنمایی منفیتابع احتمال اولیه لابلسانی (جمله دوم) بدست آید.

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



شکل ۲-۴۵: روش ICA فیلترهای محلی و راستدار مشابه میدان گیرنده‌ی سلولهای ساده V1 تولید می‌کند. 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.

۳-۴-۵-۲- ملاحظات

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

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

توجه شود که توزیعی که در شکل ۲-۴۴ راست دیده می شود یک نمونه دوبعدی از چنین توزیعهایی است. مدلهایی مانند ICA و کدگزاری پراکنده، کدهای کارا را تشکیل می دهند ولی برخلاف واپتینینگ (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/RGHa 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 مرتبه بالاتر حتماً باقی می ماند که در بخش بعد بررسی می شود.

۲-۵-۵-۲- توسعه های بیشتر

۱-۵-۵-۲- مدلهای غیرخطی

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 تمايل 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 موجک نیز دیده می شود، که فرم دیگری از کدگزاری 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 مدلها غیرخطی برخی از خصوصیتهای پیچیده پاسخهای نورونهای V1 مانند کنترل بهره و نامتغیری به فاز را شرح می دهد که معتبر بودن فرضیه کدگزاری 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 توبوگرافیک نورونهای ICA به خدمت گرفته شود(Hyvärinen, Hoyer et al. 2001).

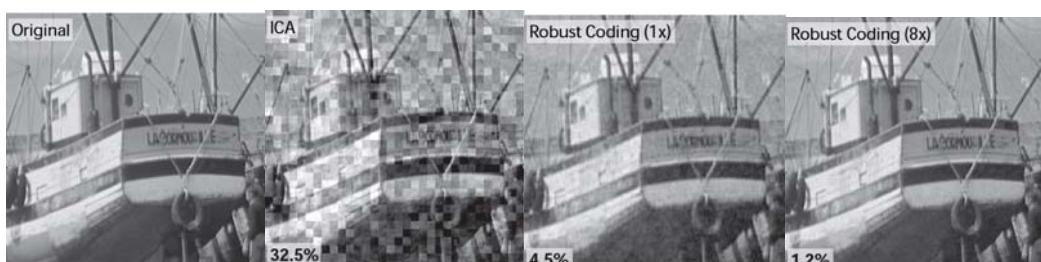
۲-۵-۲- دخیل کردن محدودیتهای بیولوژیکی

کدگزاری 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 ماهیت چند مقیاسه بازنمایی بینایی اولیه است، مخصوصاً تنظیم فرکانس مکانی در سلولهای ساده V1 که در سه اکتاو در نقاط دور از مرکز شبکیه رخ می دهد(De Valois, Albrecht et al. 1982; Anderson and DeAngelis 2004) در مقابل، کدگزاری پراکنده و بازنمایی های تک مقیاسی را بدست می هند که در شکل ۴۵-۲ مشهود است. البته کدگزاری 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 کدگزاری پراکنده و ICA اگر وجود داشته باشد هم تعداد آنها ناچیز است. بالأخص بازنمایی های چندمقیاسه همچنان در RCG و Derrington and Lennie 1984; Croner and Kaplan 1995; Lee, Mumford et al. 2000 یافت می شود(LGN) در حالی که فیلترهای واپتینینگ تنظیم تک مقیاسه بدست می دهد(شکل ۴۱-۲). سایر شکستهای پیش بینی به کمک کدگزاری 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 کدگزاری بهینه و ICA است(Ringach 2002).

یک روش برای حل این مشکل پالایش مدل با تعییه 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 تخمين زده می شود که به کمی یک بیت در هر اسپایک نورون باشد (Fairhall, Lewen et al. 2001;).

(Dhingra and Smith 2004). در چنین شرایط نویزی ای، استفاده از کدگزاری کارا به تنها می تواند مضر باشد زیرا افروزنگی 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 کدگزاری شود. از لحاظ فیزیولوژیکی اعتقاد رایج بر اینست که یک گلوگاه روی اعضای بینایی وجود دارد و این ورودی باید با ناهمبسته سازی توسط PCAنمایش داده شود. به حال این امر تنها برای نواحی جانبی شبکه صدق است. در ناحیه فوویا، تعداد سلولها همواره بدون هیچ گلوگاهی افزایش می یابد که احتمال وجود بازنمایی افزونه را به ذهن مبتادر می کند.

اخیراً مدلهای 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 Doi and Lewicki 2005; Doi and Lewicki 2006) در این مدل که به آن کدگزاری مقاوم^{۱۳۸} گفته می شود، هدف محاسباتی بجای استنتاج متغیرهای نهفته، کمینه کردن خطای بازسازی است و بنابراین تعداد واحدهای کدگزاری دلخواه است و اجرای ندارد که با تعداد واحدهای مفروض مدل برابر باشد. این امر امکان مدل کردن هر ناحیه شبکیه 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) سوچیج 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 و موجک بهبود چشمگیری داشته است که این مسئله بخصوص برای بازنمایی فراکامل $\times 8$ دیده می شود (شکل ۲-۴۶). از لحاظ نظری می توان نشان داد که خطای بازسازی با افزایش تعداد واحدهای کدگزاری کننده می تواند تا حد دلخواه کوچک شود (Doi, Balcan et al. 2006).

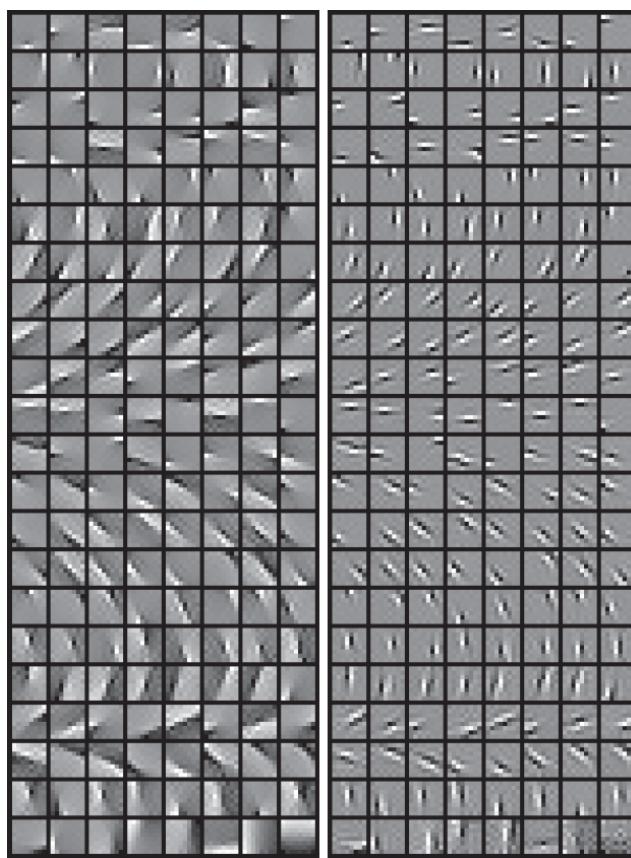


شکل ۲-۴۶: نتایج بازسازی با واحدهای دارای ظرفیت محدود. در این آزمایش نویز کانال اضافه می شود به طوری که ظرفیت اطلاعاتی کانال به یک بیت به ازای هر ضریب برسد. در هر تصویر بازسازی شده، درصد خطای بازسازی نیز ذکر شده است.

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 پراکننده باشد و برای بهبود کارایی کد، قاعده مندی مراتب بالاتر را نیز فاش کنند. 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 کدگزاری کننده است و نیز تنظیم فرکانس مکانی بیشتری مشابه سلولهای ساده V1 از خود نشان می‌دهد. این مدل یک چارچوب ابتدایی را برای بکارگیری افرونگی در بازنمایی با داشتن تعداد دلخواهی واحد کدکننده مهیا کرده است.



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

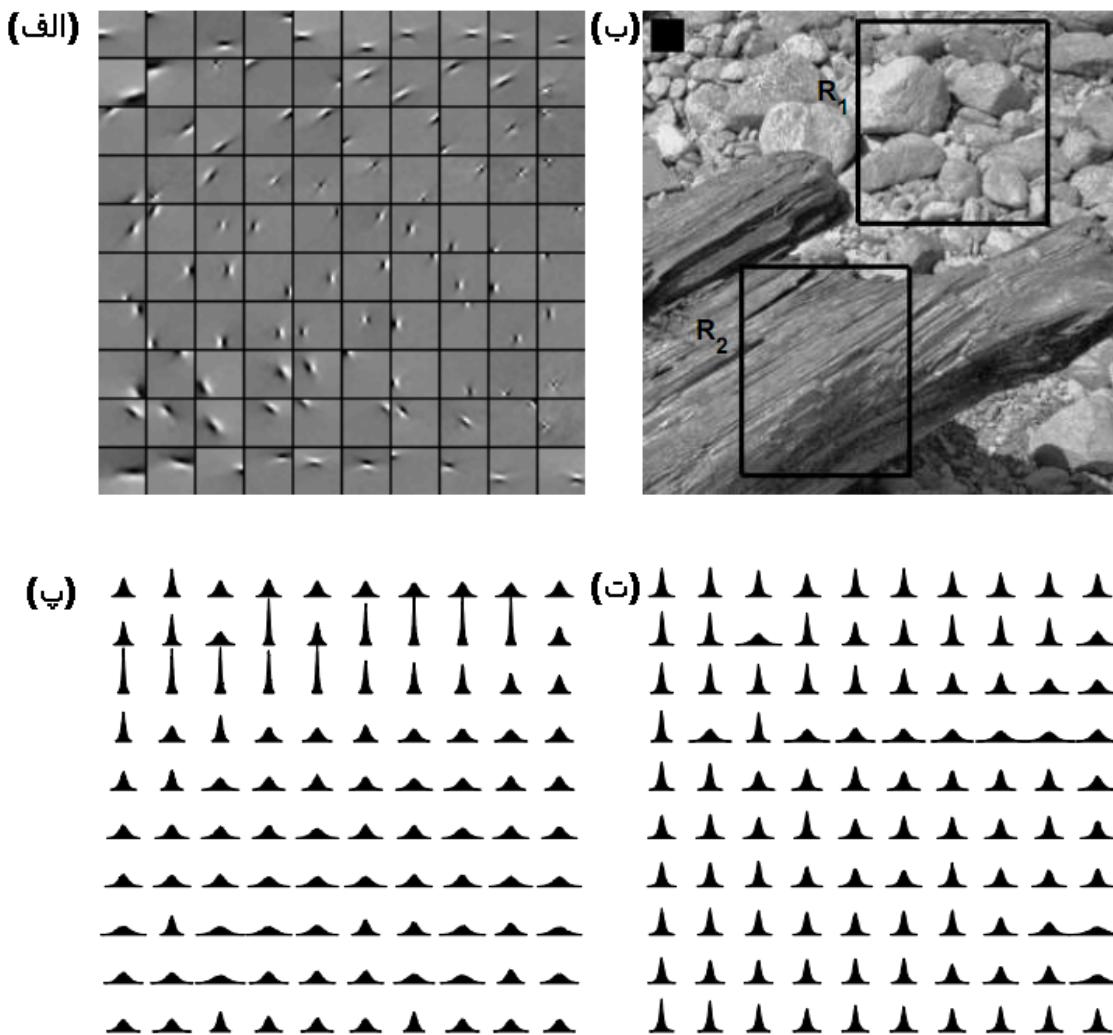
۱۳۹ ۵-۵-۳- وابستگیها بر اساس زمینه

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

¹³⁹ context

پایان نامه کارشناسی ارشد کورش مشگی - پاییز ۱۳۸۹
دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر

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

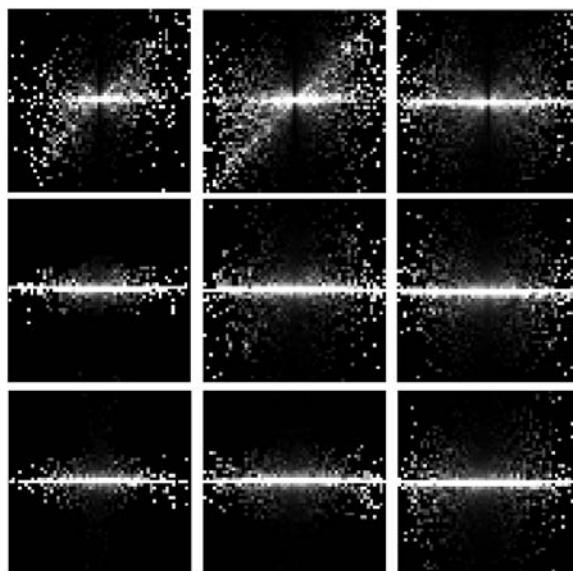


شکل ۴-۲: وابستگی پاسخ مؤلفه‌های مستقل به زمینه در تصویر. (بالا چپ) مجموعه‌ای از مؤلفه‌های مستقل استخراج شده از تصویر. (بالا راست) پاسخ مؤلفه‌های مستقل به تکه‌هایی از تصویر از نواحی R_1 و R_2 مورد تحلیل قرار گرفتند. (پایین چپ) پاسخ مؤلفه‌های مستقل به ناحیه R_1 (پایین راست) پاسخ مؤلفه‌های مستقل به ناحیه R_2 پاسخ مؤلفه‌های مستقل با جهت مشابه در زمینه‌های ذکر شده با توزیع احتمال پیش‌فرض تفاوت دارد.

۴-۵-۲- وابستگیها بر اساس کلاس اشیا

افزونگی موجود در تصاویر به کلاس تصاویر مورد بررسی بستگی دارد و بسیار پیچیده تراز وابستگی در همسایگی محدود می‌باشد (Malmir and Shiry 2009). شکل ۴-۲ توزیع احتمال توأم پاسخ فیلترهای خطی گبر را نشان می‌دهد. این توزیع بر روی تصاویری از کلاس‌های مختلف اشیا جمع آوری شده است. سه جفت فیلتر با فواصل نسبی متفاوت از یکدیگر در سه

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



شکل ۴۹-۲: توزیع احتمال توام پاسخ فیلترهای گیر به مجموعه تصاویر از کلاس‌های مختلف. هر سطر پاسخ یک زوج فیلتر با فاصله نسبی متفاوت را ارائه می‌کند. نمودارهای هر ستون با استفاده از مجموعه تصاویر مختلفی محاسبه شده اند. سه مجموعه از تصاویر برای محاسبه توزیع توام پاسخ ها استفاده شده‌اند: تصاویر صورت، تصاویر هوایپیما 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 سمت راست نیز با استفاده از تصاویر هوایپیما محاسبه شده اند. زوج فیلتر سطر اول دارای فاصله ۷ واحد، سطر دوم فاصله ۱۵ و سطر سوم فاصله ۳۰ واحد می‌باشند. تصاویر از مجموعه ۱۰۱ شی استخراج شده اند

اولین نکته‌ای که در شکل ۴۹-۲ بارز است تفاوت توزیع توام برای دو مجموعه فیلتر در ستون‌های مختلف است. برای زوج فیلتر با کمترین فاصله نسبی، وابستگی واریانس روی هر سه مجموعه تصاویر مشاهده می‌شود. در سطر دوم، زوج فیلتر برای مجموعه تصاویر صورت و هوایپیما وابستگی را نشان می‌دهند. 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 هوایپیما مشخص می‌باشد. برای دو مجموعه دیگر تصاویر، این فیلترها به نظر مستقل می‌آیند (Malmir and Shiry 2009).

مدل‌های قبلی که بر اساس وابستگی واریانس بین پاسخ فیلترهای خطی گسترش یافته بودند، نمی‌توانند این پدیده را توجیه کنند. در این مدل‌ها، دو فیلتر یا به هم وابسته هستند یا از هم مستقل اند و این موضوع به مجموعه تصاویر مورد استفاده ارتباطی ندارد. در شکل ۴۹-۲ مشاهده می‌شود 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. می‌توان گفت که اندازه‌ی همسایگی که در آن وابستگی واریانس وجود دارد با مجموعه تصاویر استفاده شده ارتباط مستقیم دارد. مدل‌های قبلی از مجموعه محدودی از تصاویر استفاده می‌کردند 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 میابد و وابستگی واریانس از میان می رود. این حالتی است که در ستون دوم و سوم سطر اول شکل ۴۹-۲ اتفاق افتاده است(مالمیر ۱۳۸۸).

هنگامی که تصاویر از کلاس یکسانی مثل تصاویر چهره انسان باشند، دامنه مکانی وابستگی واریانس گسترش میابد. 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 ۴۹-۲ 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 یافته(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 ۴۹-۲ 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 مندیهای تصاویر طبیعی، اهداف محاسباتی مانند کدگزاری 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 کدگزاری 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 اپتیک و چینش حسگرهای چشم، حرکات چشم و ...) است. همچنین پیشنهاد شده است که

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

فصل سوم

[بررسی کارهای پیشین]

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

۱-۳-۱-۳- مدل‌های نظری عملکرد قشر بینایی مغز

۱-۱-۳- مقدمه

دو روش کلی برای مطالعه آمارگان سیگنالهای طبیعی وجود دارد. مدل‌های پایین به بالا به مطالعه قاعده مندیهای آماری تجربی بسته آمده از نگاشتهای کم بعد خطی و غیرخطی می‌پردازند. این روشها نورونهای قشر مخ را به عنوان عواملی برای انتخاب و دستکاری نگاشتها می‌نگرند که هدف آنها بهینه سازی معیارهای احتمالاتی و نظریه محاسباتی است (Shannon and Weaver 1949) که از این قبیل معیارها می‌توان به پراکندگی (Field 1987) و کدگزاری کارا (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 می‌کنند و نورونهای قشر مخ را به صورت یک دستگاه مختصات می‌بینند که پارامترهای رویه آماری سیگنال را تعیین می‌کنند.

۲-۱-۳- کدگزاری بهینه

چه اصول محاسباتی زیربنای پردازش در قشر بینایی مغز را تشکیل می‌دهد؟ یکی از خصوصیات سیگنال‌های طبیعی که همواره مورد توجه بوده است افزونگی موجود در این سیگنال‌ها می‌باشد. در دهه‌های ۵۰ و ۶۰ میلادی و پس از معرفی نظریه اطلاعات توسط شanon، کارکرد این نظریه در سیگنال‌های 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 ورودی‌های حسی به مغز انسان دارای حجم بسیار زیادی می‌باشند. 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 می‌باشند و حجم اطلاعاتی که می‌توانند در هر لحظه پردازش کنند از حجم داده‌های حسی که وارد مغز می‌شوند بسیار کمتر است. مطابق نظریه کد کردن بهینه، وظیفه قشرهای حسی، پردازش داده‌های ورودی است به نحوی که حاصل این پردازش، افزونگی نداشته باشد و حجم آن متناسب با اطلاعاتی باشد که منتقل می‌کند. به عنوان مثال، برای یک موجود زنده، داده‌هایی که از طریق بینایی به مغز آن وارد می‌شوند نسبت به اطلاعات مهم در تصاویر مانند شکارچی، طعمه یا جفت، دارای حجم بیشتری می‌باشند. قسمت‌های تصمیم گیری و حرکتی مغز نمی‌توانند 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 قسمت‌های تصمیم گیری در مغز فرستاده شوند. این فرضیه که به نظریه کدگزاری بهینه معروف است، یک اصل کلی است که می‌تواند 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 می‌دهد 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 کدگزاری 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 کارتر 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 مرزهای است. ظاهر و بهم بندی این عناصر قاعده مندیهای آماری را در ورودیهای حسی پیدید 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 کند بنابراین یافتن یک توصیف فشرده چندان آسان نیست. با اینحال کدگزاری کارا یک چارچوب آماری را پیدید می آورد که طبعاً با بسیاری از جنبه های حسگری در دنیای غیرقطعی مانند وجود نویز کنار می آید و کمیتهايی 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 1987; Daugman 1989). 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 کدگزاری 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 انجمانی (Zetsche 1990; Field 1994) کارتر از روشهای سنتی هستند و مطالعات اولیه با توجه به اینکه خروجی فیلترها نماینده میدان گیرنده سلولهای ساده هستند، این بازنمایی را به قشر خاکستری مخ نسبت می دانند (Field 1987; Daugman 1989; Zetsche and Barth 1990; Field 1994).

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 کانوالو فیلترهای خطی با تصاویر پرداخته اند که این فیلترهای خطی معمولاً به گونه‌ای انتخاب می‌شوند که از لحاظ کیفی با میدانهای گیرندگی نورونهای شبکیه یا قشر مخ مطابقت داشته باشد. میدان گیرندگی بخش اولیه قشر بینایی Hubel and Wiesel (1962) و از این‌رو فیلترهای خطی که بتوانند این پارامترها را در خود بگنجانند و کنترل کنند مورد توجه قرار گرفته اند. از جمله این فیلترها می‌توان به فیلتر گابور اشاره کرد. فیلتر گابور از ضرب یک تابع سینوسی درتابع نمایی به دست می‌آیند و نشان داده شده که مدل مناسبی از گزینندگی نورون‌های V1 می‌باشد (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 می‌کنند.

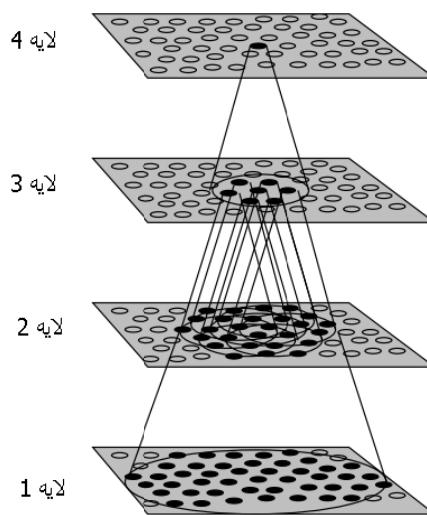
این میدانهای گیرندگی قرابت بسیاری به تجزیه چندمقیاسه موجک^{۱۴۶} دارند که در جامعه بینایی رایانه‌ای به توفیق زیادی دست یافته است. برای تصاویر طبیعی، مشاهدات تجربی فعالیتهای یک فیلتر خطی توزیعی با کوترسیس بالا (بنابراین پراکنده) را از خود نشان می‌دهد (Field, 1987). گروه‌هایی از فیلترهای خطی (که با توجه به پارامترهایی نظیر راستا، فرکانس، Wegmann and Zetsche 1990; Simoncelli 1997؛ Fazl, 1997؛ Buccigrossi and Simoncelli 1999؛ Schwartz and Simoncelli 2001؛ Brehm and Stammler 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 می‌شود.

VisNet - ۱-۳-۱-۳

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

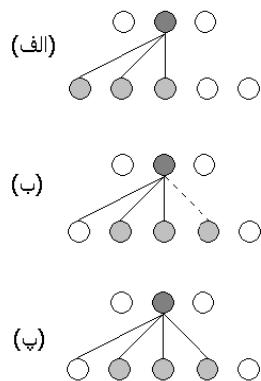
¹⁴⁶ multi-scale wavelet decomposition

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 نامتفاوت اشیا توسعه یافته اند.

در این مدل برای یادگیری پاسخ نامتفاوت اشیا، قانون یادگیری تبدیلات پیوسته به کار برد شده است (Rolls and Stringer 2006). این قانون برگرفته از رفتار نورون های قشر بینایی مغز در مواجهه با محرك های مختلف می باشد. شکل ۲-۳ این قانون را توضیح می دهد. ایده اولیه بر اساس این مشاهده است که پاسخ سطح بالای یک نورون به محرك تا مدت کوتاهی پس از مشاهده محرك باقی می ماند (Földiák 1991). مطابق این قانون، یادگیری پاسخ نامتفاوت 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 ۲-۳: 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 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 ^{۱۴۸} که به قانون ردیابی موسوم است، برای به روزرسانی اتصالات بین نورون‌ها گسترش یافته است. این قانون گسترش یافته‌ی قانون هبی می‌باشد:

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

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

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

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

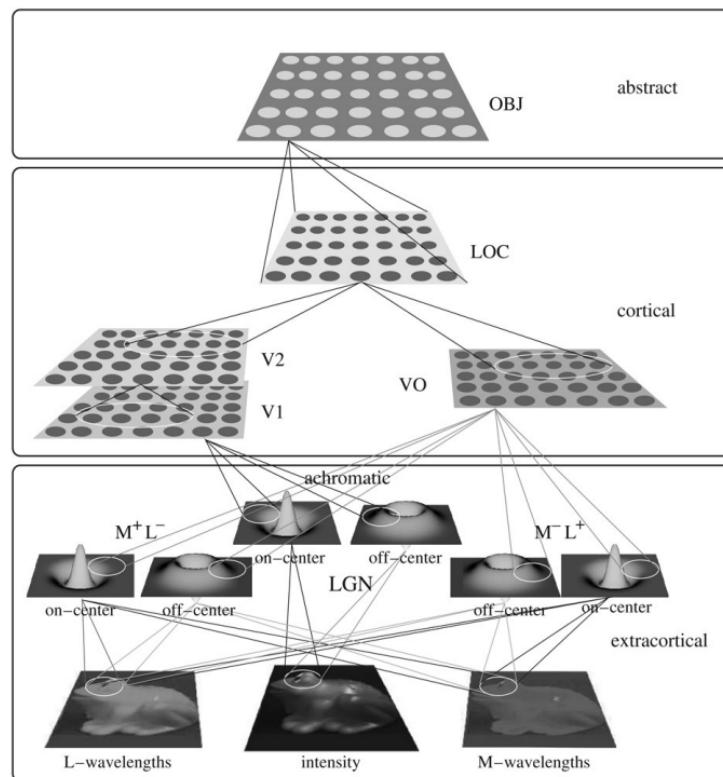
۳-۲-۳-۱- مدل نقشه‌های قشری مغز^{۱۴۹}

این مدل که به نام RF-Lissom نیز معروف است، بر اساس ایده شبکه‌های خودسازمانده گسترش یافته است (Plebe and Domenella 2007). مدل اصلی برای شبیه‌سازی ترکیب چینش نورون‌های 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 نورون‌های نواحی مختلف در قشر بینایی را شبیه‌سازی کند. خصوصیت اصلی این مدل در کلی بودن مکانیزم‌های ارتباطی ارائه شده بین نورون‌ها می‌باشد. در هر لایه، نورون‌ها از نورون‌های لایه قبل و نورون‌های مجاور خود ورودی دریافت می‌کنند. در تمام این اتصالات، قوانین خودسازمانده برای یادگیری مقادیر وزن‌ها 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 نقشه‌های قشری مشابه با الگوی موجود در قشر بینایی مغز می‌باشد. به عنوان مثال، لایه V1 در این مدل از مجموعه‌ای از نورون‌های LGN ورودی دریافت می‌کند. نشان داده شده که با قانون یادگیری خودسازمانده، در هر لایه V1 ترکیبی مشابه با ناحیه V1 قشر بینایی مغز پدیدار می‌شود. در هر لایه V2، نورون‌ها در یک روال یادگیری، حرکت‌هایی به شکل زاویه را یاد گرفتند (Plebe 2007). همچنین در لایه‌ی آخر مدل، نورون‌ها همانند نورون‌های قشر IT به صورت خوش‌بندی شده به اشیا مختلف حساسیت پیدا کرده اند.

¹⁴⁸ Trace rule

¹⁴⁹ Cortical maps



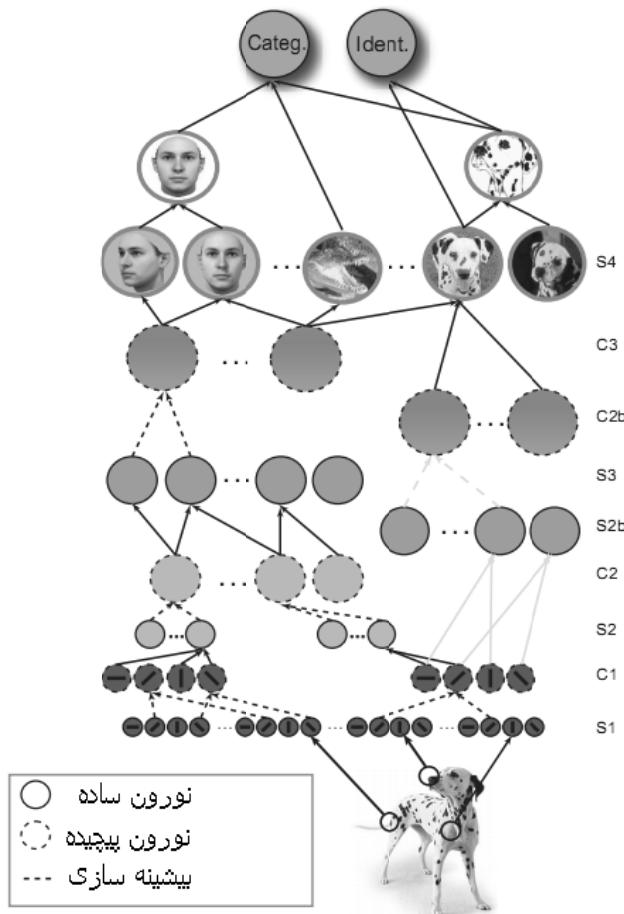
شکل ۳-۳: مدل نقشه های قشری مغز. اتصالات بین لایه ای و درون لایه ای در این مدل با استفاده از قوانین خودسازمانده یاد گرفته می شوند.

۳-۳-۱-۳ HMax مدل

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 است 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 عملگر 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 ماکریتم 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.

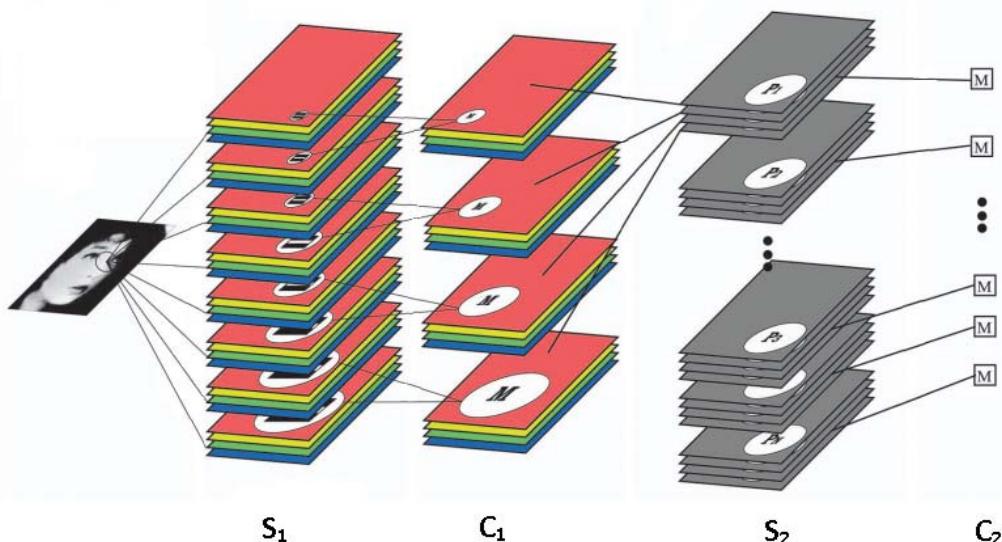
^{۱۵۰} Max operation

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 نامتفاوت نشان می دهند. لایه های ساده و پیچیده در این مدل به ترتیب با C_i و S_i مشخص شده اند که i شماره لایه را تعیین می کند.



شکل ۳-۳: 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 نامتفاوت 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 دارد. 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 نامتفاوت به تبدیلات جهت و اندازه در اشیا 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 C_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.

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

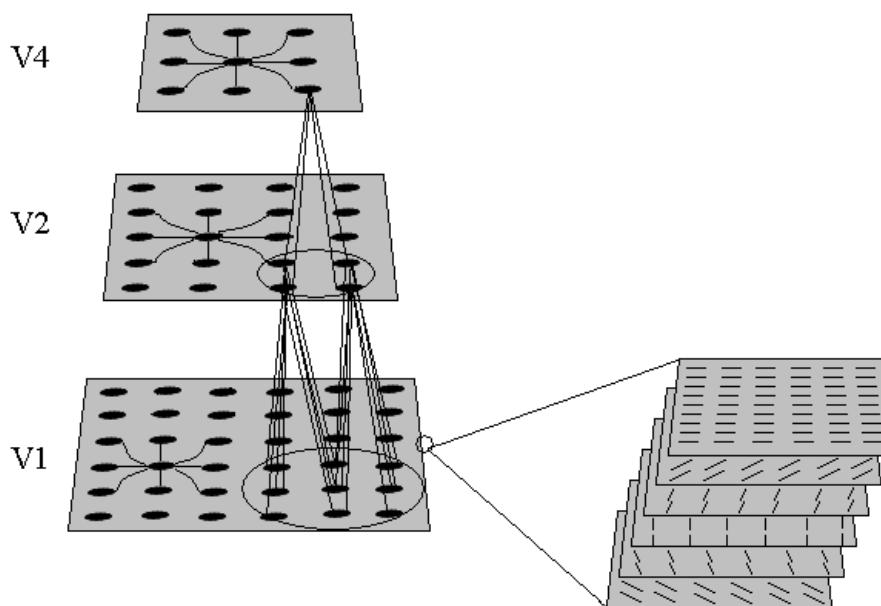
این مدل (Malmir and Shiry 2009) یک شبکه عصبی چند لایه است که به صورت سلسله مراتبی سازماندهی شده است. شکل ۳-۶ این مدل را نمایش می دهد. مدل ارائه شده دارای سه لایه متناظر با لایه های V1، V2 و V4 می باشد. در هر لایه از این مدل، نورونها از لایه پایینتر و از نورونهای مجاور خود ورودی دریافت می کنند. وزن های اتصال به لایه پایینتر، گزینندگی نورون را مشخص می کنند و اتصالات افقی در رفع افروزنگی 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 گیرندهای نورونهای هر قشر حدوداً ۲/۵ 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 افروزنگی 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 می دارد که هرگاه محرک نورون حسی از نظر مکانی گسترش باید به نحوی که نورون های مجاور آن را نیز فعال کند، پاسخ نورون مرکزی کاهش می باید. این کاهش در پاسخ با تقسیم کردن مدل شده است، یعنی پاسخ نورون بر مقدار پاسخ نورون های اطراف تقسیم می شود. مدل ارائه شده با رفتار نرمال سازی مشاهده شده در قشرهای سطح بالای مغز Carandini, Heeger et al. 1997; Reynolds, Chelazzi et al. 1999; Zoccolan, Cox et al. 2005 مطابقت دارد)

(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 می باشد. در هر لایه اتصالات افقی و اتصالات ورودی از لایه پایین تر ورودی نورون را تعیین می کنند.

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

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

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

شود که برای اینکار از روش شب نورده نزولی استفاده می شود و مقادیر وزنهای جانبی، واریانس نورونها و وزنهای بین لایه ای بروزرسانی می شود.

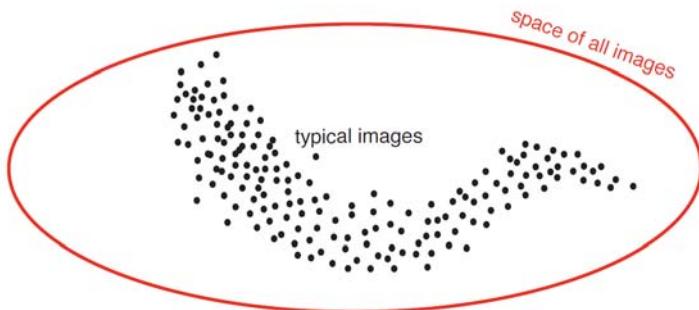
این مدل پدیده هایی مانند ساخنار توبوگرافیک لایه V1، جمع کردن کانتور در اثر اتصالات جانبی V1، گزینندگی کانتور و بافت در لایه V2، گزینندگی زاویه در V2، کد کردن کانتور در لایه V4، و تشخیص اشیا در زوایای دید و نورپردازیهای گوناگون را توجیه می کند.

۳-۱-۴-۳- 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 نورونهای 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 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 تخمين 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 بیتهاي موجود برای کدگزاری ناحیه هایی از فضای تصویر که محتملترند دیده می شود. و مسائلی مانند بهبود تفکیک پذیری تصویر با 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 گذیری، 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 ۷-۳) 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 ۷-۳: مجموعه عام تمام تصاویر طبیعی. می‌توان 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 خصوصیاتی آماری که آنقدر پایه ای و ابتدایی باشند که لزوماً برای تمام صحنه های طبیعی به کار بیایند. دوم مدلها پارامتری که در آن مجموعه ای از آبرپارامترها^{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 می‌توان یک مدل احتمالاتی برای تصاویر تولید کرد و آزمود؟ روش‌های بسیاری برای اینکار بوجود آمده است ولی در اینجا برای سهولت فهم یک روش تجربی برای تصاویر گستته سازی شده (پیکسلی) شرح داده می‌شود. دوربینهای دیجیتالی که امروزه یافته می‌شوند عکس‌هایی می‌گیرند که معمولاً حاوی میلیونها پیکسل هستند. استفاده از هیستوگرام تصاویر کاری بیهوده است زیرا اطلاعات بسیار کمی درباره محتوای تصویر دربر دارد. همانطور که در شکل ۸-۳ دیده می‌شود تصاویری با محتوای یکسان ممکن است دارای هیستوگرامهایی متفاوت باشند و بر عکس تصاویری دارای محتویاتی کاملاً متفاوت ممکن است هیستوگرامهایی یکسانی داشته باشند.

اولین ایده خامی که به ذهن متبادر می‌شود بررسی مجموعه بزرگی از تصاویر به منظور بدست آوردن توزیع آنهاست. ولی در نگاه مجدد به این راه حل می‌توان به بی‌فایده 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 می‌یابد، که این پدیده به طلسنم 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 می‌شود(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 می‌شود مستقل از توزیع پیکسلهای خارج از همسایگی است. فرض دوم برای ساده سازی از تحمیل تقارنها با نامتغیرهای خاصی به ساختار احتمال بدست می‌آید. رایجترین مورد این نامتغیریها، نامتغیری به ازای جابجایی است (که همگنی^{۱۵۶} یا پایداری اکید حستی^{۱۵۷} نیز نامیده می‌شود)

¹⁵² soft

¹⁵³ hyper parameters

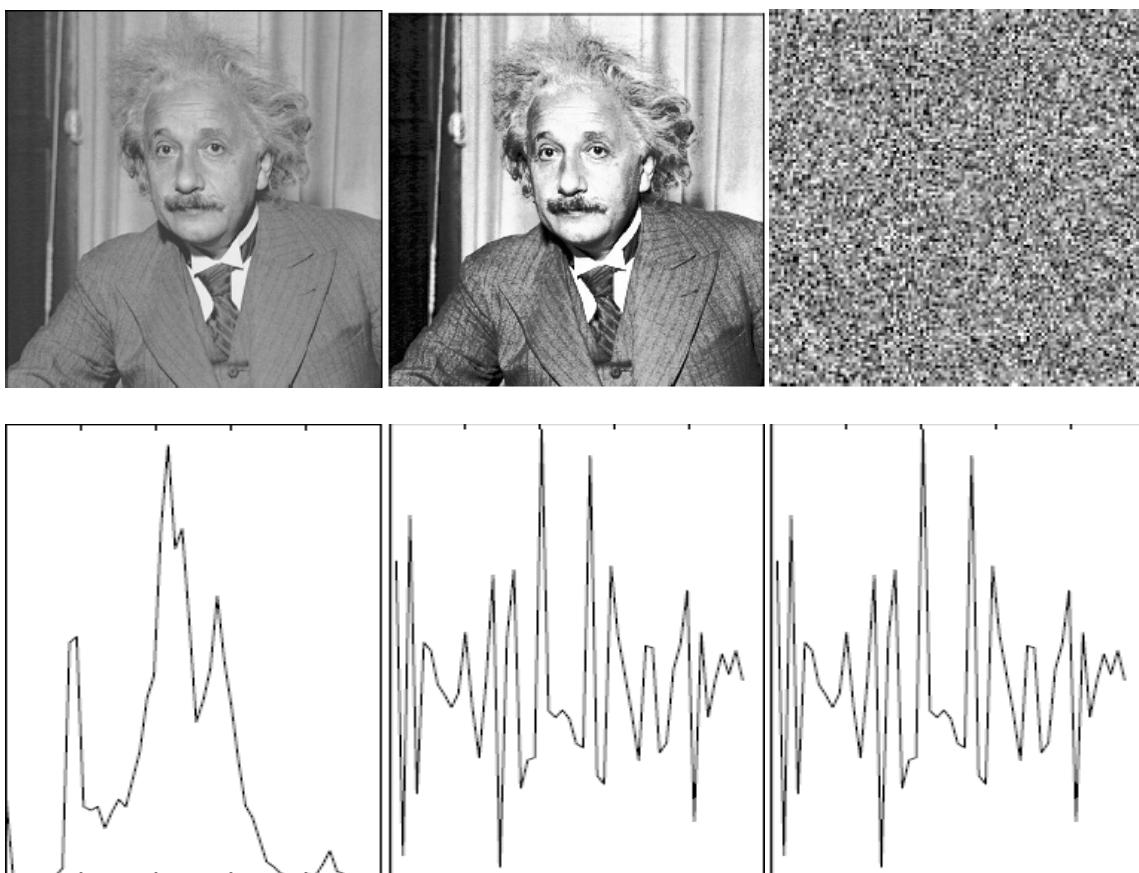
¹⁵⁴ curse of dimensionality

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 جایجایی افقی یا عمودی دوربین، تنها شدت روشنایی پیکسلها در آرایه پیکسلها شیفت پیدا می کند. باید توجه داشت که نامتغیری به ازای جایجایی در مرزها خوش تعریف نیست و معمولاً در پردازش تصویر این مناطق توجه ویژه ای را می طلبند.



شکل ۳: هیستوگرام تصاویر، اطلاعات بسیار کمی درباره محتواه تصویر دربر دارد.

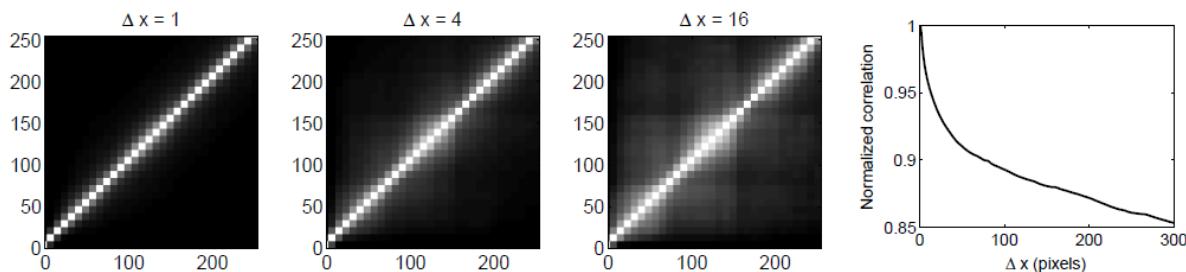
فرض رایج دیگری از این دست، نامتغیری با مقیاس است: تغییر اندازه تصویر ساختار احتمالات را تغییری نمی دهد. این فرض نیز بطور غیر رسمی اینگونه توجیه می شود که تنظیم فاصله کانونی دوربین (زوم) و یا لنز آن (جدا از اعوجاجات پرسپکتیوی) اندازه تصویر را تخمین می زند. همانند نامتغیری به ازای جایجایی، نامتغیری به 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

داده ها نشان داده می شود ارائه می شود ولی آزمون حقیقی زمانی صورت می گیرد که مدل برای یکی از کاربردهای چالش برانگیز پردازش تصویر مانند فشرده سازی یا رفع نویز استفاده شود.

برای حفظ تمرکز بحث تنها به تصاویر عکسبرداری شده، گستته و خاکستری فام پرداخته می شود. بسیاری از اصول به سادگی قابل تعمیم به عکسهای رنگی (Buchsbaum and Gottschalk 1983; Ruderman, Cronin et al. 1998) و یا دنباله ای از تصاویر موقتی (Dong and Atick 1995) یا تصاویر تخصصی تر مانند پرتره ها، منظره ها، و بافتها نیز هستند. علاوه بر آنها روشهای کلی می توانند به تصاویر غیر مرئی مانند تصاویر پزشکی، تصاویر فروسرخ، رادار و سایر انواع تصاویر بعد، و تصاویر نجومی اعمال شوند.

مدل گوسی: مدل کلاسیک آمارگان تصویر توسط مهندسان تلویزیون در دهه ۱۹۵۰ تشکیل شده است (برای مرور رک. 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 گیریهای نظری آنچه در شکل ۹-۳(چپ) مشاهده می شود تایید می شود. هر نمودار پراکندگی مقدار یک جفت پیکسل با تفاوت مکانی داده شده را بدست می دهد. 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 می شود هستند.



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

آنچه در این نمودارها بخوبی خودنمایی می کند 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 می کند. این رفتار در شکل ۹-۳(راست) خلاصه شده است که خودهمبستگی (همبستگی پیکسلها به عنوان تابعی از فاصله آنها) را نشان می دهد. آمارگان همبستگی نشان داده شده در شکل ۹-۳ قید محکمی بر روی ساختار تصاویر قرار می دهد ولی یک مدل احتمالاتی کامل را فراهم نمی آورد؛ 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

طبيعي، انتخاب توزيعي با بيشترین آنتروبي با حفظ محدوديت کواريانس است(Jaynes 78). حل کردن مسئله برای اين توزيع نسبتاً راحت است و به يك توزيع گوسی چند متغيره 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 پيكسلهای تصوير (که برای سهولت با ميانگن صفر در نظر گرفته شده اند) و $C_x = E\vec{x}\vec{x}^T$ ماتريس کواريانس است. توزيعهای گوسی در مختصرترین حالت خود با يك دوران به دستگاه مختصاتی بيان می شوند که ماتريس کواريانس آنها قطری باشد. با استفاده از جبر خطی استاندارد اين هدف تامين می شود:

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

که در آن E ماتريس متعامد حاوی بردارهای يكّه C_x است به طوري که:

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

که در آن D يك ماتريس قطری است که حاوی مقادير يكّه متناظر است. هنگامی که توزيع احتمالاتی \vec{x} پايدار باشد (با فرض حل کردن تناوبی مشكل مرزاها) ماتريس کواريانس C_x ، حلقوی خواهد بود. در اين حالت خاص می دانيم که تبديل فوريه يك تبديل قطری ساز برای ماتريس E خواهد بود و حصول رابطه معرفی شده در معادله ۷-۳ را تضمین می کند. برای اينکه مدل تصوير گوسی كامل شود، تنها باید وروديهای ماتريس قطری D را بدانيم که با مؤلفه های فرکانسي تبديل فوريه متناظر است.

دو راه برای حصول يك جواب وجود دارد. نخست با نادیده گرفتن اخطاری که زودتر داده شد، می توان فرض کرد که آمارگان تصوير به ازاي مقیاس نامتغیرند. مخصوصاً باید فرض کرد که خواص آمارگان درجه دوم (کواريانس) تصوير نسبت به تغيير اندازه تصوير نامتغیر است. می توان نامتغیر بودن به ازاي مقیاس را در دامنه فوريه چنین نشان داد:

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

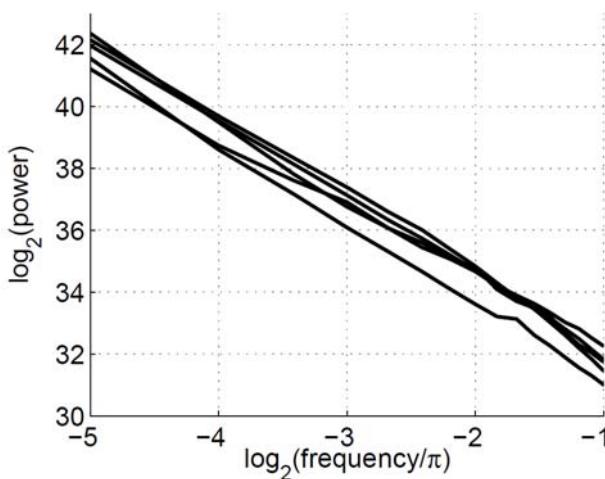
بدين معني که تغيير اندازه محور فرکانس شكل تابع را تغيير نمی دهد و تنها طيف را در يك ثابت ضرب می کند. تنها تابعی که يكساني را تامين می کند قوانين تواني¹⁵⁸ است:

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

که توان آن نرخ نزول طيف را کنترل می کند. بنابراین فرضهای دوگانه تامتصيری به ازاي جابجایی و مقیاس ساختار کواريانس تصاویر را به مدلی تنها با يك پارامتر مقيد می کند. از سوی ديگر شكل طيف توان را می توان بطور تجربی بدست آورد

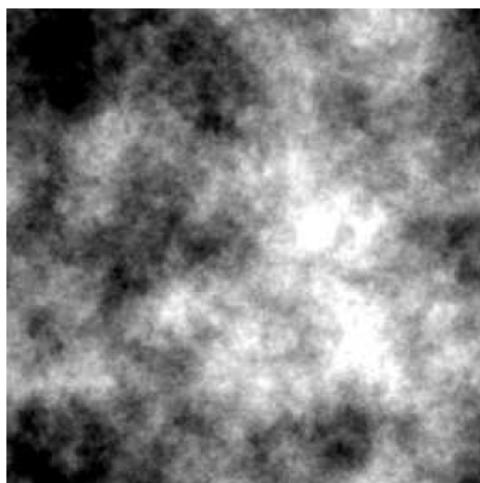
¹⁵⁸ power laws

Deriugin 1957; Field 1987; Tolhurst, Tadmor et al. 1992; Ruderman and Bialek 1994; Van der Schaaf and Van Hateren 1996 تخمین زده می شود و بر درستی خصوصیت نامتغیری به ازای مقیاس در آمارگان مرتبه دوم صحه 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 ۱۰-۳ 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 ملزمات اولیه پردازش تصویر است.



شکل ۱۰-۳: تخمین طیف توانی برای پنج تصویر نمونه (که در شکل ۹-۳ 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 ۱۱-۳ 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 ۱۱-۳: تصویری نمونه که بطور تصادفی از مدل طیفی گوسی انتخاب شده است.

به عنوان یک آزمون مستقیم می‌توان 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 ۱۱-۳ دیده می‌شود. در واقع این آزمایشها که هر بار تصاویری ابر مانند تولید می‌کند نشانگر اینست که محدودیت کواریانس برای اخذ ساختار غنی تری از ویژگیهایی که در اکثر تصاویر طبیعی یافت می‌شوند ناکافی هستند.

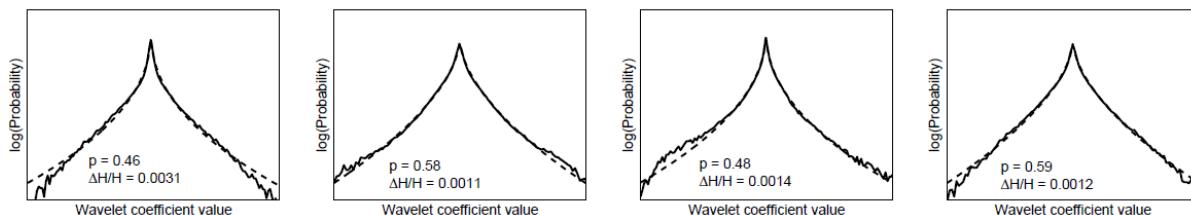
مدلهای توابع پایه خطی: برای دهه‌ها، ناقص بودن مدل گوسی مسئله‌ای آشکار و آزاردهنده بود ولی هرگونه پیشرفتی در زمینه مقید کردن فازهای فوریه به نظر بسیار دشوار می‌آمد. رابطه بین مؤلفه‌های فاز به آسانی اندازه گیری نمی‌شود که این امر از سوی بخاطر دشواری کار کردن با آمارگان توأم متغیرهای حلقوی است و از سوی دیگر به دلیل اینست که وابستگی بین فاز فرکانس‌های متفاوت به نظر نمی‌آید که بخوبی توسط مدلی که خود از فرکانس محلی استفاده می‌کند مدل شود(Simoncelli 2005).

در دهه ۱۹۸۰ یک جهش عظیم با اقبال نویسندها به شرح ویژگیهای غیرگوسی در تصاویر رخ داد. یک مدل آماری گوسی چند متغیره دارای این خاصیت است که تمام توزیعهای شرطی و حاشیه‌ای آن نیز گوسی است. ولی این نویسندها 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; Zetzsche 1990). توزیع فعالسازی تک تک فیلترهای خطی هنگامی که با تصویر کانوالو می‌شود دارای کورتوسیس بالایی است؛ بدین معنی که پاسخ فیلتر اغلب صفر است ولی ندرتاً فیلتر به ساختارهای خاصی در تصویر پاسخی قوی می‌دهد(Field 1987). این توزیعهای حاشیه‌ای در حوالی صفر قله بسیار تیزتری دارند و دنباله^{۱۵۹} های آنها گسترده‌تر از توزیع گوسی با واریانس مشابه است. مثلاً شکل ۱۲-۳ هیستوگرام سه تصویر را نشان می‌دهد که با تابع گابور^{۱۶۰} (یک شبکه سینوسی محدود شده با توزیع گوسی) فیلتر شده‌اند. توجیه منطقی این رفتار اینست که تصاویر معمولاً دارای نواحی همواری هستند که توسط برخی از ویژگیهای محلی مانند خطها، لبه‌ها و یا گوشها از هم جدا شده‌اند. 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 غیرگویی تر هستند می انجامند.



شکل ۱۲-۳: هیستوگرام لگاریتمی یک زیرباند موجک برای چهار تصویر نمونه (که در شکل ۹-۳ توصیف شده بود) برای هر هیستوگرام، دنباله‌ها 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 $\text{۱} - \frac{\text{۱}}{\text{۱} + \text{exp}(\frac{x - \mu}{\sigma})}$ توصیف می‌شوند. نوشته‌های زیر هر هیستوگرام مقدار بیشترین درستنمایی p را نشان می‌دهد که برای تطبیق نمودار استفاده می‌شود و نیز آنتروپی مربوط به مدل و هیستوگرام (افتراق کولبک-لیبلر آنها) را به عنوان کسری از کل آنتروپی، هیستوگرام نشان می‌دهد.

در دهه ۱۹۹۰ محققان بیش از پیش به تبدیلات خطی برای توصیف تصاویر طبیعی روی آوردن. 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 گرینندگی نورون های V1 نداشتند.

یکی از مشکلات تحلیل مؤلفه های پایه در این است که مؤلفه های بسته 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 وابستگی های غیر خطی را تائید 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. لب 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 انحنای نیز اشاره کرد. آماره های مرتبه دوم طیف درشتی^{۱۶۱} یک تصویر -ونه فاز آنرا- در بر دارد. درشتی یک آماره مرتبه دوم است و طیف درشتی یک تصویر لزوماً یک سری همبستگی با موجهات 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 مرتب بالاتر، به طیف فاز مریوطند(Bartlett 1998).

¹⁶¹ amplitude spectrum

پایان نامه کارشناسی ارشد کورش مشگی - پاییز ۱۳۸۹ دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر

مکانیسمهای کدگزاری 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 نویسندها روش‌هایی برای بهینه سازی پایه ای از فیلترها جهت بیشینه سازی غیرگوسی بودن پاسخها تولید کردند(Bell and Sejnowski 1997; Olshausen and Field 1997). تحلیل مؤلفه های مستقل و کدگزاری پراکنده از جمله روش‌هایی بود که برای جبران نقص های تحلیل مؤلفه‌های اصلی گسترش یافت. این روشها معمولاً با بهینه سازی آمارگان مراتب بالا مانند کورتوسیس (گشتاور چهارم تقسیم بر مربع واریانس) کار می کرد. در این تحلیلهای، 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 کدگزاری بهینه و بازنمایی تصویر در قشر مخ با معرفی این دو روش مدلسازی تنگاتنگ تر شد. این دو مدل، یعنی تحلیل مؤلفه های مستقل(Bell and Sejnowski 1997; Comon 1992) و کدگزاری پراکنده(Olshausen 1996; Olshausen and Field 1997) نشان می دادند که از میان تمام کدهای خطی ممکن، فیلترهایی که نماینده میدان گیرنده سلولهای ساده هستند، برای تصاویر طبیعی انتخاب بهینه هستند.

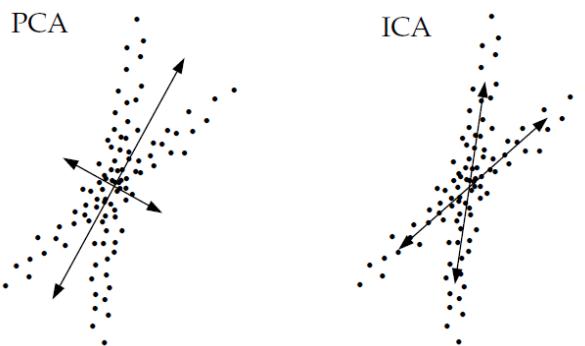
در مدل ICA هر تصویر \vec{x} (در اینجا برداری از شدت روش‌نایی پیکسلها) توسط یک ترکیب خطی از تابعهای پایه \vec{A}_j که با ضرایب s_j وزنده می شوند بازنمایی می شود:

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

تلاش ICA بر این است که توابع پایه ای مانند 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 A مربعی و معکوس پذیر باشد) و ضرایب از رابطه $\vec{x} = A^{-1}\vec{s}$ 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 توزیعهایی 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، وابستگهای مراتب بالاتر را نیز در ورودی تفکیک می کند. همانگونه که اشاره شد، تحلیل مؤلفه اصلی یک راه کدگزاری 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 ۱۳-۳ 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 متقابل 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 کدگزاری 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 کوتوله‌سیس بالا خواهد بود.



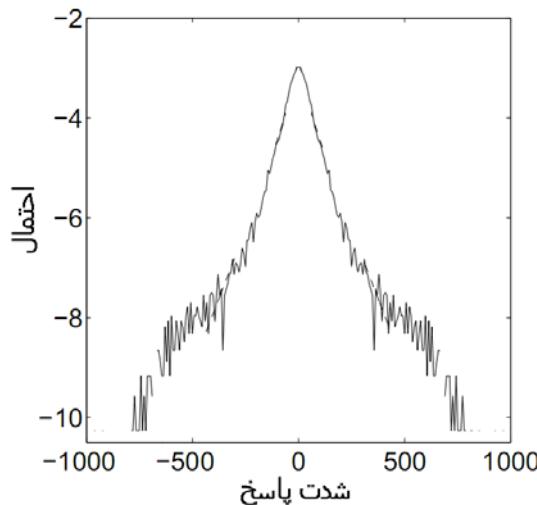
شکل ۱۳-۳: مثالی از توزیع داده‌های دوبعدی و محورهای تحلیل مؤلفه‌های اصلی(چپ) و تحلیل مؤلفه‌های مستقل(راست) متناظر

بل و سجنوسکی^{۱۶۲} الگوریتمی برای تفکیک مؤلفه‌های 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 آنتروپی 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 در حالیکه ICA آنتروپی توأم خروجیها را بیشینه می سازد. الگوریتم بل و سجنوسکی ماتریس وزن A^{-1} را با روش شبیه نوردهی صعودی روی آنتروپی توأم \vec{y} یکتابع انتقال غیرخطی داده‌های ورودی، $g(A^{-1}\vec{x}) = \vec{y}$ ، پیدا می کند. بیشینه کردن آنتروپی توأم خروجی هم ارز بیشینه کردن اطلاعات متقابل بین ورودی و خروجی (یعنی بیشینه کردن انتقال اطلاعات) است. دلیل این امر اینست که $H(\vec{y} | \vec{x}) = H(\vec{x}) + H(\vec{y}) - H(\vec{y} | \vec{x})$ ، که در آن تنها $H(\vec{y})$ به ماتریس وزنهای A^{-1} مرتبط است چون ($\vec{y} | \vec{x}$) نویز است. بیشینه کردن آنتروپی توأم خروجی، اطلاعات مشترک بین تک تک عناصر را تشویق به کوچک بودن می کند(در ک. فصل ۲). هنگامی که تابع انتقال غیرخطی g با توزیع تجمعی سیگنالهایی که مربوط به داده‌های \vec{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 چندبعدی با نظریه بارلو درباره کدگزاری کمینه آنتروپی سازگار است. چنان که گفته شد بیشینه کردن آنتروپی 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 آنتروپی توأم ثابت است (و در بیشینه مقدار خود قرار دارد)، راه حلی که اطلاعات متقابل را کمینه می کند، آنتروپی های جانبی (تک تک) هر واحد خروجی را نیز کمینه می کند.

¹⁶² Bell and Sejnowski

باید توجه شود که کاهش افزونگی کد نورونی تنها یکی از اصلهایی است که می‌تواند 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 نورون‌های 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 می‌باشد. در بقیه موارد پاسخ نورون از نظر مقداری بزرگ می‌باشد. شکل ۱۴-۳ نمودار توزیع پاسخ یک نورون نمونه را نشان می‌دهد.



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

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

اتیک^{۱۶۴} کدهای فشرده و ناهمبسته ای مانند PCA را به دلیل کارایی کدگزاری می‌پسندید (Atick 1992). فیلد^{۱۶۵} کدهای پراکنده را برای چنین منظوری ترجیح می‌دهد (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 دنباله‌های آن ندرتاً رخ می‌دهند. بنابراین بیشینه کردن پراکنده‌گی توزیع یک پاسخ معادل کمینه کردن آنتروپی است و کدهای پراکنده مزایایی مشابه کدهای آنتروپی دارند که از آن جمله می‌توان به تفکیک افزونگی‌های مرتب بالا و نیز افزونگی مرتبه دوم اشاره کرد. در چنین کدهایی، افزونگی بین عناصر ورودی به افزونگی درون الگوهای پاسخ تک تک ورودیها تبدیل می‌شوند که بدینسان تک تک خروجی‌ها، مگر در موارد محدودی، پاسخ یکسانی می‌دهند.

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 تأیید کدهای آنتروپی کمینه بارلو خواهد بود. کدهایی که تعداد نورونهای فعال را کمینه کنند برای تشخیص هم وقوعیهای مشکوک می‌توانند مفید واقع شوند. از آنجا که پاسخ غیرصرف‌هر واحد سا توجه به توزیع آن-نسبتاً متوقع نیست، روابط مراتب بالا بطور فزاینده‌ای کمیاب می‌شوند، و نتیجتاً هنگامی که در محرك ظاهر شوند اطلاعات زیادی را در بر خواهند داشت. فیلد تباین این روش را نسبت به کدهای فشرده‌ای 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 فعل شدن واحد، و نه میزان فعالیت آن نشان داده می‌شود. 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 شده‌اند. مزیت دیگر کدگزاری پراکنده (از جمله برای بازنمایی چهره) قابلیت ذخیره آن در حافظه‌های انجمانی است. شبکه‌هایی با ورودی پراکنده حافظه بزرگتری دارند و بازیابی آنها با قسمتی از اطلاعات کاراتر است. فیلد شواهدی را ارائه کرده است که فیلترهای گابور جهتدار هنگامی که روی تصاویر طبیعی اعمال شوند کد پراکنده تولید می‌کنند 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 کدگزاری پراکنده با فیلترهای گابور اساساً به طیف فاز داده‌ها وابسته است (Bartlett 1998). اولشاون و فیلد¹⁶⁶ یک شبکه را برای بازسازی تصاویر طبیعی از روی ترکیب خطی تصاویر پایه مجھول آموزش دادند (Olshausen 1996). معیار مورد استفاده در این روش، کمینه میانگین خطای مربعی است که به تنها می‌تواند به یک ترکیب خطی از مؤلفه‌های اصلی تصویر همگرا شود. هنگامی که یک معیار پراکنده‌گی به تابع هدف اضافه شود، تصاویر پایه مشابه خصوصیات پاسخ سلولهای ساده V1 محلی هستند، راستای خاصی دارند، و از لحاظ مکانی متضادند¹⁶⁷. 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 پراکنده‌گی فعالیت ضرایب حاصل از نگاشت خطی را اندازه گیری می‌کرد. بهینه سازی این معیار در تصاویر طبیعی توابع پایه‌ای را نتیجه داد که شباهت بسیاری به گزینندگی نورون‌های V1 داشتند. مدل کدگزاری پراکنده نویز افزایشی گوسی را نیز در مدل خود گنجانده است، $E = A\vec{s} + \vec{x}$ ، 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 فراکامل¹⁶⁸)، فعالیت نورونها دیگر تابعی قطعی از داده‌ها نمی‌باشد زیرا حالتهای 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 پایه‌های بهینه است و محاسبه احتمال ثانویه $p(\vec{s} | A, \vec{x})$ حالت ضرایب را دشوار می‌کند ولی از سوی دیگر به مدل امکان کار کردن با داده‌های حسی و تعداد متغیر نورون برای بازنمایی شکل را می‌دهد.

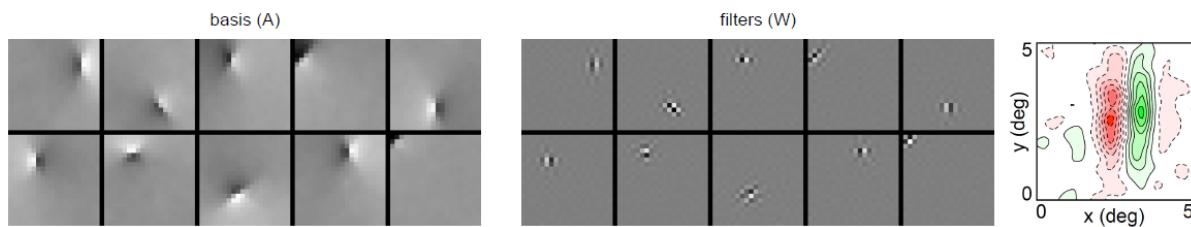
هر دو مدل وقتی با تکه‌های کوچکی از تصاویر طبیعی سازگار می‌شوند و بیشگاهی راستاداری را می‌سازند که میدان گیرنده‌گی سلولهای ساده را تداعی می‌کنند (شکل ۱۵-۳). این نتایج مستقیماً از اصل کدگزاری بهینه مشتق می‌شوند و فرضیه جدید را درباره عملکرد نورونهای V1 ارائه می‌دهند: بجائی عمل کردن به عنوان لبه یاب و یا تحلیل تصویر از روی

¹⁶⁶ Olshausen and Field

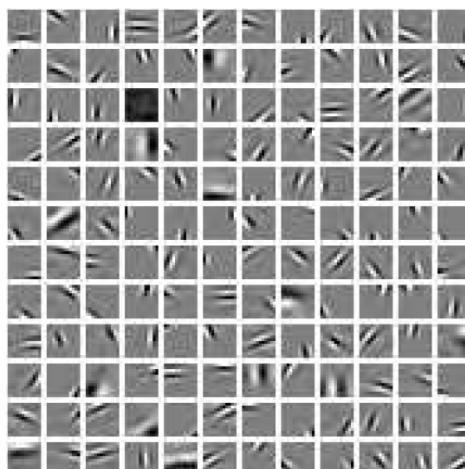
¹⁶⁷ spatially opponent

¹⁶⁸ overcomplete

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



شکل ۳-۱۵: روش تحلیل مؤلفه‌های مستقل وقتی روی تکه تصویرهای طبیعی اعمال شود میدان گیرندگی شبکیه سلولهای ساده بدست می‌دهد. (چپ) زیرمجموعه‌ای 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 (محركی 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 حاشیه‌ای معیار کورتوسیس. شباهت این توابع با گزینندگی نورون‌های V1 انکار ناپذیر است.

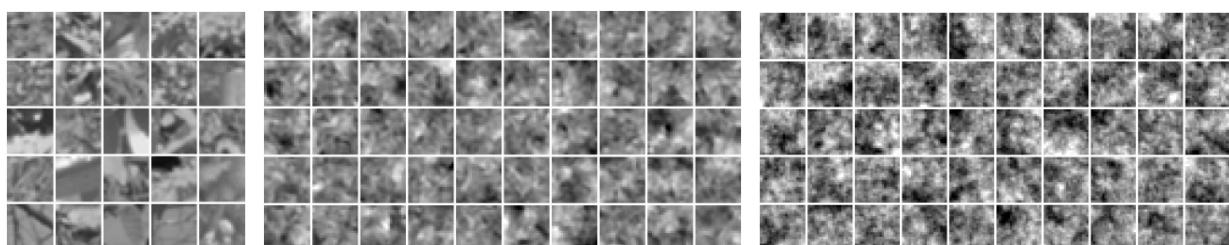
اهداف محاسباتی و قیود نورونی دیگری که برخی از آنها به کاهش افزونگی مرتبط‌بند نیز به تصاویر طبیعی اعمال شده اند. مدل‌های خطی که هزینه انرژی^{۱۶۹} را کمینه می‌کنند(Baddeley 1996)، فعالیت نورونها را ملزم می‌کنند که در

^{۱۶۹} energy expenditure

طول زمان به نرمی تغییر کنند (Hurri and Hyvärinen 2003)، یا شبیه تعامل^{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 آکتاو هستند (Simoncelli 2005) و یک دسته از آنها در شکل ۳-۱۶ آورده شده است. در نتیجه چندین توجیه عملکردی برای شکل میدان گیرندگی نورونها 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 نگاشت خطی از تصاویر، سعی در کم کردن میزان وابستگی بین ضرایب حاصل از این نگاشت خطی داشتند. با وجود اینکه این مدلها توانسته اند با موفقیت یک توجیه عملکردی برای پردازش تصویری اولیه در سیستم بینایی ارائه دهند، دچار محدودیتهايی نيز هستند.

نخست آنکه به دلیل خطی بودن مدلهاي مولدزيربنايی آنها، تنها آن دسته از وابستگيهای آماری را در بر می گيرند که از ترکيب خطی کميتهای مستقل حاصل شده اند. اگر اين مسئله برای تمام ساختارها برقرار بود انتظار می رفت که اولاً تصاویری که از اين مدلها نمونه گيري می شوند از نمونه های تصاویر طبیعی غیر قابل تمایز باشنند و ثانیاً ضرایبی که برای کدگزاری تصاویر استفاده می شوند (و مستقل فرض شده اند) واقعاً مستقل باشنند. ولی به سادگی می توان نشان داد که قسمت اول صادق نیست. هرچند که تصاویری که از مدلها نمونه گيري 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 قادر ساختارهای تصویری مانند کانتورهای طولانی، بافت‌های عادی، و نواحی همگن هستند (Karklin and Lewicki 2008).

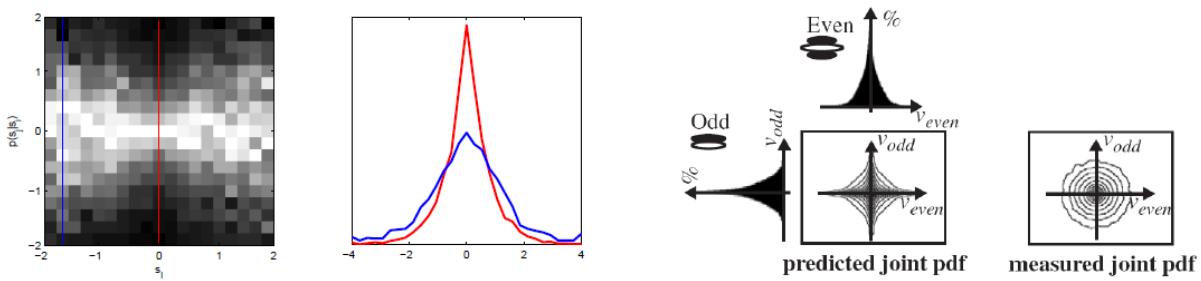


شکل ۳-۱۷: تصاویری که از یک مدل گوسی (راست) و یک مدل ICA با ضرایب پراکنده (وسط) تولید شده اند. هیچکدام از آنها مشابه تکه های تصاویر طبیعی (چپ) نیستند. (ر.ک. به متن)

دوم آنکه نشان داده شده است که ضرایب مدلهاي 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

^{۱۷۰} quasi-orthogonality

موقعیتهای مجاور درون تصاویر هستند یا راستاهای نزدیکی دارند قویترین همبستگی را از خود نشان می‌دهند. شاهد دیگری برای وابستگی مراتب بالا اینست که ضرایب به خوبی توسط توزیعهای مجذعاً توصیف نمی‌شوند یعنی توزیع توأم آنها را نمی‌توان از ضرب توزیعهای حاشیه‌ای بدست آورد (Zetzsche and Rhrbein 2001). این نکته مبنی آنست که حتی با داشتن یک دیدگاه صرفاً آماری، کدگزاری 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 می‌کند (Lyu and Simoncelli 2009).



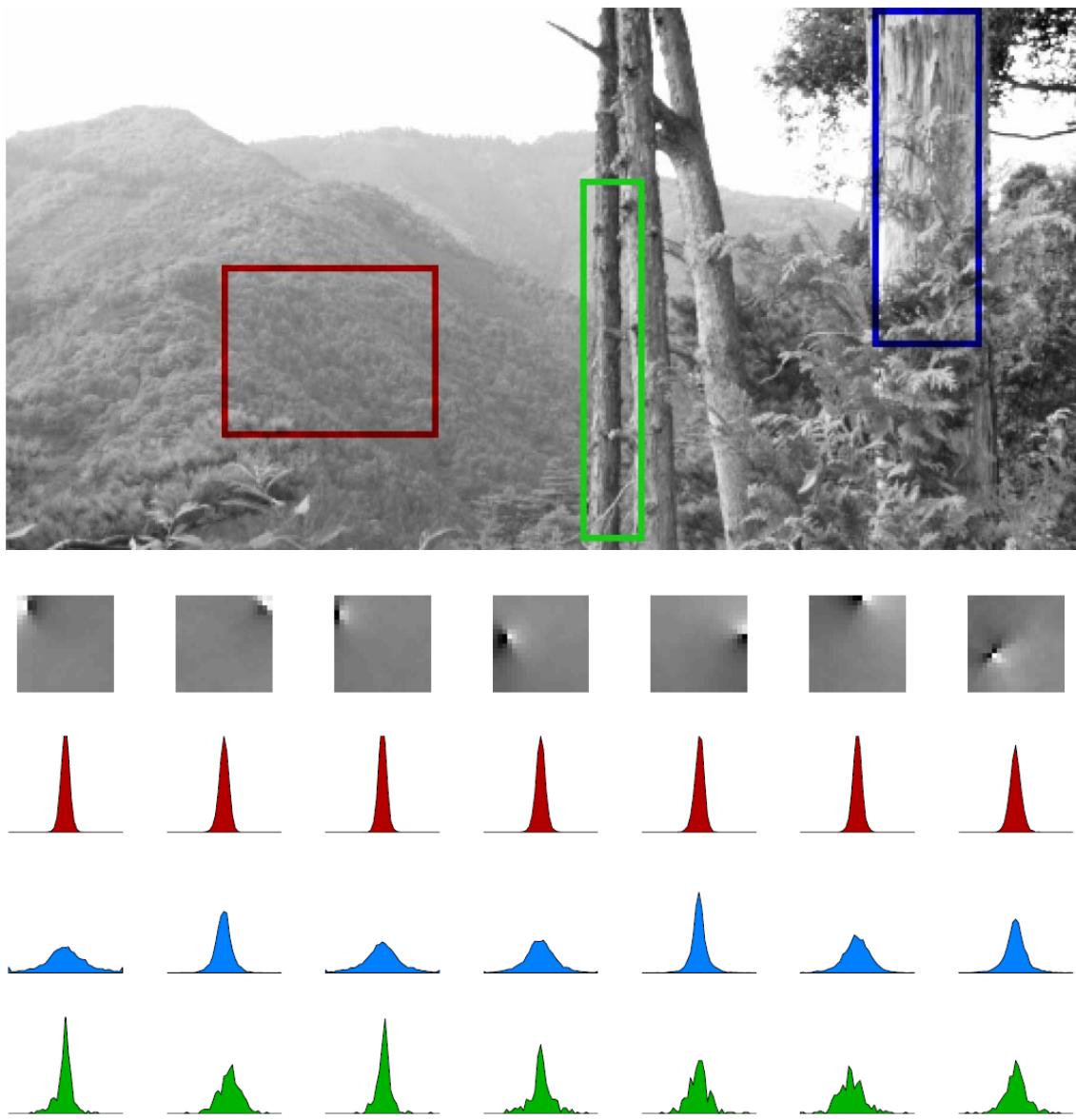
شکل ۳-۱۸-۳: وابستگی بزرگی بین زوجهای ضرایب (چپ) هیستوگرام شرطی، هر برش عمودی توزیع s_j که روی مقادیر s_i شرطی است را نشان می‌دهد (که به طور قائم قرار دارد و شدت روشناختی آن برای هر ستون به طور مستقلی نرمال شده است) (شکل پایینی این توزیع نشان می‌دهد که واریانس s_j با مقادیر بزرگتر $|s_i|$ افزایش می‌یابد.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

سومین کمبود این مدلها اینست که آنها فرض می‌کنند که آمارگان داده‌ها فرق نمی‌کنند و به عبارت دیگر توزیعهای احتمالاتی آنها ایستا^{۱۷۱} است. برای مثال به محض اینکه پارامترهای مدل ICA تطبیق یافته‌ند، توابع احتمال اولیه و پایه ثابت می‌شوند و به یک توزیع ایستا روی تمام مجموعه تصاویر منجر می‌شوند. این به شکل تابع احتمال اولیه که توزیع احتمال اولیه آنها سازگار پذیر و یا کاملاً غیرپارامتری است نیز ثابت است. با اینحال آمارگان تصاویر بر اساس محتوا آنها (شکل ۳-۱۹)، دگرگونی خصوصیات فیزیکی محیط یا شرایط اخذ تصویر تغییر می‌کند. همچنین گزارشاتی درباره تغییرات در آمارگان تصویر به عنوان تابعی از پهنه‌ای همسایگی مکانی مجموعه ضرایب موجود (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 داده‌ها روی بدنه عظیم تصاویر آموزشی بدهست می‌دهد ولی تغییراتی که در زیرمجموعه‌های مجموعه داده‌ها دیده می‌شود را منعکس نمی‌کند.

همانطور که در شکل ۳-۱۷ نیز مشاهده شد، تبدیل خطی تکه‌های تصویر بازنمایی قابل قبولی را از ساختارهای بینایی که در تصاویر طبیعی دیده می‌شود را نشان نمی‌دهد. بافت‌ها، لبه‌ها، گوشه‌ها، و سایر ویژگیهای ظاهرآمهم در بازنمایی متمایز نشده‌اند. قسمتهای ثابت در تصاویری از یک نوع خاص (مثلاً بافت‌های همگن) باعث نوسان شدید خروجی فیلترهای خطی می‌شوند. 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 نشده‌ای را

¹⁷¹ stationary

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



شکل ۳: واریانسهای ضرایب ICA از یک محتوا تا محتوا 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 محرك مورد نظر در ناحیه دریافت خود حساس نیستند. به این ترتیب، نمی توان با استفاده از یک ترکیب خطی از تصویر ورودی به خصوصیاتی

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 راستدار 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. مدلهای جدایی پذیر خطی شالوده خوبی برای مدلهای آماری قویتر از تصاویر طبیعی فراهم کرده بودند و چندین توسعه از آنها برای اخذ وابستگیهای 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 (زیرفضاهای) 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 توبوگرافیک^{۱۷۴} معروف است، مجموعه های افزای توابع پایه وابسته با یک چینش توبوگرافیک که وابستگیهای 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 بازهم ویژگیهای راستدار و محلی است که اینبار در مجموعه هایی مرتبط یا نقشه های توبولوژیکی مرتب شده اند. واحدهای همسایگی سرکشی شده ویژگیهای پیچیده تری از تکه های تصویری بازنمایی می کند و برخی ویژگیهای نامتعین سلولهای پیچیده را نشان می دهد (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 توبوگرافیک بتوانند آنرا 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 مندیهای آماری عمومی تصویرهای پیچیده مانند بافتها رنج می بردن.(Karklin 2007)

ایرادات وارد شده به مدلهای مبتنی بر توابع پایه خطی دامنگیر کدگزاری پراکنده نیز می شود. البته درباره کدگزاری پراکنده باید گفت که هنگامی که بازنمایی فراکامل باشد یا سطح نویز چشمگیر باشد، مرحله کدگزاری (بدست آوردن بازنمایی بهینه برای تصویر داده شده) غیرخطی است و ضرایب تابعهای پایه برای بازنمایی هر چه بهتر تصویر و بیشینه کردن پراکنده گی 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 VI (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 زمینه‌های مختلفی در حال تهیه پایه‌های متعامد یکه چند مقیاسه برای سیگنال و تصویر بودند که امروزه به نام موجک^{۱۷۵} مشهورند. این پایه‌ها تخمینهای خوبی از پایه‌های 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 یکنواخت^{۱۷۶} 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 یافته^{۱۷۷} (که به نامهای تابع نمایی کشیده شده یا تعمیم یافته^{۱۷۸} 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 $(\frac{1}{p})\Gamma(\frac{1}{p})$ است. توان 2 با توزیع گوسی متناظر است، و توان 1

به توزیع لاپلاسین مربوط است. به صور کلی مقادیر کوچکتر p به توزیعی منتهی می شوند که هم در صفتراکمتر است و هم دنباله های گستردۀ تری دارد (شکل ۱۰-۲). هر کدام از هیستوگرامهای شکل ۱۲-۳ 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 مدعای توسعه معیارهای

¹⁷⁵ wavelet

¹⁷⁶ independent and identically distributed (iid)

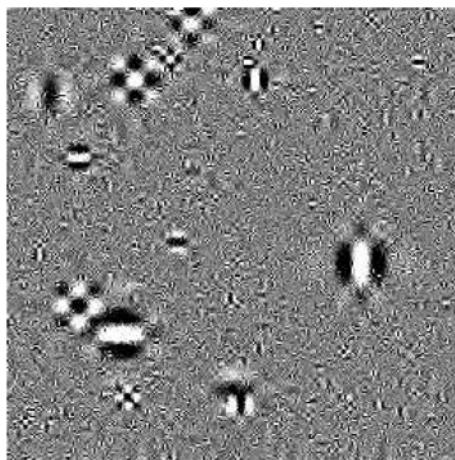
¹⁷⁷ generalized Gaussian

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 $(0.0/4)$ است و عامل δ 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 توزیعهایی که بالاتر به آن اشاره شد از یک توزیع گوسی با واریانس مشابه بسیار کمتر است و مستقیماً به بالاتر رفتن کارایی کدگزاری می انجامد. در رفع نویز نیز استفاده از این مدل به عنوان توزیع اولیه تصویر منجر به بهبود Simoncelli 96, Chipman 97, Abramovich 98, Moulin 99, ۱۷۹ فاخشی نسبت به مدل گوسی می شود (Simoncelli 99). تخمین زنهایی که از این روش بدست می آیند تابعهای هسته ای ψ 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 موقیتها، 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 متعدد یکه باشد، به آسانی می توان نمونه های آماری از مدل گرفت. شکل ۳-۲۰ نتیجه نمونه گیری از ضرایب بازنمایی موجک را مستقل از توزیعهای گوسی نرمال شده نشان می دهد. پارامترهای توزیع برای هر زیرباند آنهایی انتخاب شده اند که با یک تصویر خاص (در اینجا تصویر انسیتین) 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 ۹-۳) دارای ساختهای بیشتری است ولی باز هم 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 انسیتین (شکل ۳-۲۱) تطبیق داشته باشد.

مدل حاشیه ای موجک می تواند با توسعه خود روی یک پایه فراکامل موجک بهبود یابد. به زبان دقیقت، ψ^{180} و همکاران نشان داده اند (Zhu 98) که با استفاده از نسخه ای از نظریه برش نمای فوریه ψ^{181} که برای بازسازی تصاویر

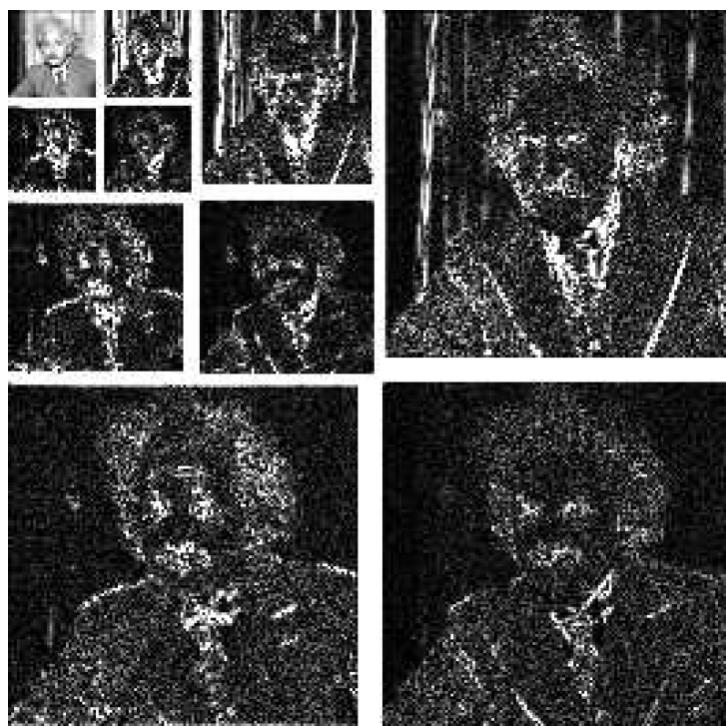
¹⁷⁹ kernel functions

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 بکارگیریشان برای تولید راه حلهایی برای کاربردهای پردازش تصویر دشوار است. برای راحت کردن مسئله، توصیفهای آماری توأم بکار گرفته شد.

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



شکل ۲۱-۳: بزرگی ضرایب چندمقیاسه موجک برای تصویر انیشتین. هر تصویر کوچک بزرگی ضرایب زیرباند را که از کانوالوشن با فیلتری در راستا و مقیاسی متفاوت بدست می آید و با فاکتوری مناسب نمونه گیری شده است نشان می دهد. ضرایبی که از لحاظ مکانی در یک باند به هم نزدیک هستند تمایل بیشتری به داشتن بزرگی های 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 موقعیتها، راستاهای 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 محتملتند. این خاصیت خوشه بندی در یک روش هیوریستیک بسیار کارا درون کدکننده موجک درختِ صفر پنهان¹⁸³(Shapiro 93)(EZW) 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 واریانسی: بیش از ربع قرن پیش، لی¹⁸⁴(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 وايانس را یک ثابت عمومی برای سراسر تصویر قلمداد می کرد.

رودمن¹⁸⁵(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 راستاهای نزدیک به استفاده از این واریانس تخمین زده برای رفع نویز خوشه توجه شد(Malfait 97, Simoncelli 97, Chang 98, Simoncelli 99, Michak 99, Wainwright 2001, Abramovich 2002).

مدل های غیر خطی مختلفی برای برطرف کردن این نوع وابستگی ارائه شده اند. به عنوان مثال شوارتز¹⁸⁶ از مکانیزم نورونی نرمال سازی با تقسیم¹⁸⁷ در رفع افزونگی استفاده کرد(Schwartz 01). نرمال سازی پاسخ یک نورون هنگامی اتفاق می افتد که محرک ورودی به نورون، بزرگتر از میدان گیرنده‌ی 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

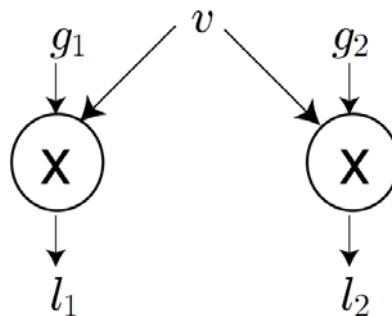
¹⁸⁴ Lee

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 ۲۲-۳ توجیه می‌شود. در این شکل، پاسخ دو فیلتر وابسته با l_1 و l_2 مشخص شده‌اند. دو متغیر مستقل گوسی به نام‌های g_1 و g_2 در یک متغیر تصادفی عددی به نام v ضرب می‌شوند تا پاسخ نورون‌های وابسته را ایجاد کنند. در این مدل، متغیر v مسئول ایجاد وابستگی بین متغیرهای l_1 و l_2 می‌باشد. در تحلیل‌های آماری نشان داده می‌شود که این متغیر تنها باعث وابستگی واریانس در نورون‌های 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 پاسخ‌ها از بین می‌رود. شوارتز از این واقعیت استفاده کرد و پاسخ نورون را به مربع پاسخ‌های نورون‌های اطراف تقسیم کرد تا پاسخ مستقل حاصل شود.



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

در مدل شوارتز، هر نورون به وسیله‌ی اتصالات افقی با نورون‌های مجاور خود ارتباط دارد. این مدل نورون‌های $V1$ را با استفاده از تحلیل مؤلفه‌های مستقل به دست می‌آورد و سپس با استفاده از اتصالات افقی بین نورون‌ها، سعی در برطرف کردن وابستگی بین آنها دارد. واریانس هر نورون به صورت حاصل‌جمع مربعات پاسخ نورون‌های 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 نرمال‌سازی را نشان می‌دهد و C_x یک همسایگی اطراف نورون x است که نورون‌های آن دارای وابستگی واریانس با x هستند. وزن اتصال افقی بین نورون‌های x و y با w_{xy}^l نشان داده می‌شود و σ_x^2 بخشی از واریانس x است که به پاسخ نورون‌های 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_y پاسخ مستقل برای نورون y می باشد. پارامترهای مدل با استفاده از تخمین درستنمایی بیشینه^{۱۸۸} روی یک مجموعه از تصاویر طبیعی به دست می آید (برای سادگی فرض شده که پاسخ نورون ها دارای توزيع نرمال می باشند):

$$\hat{w}_{xy}^l, \hat{\sigma}_x = \arg \max_{w_{xy}^l, \sigma_x} \prod_i p(L_x(i) | L_y(i), y \in C_x) \quad (14-3)$$

که در معادله بالا p تابع توزيع احتمال توان 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. مدلها 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 توانم دو فیلتر خطی وقتی با همان تصویر کانوالو 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 توزيع توان شرطی منعکس می شود و این توزيع شکلی مشابه پایپیون دارد که خود نشان دهنده وابستگی واریانس است (Buccigrossi & Simoncelli, 1999; Schwartz & Simoncelli, 2001) یا با بررسی توزيع حاشیه ای در برابر توزيع توان می توان به این نتیجه رسید (Zetzsche et al., 1993; Zetzsche & Nuding, 2005). هوانگ و مامفورد^{۱۹۱} نمودار کانتور توان را برای یک پایگاه داده بزرگ تصاویر بررسی کردند و وابستگی توان را به صورت یک توزيع 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 مدلها قبل می توان مدل بسیار صریحتری از طریق بررسی آمارگان ضرایب ساخت (Simoncelli 97). سطر بالای شکل ۳-۲۳. هیستوگرام توان چند جفت مقاوت از ضرایب موجک را نشان می دهد. همانند توزيعهای حاشیه ای، برای اینکه این هیستوگرامهای توان که از جفت‌های مختلفی از ضرایب که در جاهای مختلف تصویر پخش شده اند را بتوانیم به عنوان نماینده‌گان توزيع آنها در نظر بگیریم، نیاز به فرض همگنی داریم. ضرایبی که از توابع پایه مجاور باشند کانتورهایی تقریباً

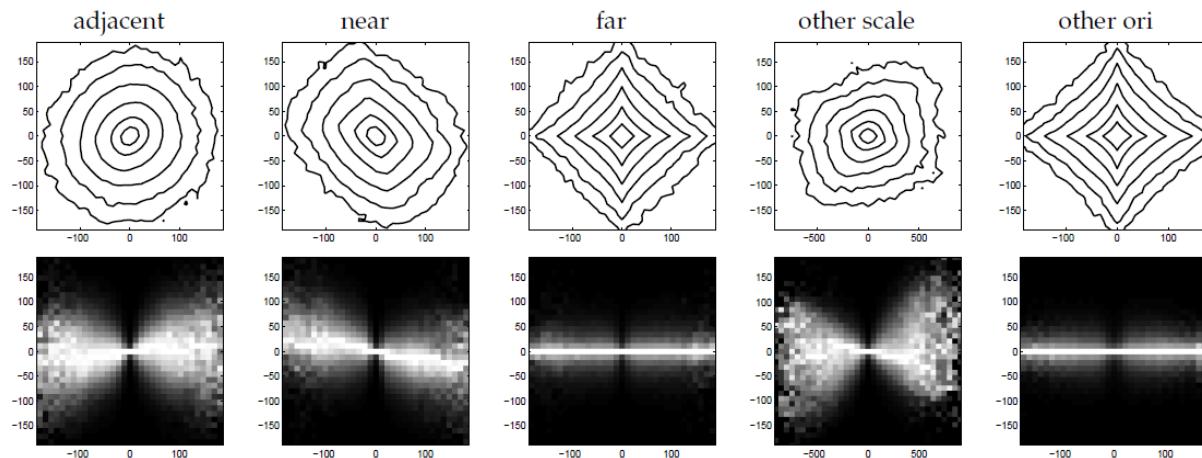
¹⁸⁸ maximum likelihood estimate

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. زشه^{۱۹۲} (Zetzsche 93) توزیعهای تجربی جفتهای رُبعی^{۱۹۳} (تبديل 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,Srivastava 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 ۹-۳ 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 زوچهایی از توابع پایه متناظرنده مقیاس و راستای یکسانی دارند ولی فاصله مکانی آنها متفاوت است. نمودار بعدی با زوچی از مقیاسهای مجاور متضاد است (با راستای یکسان و موقعیت مکانی تقریباً مشابه) و نمودار راست با زوچی 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 مناسب 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 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 باز هم وابستگی آماری از خود نشان می دهند!

¹⁹² Zetzsche

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. مدل هایی 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 توجیه است.

یکی از مطالعات جدیدی که خصوصیات آماری تصاویر را مورد بررسی قرار داده، وابستگی بین واریانس ها بر اساس زمینه^{۱۹۷} 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 ۱۹-۳ 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 ۲۳-۳ دیده می شود به طور اعجاب آوری برای بسیاری از تصاویر پایدار است. به علاوه شکل کیفی این روابط آماری برای جفت‌های ضرایب موجود در موقعیت‌های مکانی مجاور و راستاهای مجاور صادق است. با نگاهی به ضرایبی که فاصله بیشتری دارند (چه از نظر مکانی و چه از نظر مقیاس) این وابستگی کمتر به چشم می خورد و به ذهن الق 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 واریانسی که با توجه به مکان نوسان می کند^{۱۹۸} مدل شوند. مثال مناسبی برای اینکار از ضرب یک بردار گوسی در یک ضریب اسکالر پنهان بدست می آید که به مدل مخلوط مقیاس گوسی^{۱۹۹} (GSM) مشهور است. این توزیعها زیرمجموعه مهمی از توزیعهای متقارن بیضوی را بازنمایی می کند که شامل آنهایی است که به عنوان تابعی از \sqrt{n} ربعی بردار تصادفی محسوب می شوند. این نوع از مدلها که در حوزه تصادفی تعییه شده اند در زمینه پردازش گفتار بسیار سودمند واقع شده اند (Brehm 87). یک مجموعه مرتبط از مدلها به مدلها خود رگرسیون واریانس-متغیر شرطی^{۲۰۰} (ARCH) مانند (Bollersley 94) معروفند و برای بسیاری از سیگنالهای حقیقی که دچار نوسانات ناگهانی و پس از آن دوره های نسبتاً

^{۱۹۷} context

^{۱۹۸} این خاصیت به heteroscedasticity معروف است.

^{۱۹۹} gaussian scale mixture model

^{۲۰۰} autoregressive conditional heteroskedastic model

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

مدلهایی مانند GSM از معماریهای انعطاف پذیری استفاده می‌کنند ولی به تبدیل خطی ثابتی مانند تجزیه چندمقیاسه موجک اتکا دارند و به کمک آنها وابستگی بین ضرایب خود را مدل می‌کنند. مدل GSM حاصل نگرش متفاوتی به خاصیت خودسازماندهی تصاویر است (&Simoncelli, 2000; Wainwright et al., 2001 Wainwright) که از Andrews and Mallows 1974 (Andrews and Mallows 1974; Buccigrossi and Simoncelli, 1974; Wainwright et al., 2001 Buccigrossi and Simoncelli, 1999; Romberg et al., 2001) ضرایب تبدیلهای خطی (مانند هرم چندمقیاسه موجک) را به صورت حاصلضرب متفاوتی آن بیش از سه دهه می‌گذرد (Andrews and Mallows 1974). مدل مخلوط مقیاس گوسی (Buccigrossi and Simoncelli, 1999; Portilla et al., 2001; Wainwright et al., 2001) بهبود بخشد که به طور ضمنی نوید پیدا کردن توصیف بهتری از آمارگان تصویر را می‌دهد. این مدلها برای توجیه وابستگیهای آماری مشاهده شده در تصاویر طبیعی طراحی شده اند ولی در عین حال می‌توانند برای شرح دادن خصوصیتهای غیرخطی نورونهای حسی نیز استفاده شوند و بر این نکته که بازنمایی نورونی برای آمارگان صحنه‌های طبیعی بهینه سازی شده اند، صحه بگذارد. برای مثال این مدلها تخمینهای محلی واریانس را برای ورودیهای فیلترهای خطی حساب می‌کنند. نرمال کردن خروجی فیلترها کد مستقلتری بدست می‌دهد که بهتر از مدل‌های خطی ساده با پاسخهای نورونی مطابق است، و در ضمن رفتارهای پیچیده سلولهای ساده نظیر اشباع پاسخ کنتراست^{۲۰۱}، بازداری فعالیت به ازای راستهای متقاطع، و فرونشانی ماسک^{۲۰۳} را نیز از خود بروز می‌دهد (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 می‌کند و یک مخلوط \vec{l} را طبق رابطه زیر می‌سازد که توزیع آن نیز در رابطه آمده است:

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

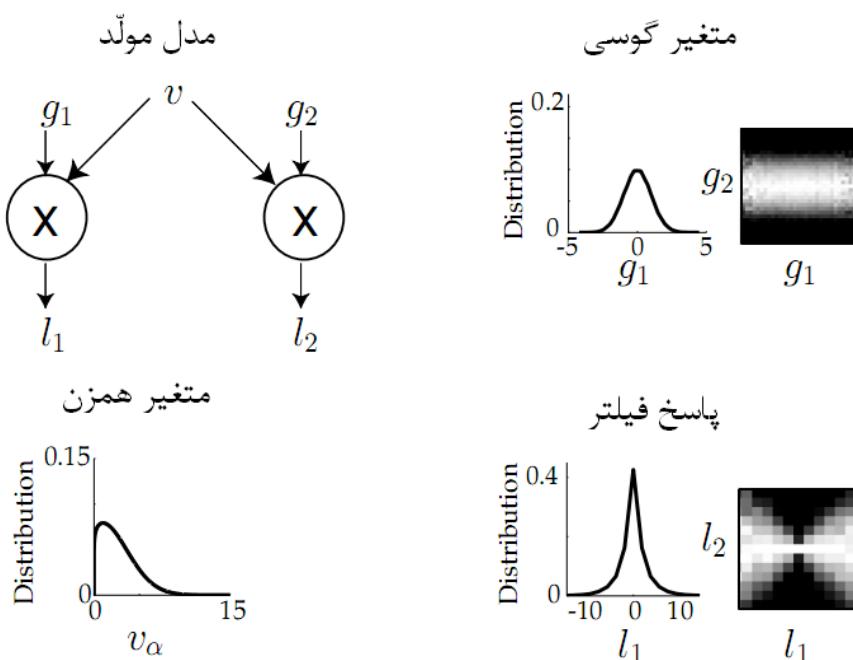
$$p(\vec{l}) = \int \frac{1}{(2\pi)^{\frac{m}{2}} |v^2 \Sigma|^{\frac{1}{2}}} \exp\left(-\frac{\vec{l}^T \Sigma^{-1} \vec{l}}{2v^2}\right) p(v) dv \quad (15-3)$$

²⁰¹ mixer variable

²⁰² contrast response saturation

²⁰³ mask suppression

که m تعداد فیلترها(بعد)، Σ ماتریس کواریانس، و متغیر v با توجه به توزيع (v) مخلوط شده است و این رابطه در يك 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 يك فیلتر خطی وقتی که روی يك 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(بالا-چپ) يك مدل مولّد 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) نیزتابع توزيع حاشیه ای و شرطی توام گوسی متغیرهای گوسی و همزن را برای داده هایی که از مدل ساخته می شوند، نشان می دهد.



شکل ۳: (بالا-چپ) مدل مولّد برای يك 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 پایپون شکل). برای آمارگان توأم شرطی، شدت روشنایی با شمارش هر ظرف^{۲۰۶} هیستوگرام بدست می آید و هر ستون برای رسیدن به تمام بازه شدت روشنایی مستقلانه نرمال می شود. (پایین-چپ) توزيع حاشیه ای متغیر همزن مدل GSM. متغیر همزن طبق تعریف باید مثبت باشد و در اینجا توزيع رایلیخ^{۲۰۷} با پارامتر $a=0.1$ برای آن برگزیده شده است. در هر حال توریع 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 يك توجیه بالا به پایین برای دو مشخصه پایین به بالا آمارگان تصاویر طبیعی که قبلاً به آنها اشاره شد بدست می دهد: تابع توزيع حاشیه ای يك فیلتر با کوتسيس بالا و آمارگان شرطی توأم دو فیلتر خطی که يك متغیر

^{۲۰۴} البته اين مدل داراي نمايشهای دیگر مانند $\vec{x} = \sqrt{zu}$ است که در کارکرد مدل تاثیری ندارد.

^{۲۰۵} Mixer variable

^{۲۰۶} bin

^{۲۰۷} Rayleigh distribution

همزن مشترک دارند (Wainwright & Simoncelli, 2000; Wainwright et al., 2001). شکل ۳-۲۴(بالا-چپ) توزیع حاشیه‌ای و شرطی توأم پاسخ دو فیلتر I_1 و I_2 را بر مبنای داده‌های ساخته شده در شکل ۳-۲۴(بالا-راست) و ۳-۲۴(پایین-چپ) نشان می‌دهد.

مدل GSM رابطه نزدیکی با روش‌های پایین به بالا آمارگان تصویر و بازنمایی قشر مخ دارد. اولاً مدل‌های کدگزاری پراکنده و بازنمایی میدان گیرنده‌گی قشر مخ معمولاً از خصوصیات زبرگوسی توزیع‌های حاشیه‌ای پاسخ‌های فیلتر استفاده می‌کنند که به طور طبیعی در مدل مولد GSM بوجود می‌آید (شکل ۳-۲۴-۳ پایین-راست). ثانیاً GSM‌ها توجیه‌ای برای مختصات فیلتر پیشنهاد می‌دهند از آن قبیل می‌توان به چارچوب حباب^{۲۰۸} (Hyvärinen, Hurri, & Vayrynen, 2003) اشاره کرد. وقتی که در مدل GSM پاسخ فیلترها از یک متغیر همزن مشترک استفاده می‌کند (شکل ۳-۲۴-۳ پایین-راست) مختصات بوجود می‌آید. ثالثاً برخی از چارچوبهای پایین به بالا مستقیماً از نسخه‌هایی از دو مؤلفه GSM استفاده می‌کنند. مثلاً مدل‌های آمارگان تصویر و کنترل بهره قشر مخ^{۲۰۹} (Schwartz & Simoncelli, 2001) به نرمال کردن با تقسیم مؤلفه‌های خروجی می‌پردازند که دارای مشخصاتی است که از لحاظ آمارگان حاشیه‌ای و توأم یادآور مؤلفه گوسی Ruderman & Bialek (Wainwright & Simoncelli, 2000) است (شکل ۳-۲۴-۳ بالا-راست). در پژوهشی^{۹۴} نشان داده شده است که پیکسلهای یک تصویر (نه حاصل کانولوشن پاسخ فیلتر با تصویر) را می‌توان 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 می‌کند.

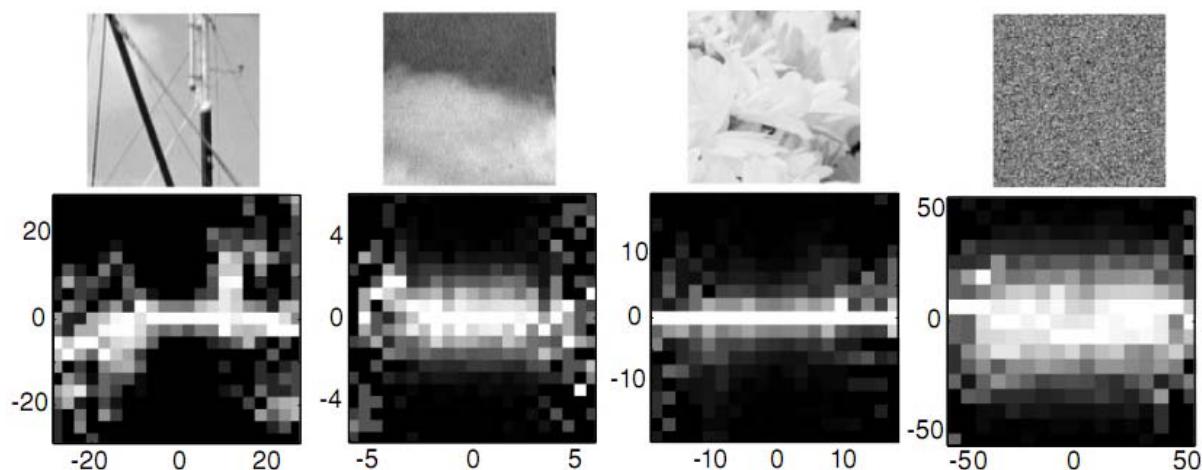
در فرمولاسیون اصلی GSM، یک همزن برای یک مجموعه از متغیرهای گوسی وجود دارد و لذا وابستگی پایپون شکل آنها همگن است. با اینحال پاسخ یک بازه از فیلترهای خطی به تکه‌های تصویری توسط ناهمگنی در درجه وابستگی آماری آنها مشخص می‌شود. در مقاله‌ای (Wainwright & Simoncelli 2000) یک چینش از پیش مشخص شده سلسله مراتبی 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 صحنه‌ها، باید چینش سلسله مراتبی از روی مثالها یاد گرفته شود. به علاوه از آنجایی که اشیا مختلف وابستگی‌های خاص خود را القا می‌کنند، برای تکه‌های مختلف تصویر پیشنهای مختلفی مناسب خواهد بود. مثلاً برای یک جفت فیلتر، قدرت وابستگی شرطی توأم ممکن است برای تکه‌های مختلف تصویر فرق کند. (شکل ۳-۲۵). این امر گویای اینست که بر پایه تکه‌های تصویر، همزنی‌های متفاوتی باید باشد فیلترهای مختلف انتساب شوند.

مدل‌های مولد GSM یک راه ساده برای پارامتری کردن وابستگی آماری واریانس فعالیت فیلترهای لایه اول در اختیار می‌گذارند به طوری که برخی از خصوصیت‌های کلیدی پایین به بالای آماری تصاویر را اخذ می‌کند. 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

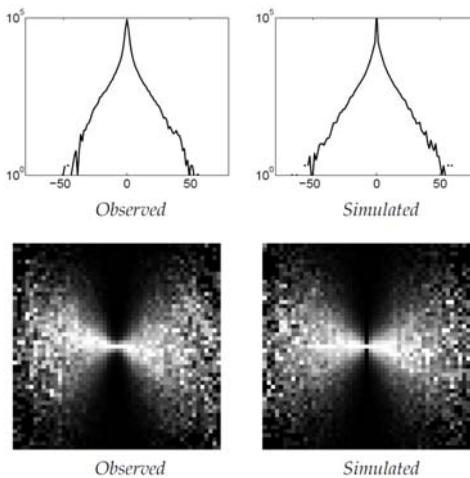
²⁰⁹ cortical gain control

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



شکل ۲۵-۳: آمارگان توان 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 شمارنامه هر ظرف هیستوگرام بدست می آید و هر ستون برای رسیدن به تمام بازه شدت روشنمایی مستقلانه نرمال می شود.

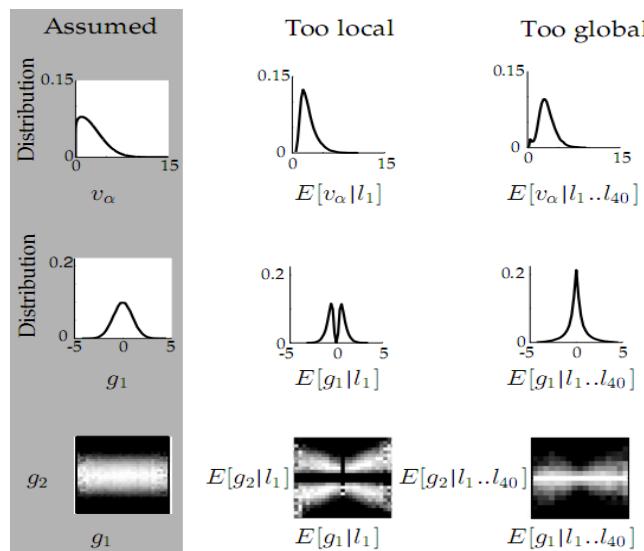
برای رفع این نقصها، توسعه هایی برای مدل GSM پیشنهاد شده است. برخی از مدل های جدید تصویر، ضرایب موجک درون هر همسایگی محلی را با استفاده از یک GSM توصیف می کنند که رفتار شدیداً زبرگویی توزیعهای حاشیه ای ضرایب موجک تصاویر طبیعی و نیز همبستگی آنها در بزرگی محلی را مدل می کند (شکل ۲۶-۳).



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

به عنوان مثال Baraniuk²¹⁰ و همکارانش یک متغیر همزن دو حالت 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 مدلهای GSM برای تصاویر با متغیرهای متغیر همزن ۷ 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) 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 (v_α, v_β, \dots) با فیلترهای خطی طوری خوش 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.

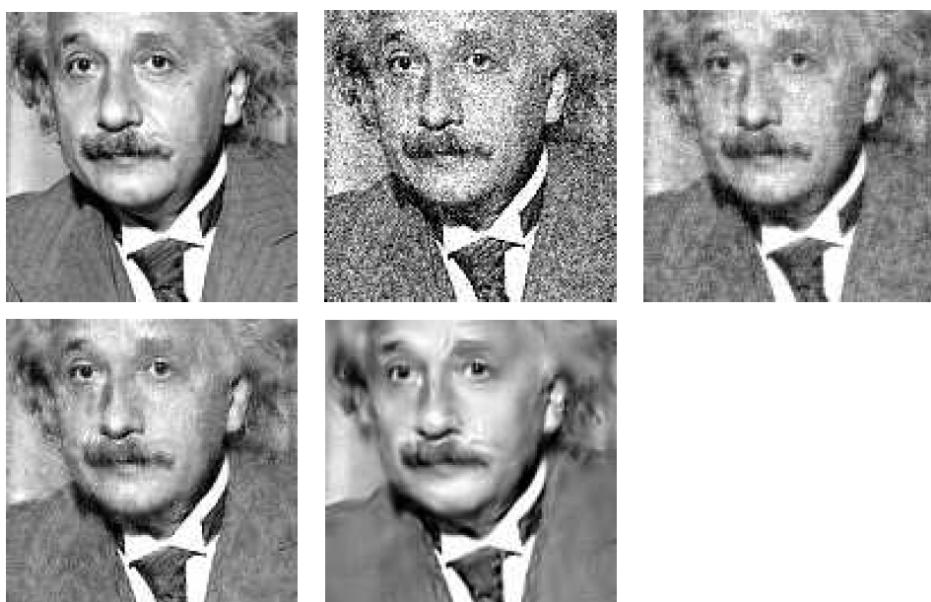
²¹⁰ Baraniuk

²¹¹ assignment problem

در این مقاله از تابعهای احتمالی رایلیخ دارای همپوشانی برای همزن‌ها استفاده شده است. روش‌های انتساب دیگری برای سایر مدل‌های تصویری استفاده شده است که از آن قبیل می‌توان به مدل‌سازی پویای درخت^{۲۱۲} (Williams & Hinton, Ghahramani, & Teh, 1999; Adams & Williams, 2003) و شبکه‌های اعتبار^{۲۱۳} (Adams, 1999; Adams & Williams, 2003) اشاره کرد. شکل ۲۷-۳ 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 ندارد) Schwartz (2006). استفاده از خانواده‌های 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 حاشیه‌ای موجک گفته شده بود پیچیده‌تر است ولی روی بازه گسترده‌ای از تصاویر و میزان نویز بهبود چشمگیری دارد (شکل ۲۸-۳). 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.



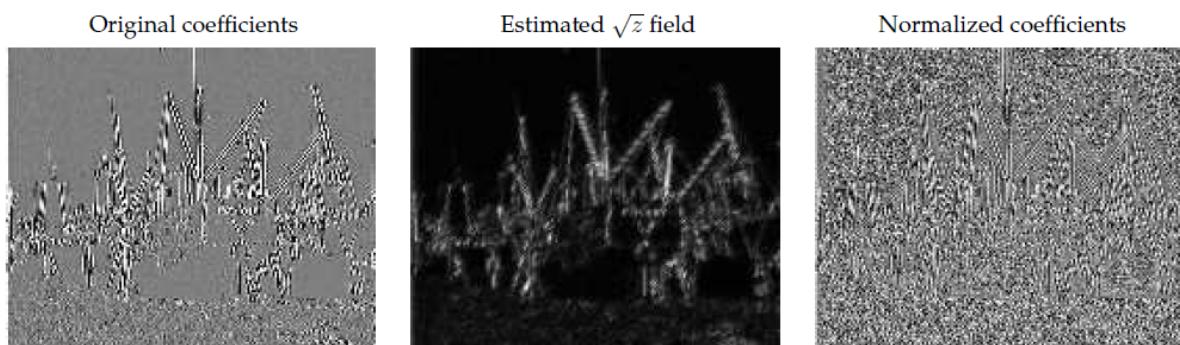
شکل ۲۸-۳: مثالهایی از رفع نویز. (بالا چپ) تصویر اصلی (انیشتین) (بالا وسط) تصویر آلوده شده به نویز سفید گوسی با $PSNR=21.45$ و نرخ قله سیگنال به نویز برابر $PSNR=22.06$. (بالا راست) رفع نویز با مدل حاشیه‌ای گوسی $PSNR=27.87$ (پایین چپ) رفع نویز با مدل حاشیه‌ای گوسی تعمیم یافته $PSNR=29.24$ (پایین راست) رفع نویز با مدل مخلوط مقیاس گوسی $PSNR=30.86$. تمام روش‌ها در یک حوزه فراکامل موجک پیاده سازی شده‌اند. در هر حالت واریانس نویز معلوم فرض شده‌اند و فرایند رفع نویز به کمک کمینه خطای مربعی بیزی انجام شده است (یعنی میانگین توزیع احتمال ثانویه). پارامترهای مدل با بیشینه سازی درستنمایی داده‌اند. 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 مدعای 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 ۲۹-۳ این فرایند را مشخص می‌کند و مثالی از زیرباند موجک، حوزه واریانس تخمین زده شده و ضرایب نرمال شده را نشان می‌دهد. با وجود حذف افزونگی‌های 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 می‌دهد. و دوم اینکه حوزه واریانس خودش هم از همگنی دور است زیرا بیشتر مقادیر مهم روی کانتورهای یک بُعدی متتمرکز شده اند (Simoncelli 2005).



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

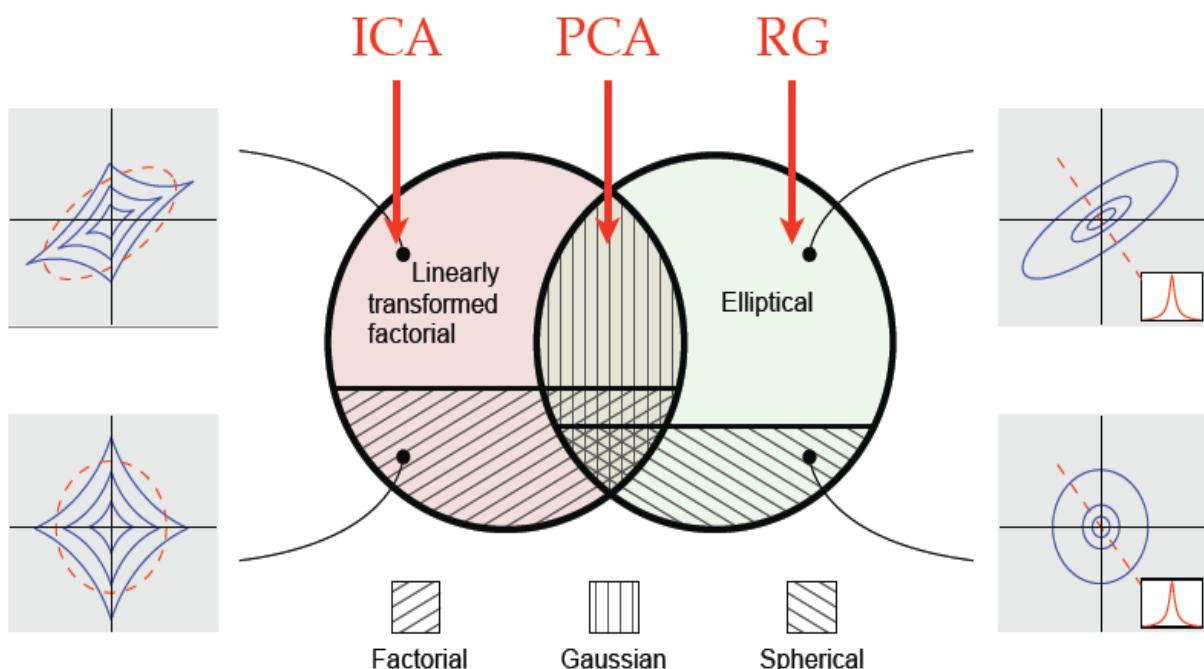
یکی از مشکلات دیگر از آنجا ناشی می‌شود که همسایگی یک فیلتر نه تنها باید توسط پهنهای مکانی آن مشخص شود، بلکه باید به مشخصات ویژگی‌های آن، محتوای تصویر و کلاس تصویر نیز توجه داشته باشد. به عنوان مثال برای تکه هایی با بافت عمودی در یک تصویر از لبه و یک تصویر از چهره باید همسایگهای متفاوتی برای فیلتر در 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).

۲-۴-۱-۳- روند پیشرفت

با گذشت نیم قرن از مدل گوسی‌افوریه، 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) مدل‌های آماری سازگار با واریانس مبتنی بر متغیرهای نهفته ناشی شده است. بهبودهایی که در اثر ظهور این ایده‌ها در کاربردهای پردازش تصویر حاصل شده است مداوم و بزرگ بوده است ولی پیاده کردن کامل این ایده‌ها و پالایش آنها همچنان در حال تکوین است.

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

داشت که این دو روش تنها زمانی حذف گارانتی ها را تضمین می کنند که مدل منبع آنها صحیح باشد (منبع جدایی پذیر با تبدیل خطی^{۲۱۵})؛ که البته در همین شرایط هم که به نظر می آید این متداول‌ترینها پاسخهای جالبی تولید می کنند، مؤلفه های بازنمایی تولید شده از استقلال آماری فاصله زیادی دارند. این شرایط در جای خود اینگونه نشان می دهد که تبدیلهای های 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 بانگذر کاسته می شود، ولی بازهم وابستگیهای مرتبه بالایی از خود نشان می دهد. حال اگر شرایطی در نظر گرفته شود که کاهش افزونگی برای کلاسهایی از مدلها منبع انجام شود که توزیعهای متقارن بیضوی^{۲۱۶} (EDS) نام دارند(شکل ۳۰-۳)، مشاهده می شود 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 ۳۰-۳: نمودار ون رابطه بین مدلها توزیع. دو دایره توزیعهای جدایی پذیر با تبدیلات خطی (آنگونه که ICA فرض می کند) و توزیعهای متقارن بیضوی(EDSها) را نشان می دهد. 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:

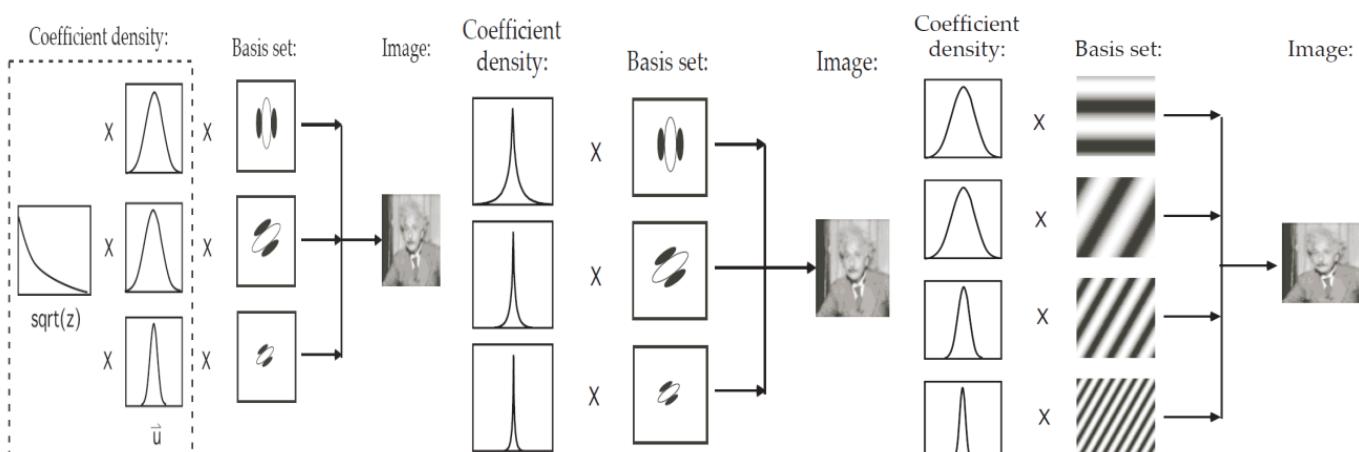
$$p(\vec{x}) = \frac{1}{\alpha |\Sigma|^{\frac{1}{2}}} f\left(-\frac{1}{2} \vec{x}^T \Sigma^{-1} \vec{x}\right) \quad (16-3)$$

²¹⁵ Linearly transformed factorial source

²¹⁶ elliptically symmetric densities

که در آن \mathcal{L} یک ماتریس مثبت-معین باشد، $(\int_0^\infty f(-r^2/2)r^{d-1}dr < \infty)$ و $f(\cdot) \geq 0$ را ارض 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 \mathcal{L} تعریف می کند. در حالت خاص وقتی که \mathcal{L} مضری از ماتریس همانی باشد، محدوده های توزیع آبرکره هستند و توزیع متقارن کروی^{۲۱۷} (SSD) نامیده می شود. اگر \mathcal{L} دارای آمارگان مرتبه دوم متناهی باشد، \mathcal{L} ضریبی از ماتریس کواریانس آنست که به طور ضمنی بیانگر اینست که یک **ESD** با یک عملگر PCA یا واپتینینگ قابل تبدیل به SSD است. مقاله (Lyu 2008d) برای این کلاس از مدلها مربع تصویر، روش جایگزین غیرخطی برای حذف افزونگی های مراتب بالاتر پیشنهاد می کند و روش خود را گوسی سازی شعاعی^{۲۱۸} می نامد.

شکل ۳۱-۳ خلاصه ای از مدلها گوسی، مدلها پایه های خطی، و مدلها 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 مدلهاست. نخست تلاش برخی از محققان در بررسی روشها برای توصیف قاعده مندیها در حوزه واریانس محلی است. این روشها شامل حوزه های تصادفی مکانی^{۲۲۰} (Hyvarinen 2003, Karklin 2003, Hyvarinen 2000) و مدلها درخت شکل چندمقیاسه (Romberg 99, Wainwright 2001) است. بسیاری از ساختارها در حوزه واریانس ممکن است خاصیتهای شبیه ویژگیهای ناپیوسته مانند لبه ها، خطوط، و گوشه ها داشته باشد. درباره این ساختارها مطالعه 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 مندیهای هندسی که از امتداد کانتور یا مرزها بست می آید صورت گرفته است (Elder 2002, Giesler 2001, Hoyer 2002, Zhu 2003).



شکل ۳۱-۳: خلاصه ای از مدلها گوسی(راست)، مدلها پایه های خطی(وسط)، و مدلها 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 زیبیت بهبود خواهد داد.

۳-۱-۴- ساختار سلسله مراتبی

همانگونه که اشاره شد مغز انسان با یک ساختار سلسله مراتبی داده های ورودی را پردازش می کنند و هر لایه از لایه های قبلی اطلاعات بیشتری را استخراج می کند و سیگنالهای پردازش شده را به لایه های بعدی خود تحویل می دهد. در مدلهای پایین به بالا نیز این ساختار حفظ می شود و طراحان مدل سعی می کنند با طراحی تابع فعالیت نورونها، میدان گیرندگی آنها و شکل بهم بندی و اتصال نورونها، به پدیده های 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 با بالا برای استفاده از مزایای آنها، و جبران ضعفهایی از مدلها که پیشتر بیان شد، برخی از پژوهشگران دست به افزودن لایه هایی به مدلهای موجود زده اند و سازمان سلسله مراتبی را به خدمت گرفته اند. در ادامه به ایده هایی سلسله مراتبی پرداخته می شود که اخیراً جای خود را در ادبیات فهم تصاویر طبیعی، باز کرده اند و در توجیه های بالاتر قشر خاکستری مخ به توفيقهایی، دست یافته اند.

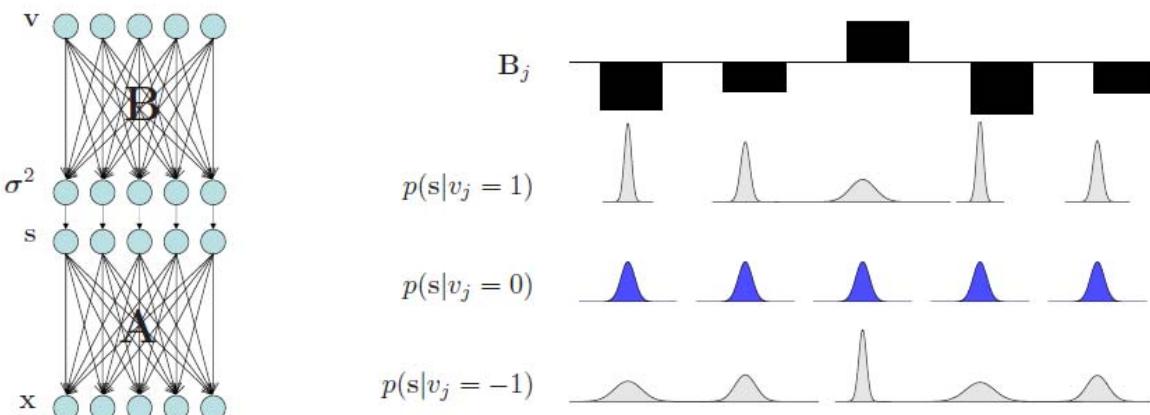
نخستین ایده ای که معرفی می شود، در حوزه ICA اضافه کردن لایه دومی به لایه اول فیلترها بود که واریانس فعالیتهای لایه اول را مشخص می کرد (Hyvärinen & Hoyer, 2000a, 2000b; Hyvärinen, 2002). کارکلین و لویکی (Karklin & Lewicki, 2003a, 2005; Park & Lee, 2005) مدلی را پیشنهاد کرده اند که در آن واریانس هر واحد در لایه اول از ترکیب جمع مجموعه ای از واحدهای 2003b, 2005 توابع پایه واریانس در لایه دوم تشکیل شده است. روش شوارتز برای مدل GSM (Schwartz, 2006) را می توان 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 کوواریانسی معرفی می شود که شالوده اصلی شبیه سازی نزونهای 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 مؤلفه های مستقل را گسترش دهد و گزینندگی نورون های پیچیده را شبیه سازی کند (Hyvärinen 2000). ایده اصلی در توسعه این مدل، وابستگی بین ضرایب مؤلفه های مستقل تصاویر طبیعی بود. در مدل ICA زیرفضا (Hyvärinen and Hoyer, 2000) ضرایب خطی دیگر مستقل فرض نمی شوند و بجای آن گروههایی از توابع پایه باهم همسایگی ها (زیرفضاهای) 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 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 پیشینه 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 بیابند.

در این مدل توابع پایه در لایه اول قرار دارند و نماینده‌گان 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 واریانسی: یکی از نقاط ضعف تحلیل مؤلفه های مستقل این است که مؤلفه های حاصل، محدود به خصوصیات نورون های قشر ابتدایی بینایی می باشد که خصوصیاتی نسبتاً ساده از تصاویر هستند. این روش قابلیت گسترش به خصوصیات سطح بالاتر را ندارد زیرا ضرایب این مؤلفه‌ها 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 ضرایب 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 ۳۲-۳). 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 ۳۲-۳: (چپ) یک بازنمایی گرافیکی از مدل سلسله مراتبی برای وابستگی واریانسی. واریانس ضرایب S تابعی از متغیرهای نهفته مرتبه بالاتر V هستند. پارامترهای مدل B تبدیلی خطی از V به σ^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 زامین ستون B بر توزیع توأم $p(s)$ با فرض آنکه تمام متغیرهای مرتبه بالای دیگر v_k ($k \neq j$) برای 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 ۱۱-۳) در نظر گرفته شده اند:

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

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:

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

در معادله فوق، B ماتریس توابع پایه واریانس و \vec{v} ضرایب این توابع را نشان می دهد. در این مدل، توزیع احتمال \vec{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 \vec{v} در معادله ۱۶-۳ می شود، ماتریس B است که ترکیب مؤلفه های واریانس را مشخص می کند.

برای تخمین ماتریس B از تحلیل درستنامایی 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:

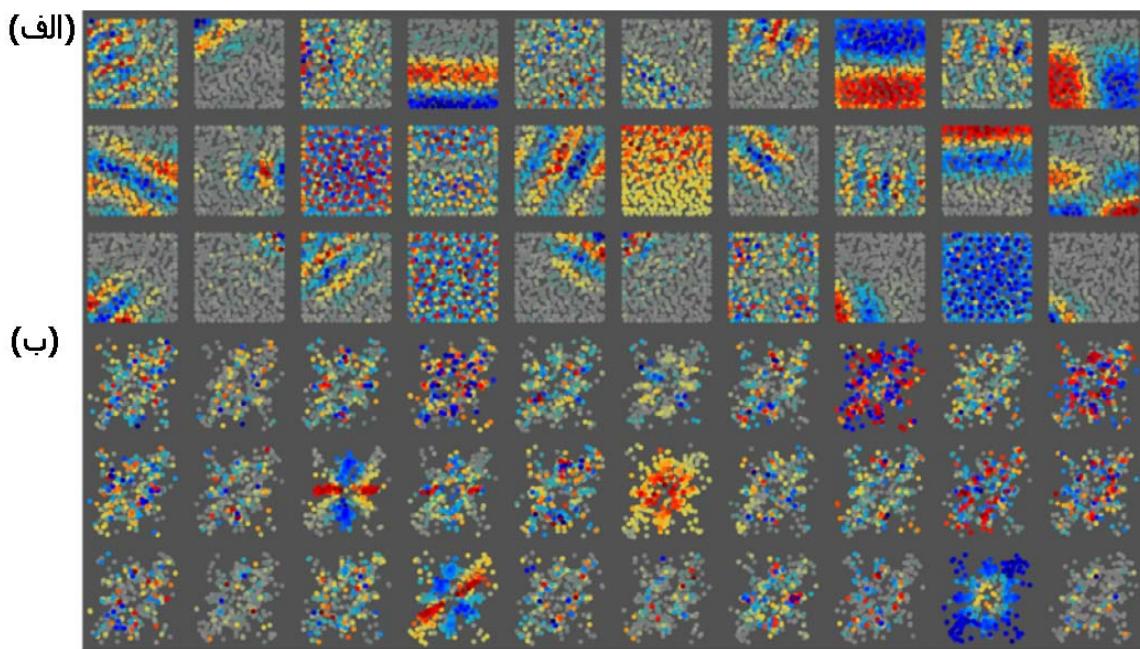
$$p(x_1, x_2, \dots, x_n) = \prod_i p(x_i | A, B) \quad (18-3)$$

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 ۳۳-۳ 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 مدلهای تحلیل تصاویر، مدلهای واریانسی (Karklin 2003a) و کواریانسی کارکلین (Karklin 2008) که در ادامه همین قسمت به آن اشاره خواهد شد) ساختارهای آماری غیرخطی را با استفاده از بازنمایی توزیع آنها کدگزاری می کند که این رهیافت احتمالاً در مغز نیز استفاده می شود. مدلهای پیشنهاد شده همچنین نسبت به روش‌های یادگیری بدون ناظر استاندارد تخمین توزیعی بهتری از تصاویر طبیعی بدست می دهند. با وجود آنکه داده های روی داده های

نورونی تطبیق داده نشده اند، تعدادی از خاصیتهای کلاسیک نورونهای ناحیه V1، و نیز اثرات ظرفی که در V2 و V4 دیده می شوند را بروز می دهند که یک توصیف کارکردی برای این رفتارهای نورونی را ارائه می کند.



شکل ۳-۳: مجموعه ای از مؤلفه های واریانس پایه که از تصاویر طبیعی استخراج شده اند. هر مؤلفه یک الگوی مشخص برای انحراف از شرط مستقل بودن را نشان می دهد. مؤلفه ها بر اساس (الف) موقعیت مکانی و (ب) جهت مؤلفه های مستقل متناظرشان مرتب شده اند.

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 می شود. در بین این توابع پایه، وابستگی واریانسی به عنوان تابعی از متغیر نهفته $\tilde{\tau}$ فرض شده است. لایه دوم این مدل متغیرهای نهفته و ارتباط آنها برای تولید تک تک واریانسها را مدل می کند. در این مدل محتمل ترین مجموعه متغیرهای $\tilde{\tau}$ و ماتریس مربوط به ترکیب آنها، B ، از استنتاج بدست می آید و سپس مؤلفه های مستقل از روی آنها در لایه اول ساخته می شود.

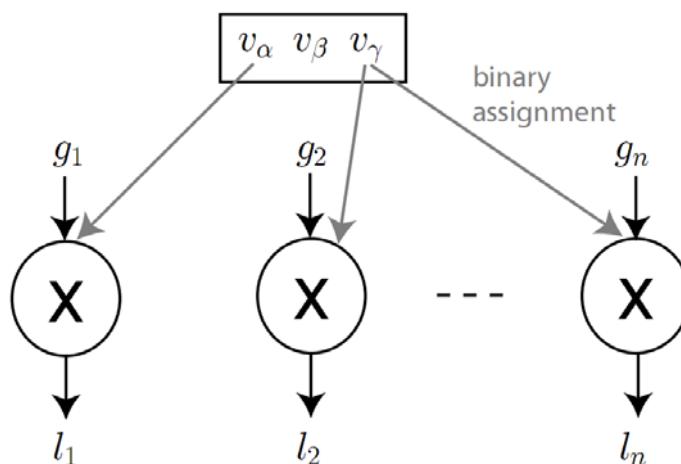
سلسله مراتب در مدل مخلوط مقیاسهای گوسی: یکی از مدل هایی که مدل کارکلین را گسترش داد، مقاله ای بود که در ۲۰۰۶ منتشر شد (Schwartz06). در این مقاله مسئله به این صورت مدل می شود: تعدادی متغیر همزن ($v_\alpha, v_\beta, \dots = \tilde{v}$) با فیلترهای خطی طوری خوش بندی شوند که هر چند فیلتر خطی، یک متغیر همزن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

²²¹ expectation maximization

پایان نامه کارشناسی ارشد کورش مشگی - پاییز ۱۳۸۹
دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر

در این مقاله، که بر مبنای مدل مؤلفه‌های مشترک واریانس می‌باشد، یک مجموعه از عوامل واریانس (متغیرهای همزن در GSM) بر لایه‌ای از ضرایب مؤلفه‌های مستقل (توزیعهای گوسی چند متغیره در GSM) اضافه شده‌اند. این مدل نیز بر اساس وابستگی واریانس بین پاسخ مؤلفه‌های مستقل می‌باشد. تفاوت این مدل با مدل کارکلین در نحوه انتساب مؤلفه‌های واریانس به ضرایب مؤلفه‌های مستقل می‌باشد. در مدل کارکلین، ضرایب واریانس با استفاده از ماتریس مؤلفه‌های واریانس با یکدیگر ترکیب می‌شوند تا واریانس ضرایب مؤلفه‌های مستقل را تعیین کنند ولی در مدل شوارتز، ضرایب واریانس با یکدیگر رقابت می‌کنند و برندۀ رقابت به مؤلفه‌ها در لایه‌ی پایین‌تر نسبت داده می‌شود.

به ازای هر ضریب واریانس و مجموعه‌ای از ضرایب مؤلفه‌های مستقل، یک مجموعه احتمال محاسبه می‌شود که احتمال نسبت دادن ضریب واریانس به مؤلفه‌های مستقل را نشان می‌دهد. پس از محاسبه احتمال به ازای تمام ضرایب واریانس، ضریب با بیشترین مقدار احتمال به هر ضریب مؤلفه مستقل نسبت داده می‌شود. شکل ۳-۴-۳ این مدل را شرح می‌دهد.



شکل ۳-۴-۳: مدل انتساب احتمالی یک مدل GSM چند بعدی. پاسخ فیلترهای $\{l_1, \dots, l_n\}$ از ضرب متغیرهای گوسی $g = \{g_1, \dots, g_n\}$ در متغیرهای $\{v_\alpha, \dots, v_\gamma\}$ که فرض می‌شود که $n < m$. می‌توان هر مخلوط l_i را به عنوان پاسخ فیلتر خطی وقتی که روی یک تکه تصویر خاص اعمال می‌شود در نظر گرفت. مسئله انتساب برای تولید پاسخ فیلتر l_i می‌پرسد که کدام متغیر همزن v_j باید به کدام متغیر گوسی g_j نسبت داده شود. مجموعه متغیرهای همزن $v_\alpha, v_\beta, v_\gamma$ در مستطیل سیاه رنگ مشخص شده است. پیکانهای خاکستری رنگ انتساب بازتری را نشان می‌دهند. با همزن v_α تولید می‌شود و با همزن v_β و با همزن v_γ تولید می‌شوند.

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

²²² Markov chain Monte Carlo sampling

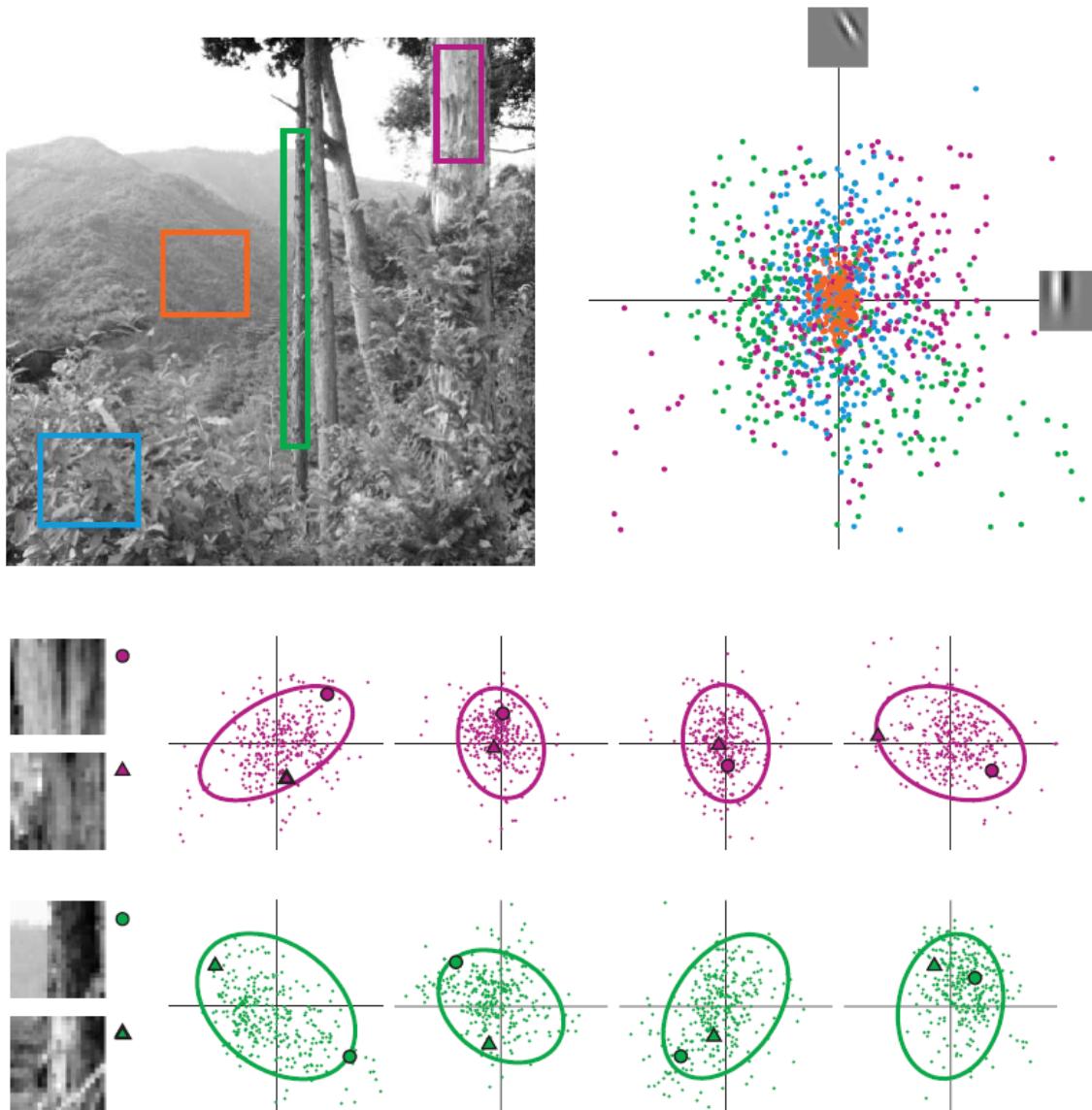
سلسله مراتب در مدل وابستگیهای کواریانسی: هنگامی که تصویری از یک صحنه پیچیده طبیعی را نگاه می کنیم، در چندین مکان مختلف (مثلاً تنه یک درخت و یا در لبه های آن) به نقاط تمرکز^{۲۲۳} خاصی بر می خوریم که حتی با وجود اینکه تک تک تصاویری که روی شبکیه می افتد ذاتاً دارای تنوع زیادی است، درکی منطقی از ساختار زیرین (بافت پوست یا کانتور لبه درخت) بدست می دهد. شکل ۳۵-۳ مسئله ای را که مغز ما به راحتی حل کرده است را شرح می دهد: نواحی تصویر که از لحاظ ادراکی متمایزند الگوهای پاسخی را می سازند که همپوشانی زیادی دارند و نمی توان 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 کشف کننده های خطی بینایی توصیف می کنند (Hubel 62, Movshon 78-35, Chichilinsky 01) و مدلهایی که حول این ایده تولید شده اند با دقت بالایی می توانند رفتار نورونها را، از شبکیه (Carandini 97) تا سلولهای ساده در قشر بینایی مخ (Hubel 62) اخذ کنند. ولی همانگونه که شکل ۳۵ نشان می دهد، نه تک تک ویژگیها و نه تبدیلات خطی نمی توانند تصاویری با یک ساختار را از تصاویر دیگر جدا کنند. در لایه های بعد سیستم بینایی احتمالاً ویژگیهای مجردتری محاسبه می شوند ولی دانش ما در این مورد فعلاً محدود است. در لایه V1، سلولهای پیچیده در مکانهای مختلفی به لبه پاسخ می دهند (Hubel 62) ولی مدلهای کلاسیک این سلولها (Movshon 78b, Adelson 85) از تشریح برخی از پدیده های غیرخطی مشاهده شده در این لایه، مانند بازنشانی اطرافیان و بازداری فعالیت به ازای راستهای متقاطع عاجزند (Bonds 89, Jones 2002, Cavanagh 2002). از آن مهمتر اینست که هیچ توجیه کارکردی برای نقش این رفتارها در ادراک صحنه های طبیعی نیز بدست نمی دهنند. در لایه های بالاتر بینایی نظیر V2 و V4، نورونها نسبت به خصوصیات تصویر مانند مکان و مقیاس بیشتر نامتغیرند (Kabotake 96, Gallant 96, Connor 2007, Hedge 2000). برای تعمیم کارای این نورونها، مدار نورونی باید یک بازنمایی تولید که از یکسو برای پراکندگی وسیع صحنه های طبیعی از یک نوع (مثلاً بافت یا کانتور) مشابه باشد و از سوی دیگر برای پراکندگی وسیعتر تمام صحنه های طبیعی دیگر متمایز باشد. کارهای نظری پیشین نشان داده است که نورونها در بخش اولیه قشر خاکستری مخ یک کد بهینه تشکیل می دهند که با آمارگان تصاویر طبیعی سازگار است (Olshausen 96, Bell 97) اما هیچ اطلاعاتی درباره این نمی دهد که چگونه نورونها روی تنوع ذاتی این عناصر تصویر تعمیم می یابند.

روش حاضر توسعه ای بر روشن کدگزاری کاراست و پیشنهاد می دهد که بازنمایی توزیعهای بیشمار آمارگانی که نواحی محلی تصویر را معین می کنند، وجه مهمی از محاسبات بصری است. نورونهای لایه های بعد، به جای اینکه شدت روشنایی پیکسلهای یک تکه تصویر از بافت یا لبه را کدگزاری کنند، توزیع تصویر (یعنی بازه و الگوی تنوع شدت روشنایی پیکسلها یا ویژگیهای تصویر) را که با تصویر ورودی همخوانی بیشتری دارد کدگزاری می کند. این نحوه کدگزاری باعث می شود 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 می شود و برای صحنه های طبیعی بهینه سازی می شود، خواص غیرخطی سلولهای پیچیده و نورونهای لایه های بالاتر را نیز در بر می گیرد و لذا یک تفسیر کارکردی از این پدیده ها را ارائه می کند.

²²³ fixation

پایان نامه کارشناسی ارشد کورش مشگی - پاییز ۱۳۸۹
دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر



شکل ۳-۳: الگوهای آماری، نواحی مختلف صحنه‌های طبیعی را تمایز می‌کند. (بالا چپ) یک صحنه طبیعی با چهار ناحیه مشخص شده (برگرفته از (Doi 2005)). (بالا راست) نمودار پراکندگی نشان دهنده الگوی پراکندگی زوجهای یابنده‌های فیلتر خطی (فیلترهای خطی راستدار) برای ۲۰۳۲۰ تکه تصویر نمونه گیری شده از چهار ناحیه مشخص می‌باشد. 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 یابندهای خطی هستند که به نواحی که لبه درخت و پوست درخت قرار دارند مربوطند (ولین ستون همان ویژگیهایی هستند که در تصویر بالا-راست آورده شده‌اند). همبستگی در هر قسمت می‌تواند با یک توزیع گوسی و کواریانس آن توصیف شود (بیضی‌ها). تفاوت در توزیعهای بین سطراها مشخصات الگوهای درون همبستگیها را آشکار می‌کند که وقتی ویژگیهای بیشتری مد نظر قرار بگیرند واضح‌تر نیز می‌شود. این الگوها می‌توانند برای تعمیم درون نواحی استفاده شوند، در حالی که همچنان بتوانند آنها را تمایز بدهنند. به عنوان مثال در هر ناحیه دو تکه تصویر مشخص شده و با مثلث و دائره نمایش داده شده است. با وجود آنکه جفت‌های تصویر بهوضوح تفاوت زیادی دارند، هر تصویر با توزیع ناحیه تصویر خود سازگار است. با بررسی مقابله‌ای توزیع در ابعاد مختلف، استنتاج نوع تکه‌های تصویر ممکن است، حتی با وجود این امر که تکه‌های 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 ۳۵-۳ آمده است) و آنها را از روی تکه های منفرد تصویر تشخیص دهد. حال باید پرسید که کدام قاعده مندیهای آماری برای اینکار مناسب است؟ همانطور که شکل ۳۵-۳ القا می کند، توزیعهای تصاویری که از لحاظ مفهومی مشابهند، الگوهای سازگاری را با درجه ای از تنوع در یک بعد 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 بعد، همانطور که در مثالها دیده می شود، به نظر ماهرانه و زیرکانه می رسد، توزیع کامل چند متغیره با صدها بعد 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 کواریانسها مشاهده شده و انتساب یک نورون به هریک 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.

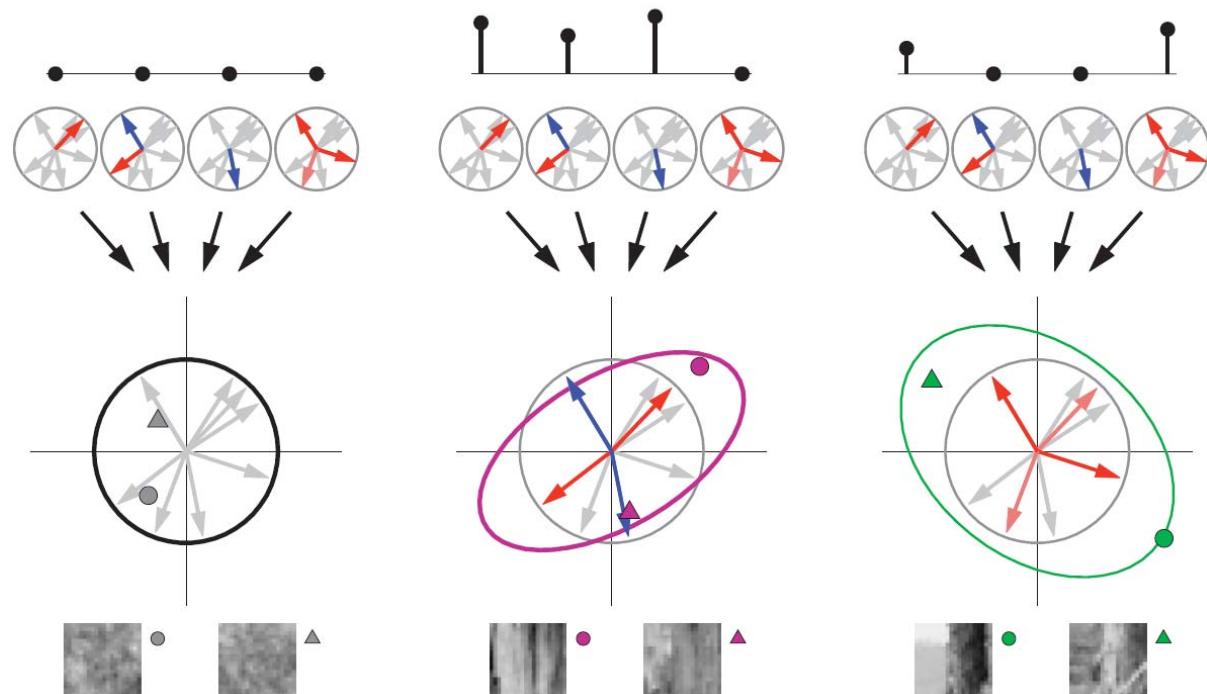
این مدل کدگزاری توزیع به صورت شماتیک در شکل ۳۶-۳ آورده شده است. این مدل همبستگی موجود در نواحی محلی تصویر را با یک توزیع گوسی چند متغیره دارای میانگین ثابت صفر و کواریانسی که تابع فعالیت نورنهاست بازنمایی می کند. این تعریف ساده آماری از سویی انعطاف پذیری در اخذ زنجیره توزیعهای تصاویر طبیعی و از سوی دیگر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

سری اول بردارهای \bar{b}_k (پیکانهای درون دایره خاکستری) هستند که ویژگیهایی از تصویر را مشخص می کند که توزیع تصویر در راستای آنها می توان از توزیع استاندارد²²⁴ (دایره سیاه رنگ) کشیده تر یا فشرده تر باشد. این بردارها بین تمام نورونهای این مدل (که با چهار دایره خاکستری که مجموعه پیکانهای 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 بعد باشد.

نورونهای مدل، زله، 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 \bar{b}_k است. وزن مشبت (قرمز رنگ) بدین معناست که نورونها به حرکتی بیشتری در آن راستا پاسخ می دهند، وزن منفی (آبی رنگ) به معنای بازه محدودتری از حرکه است، 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 کدگزاری 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 (مثالاً برای جفت تکه های تصویر شکل ۳۵-۳).



شکل ۳-۳: مدل کدگزاری توزیع. بجای کدگزاری مقادیر دقیق پیکسل تصویر ورودی(پایین)، مدلهای 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 ۳۵-۳) در هر ستون الگوی فعالسازی برای هر دو ورودی یکسان است.

با تطبیق پارامترهای \tilde{b}_k و w_{jk} مدل به داده، می توان کارترین 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 محلی بودن صورت نگرفته است (کدگزاری برای هر تکه تصویر مستقلًا صورت گرفته است) و این وظیفه مدل است که یک بازنمایی فشرده را از تمام تکه ها یادگیری کند و خودخود کشف کند که کدام بازنمایی خواص آماری از یک نوع را به اشتراک می گذارد.

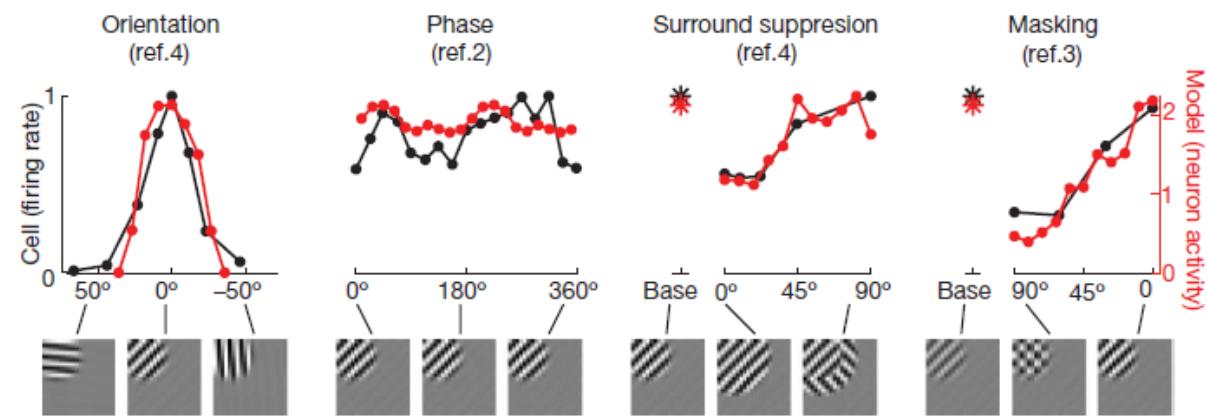
اگر طبق فرضیه ها نورونهای بخش بینایی قشر خاکستری مخ الگوهای همبستگیهای نواحی محلی را کدگزاری کند و خصوصاً با آمارگان صحنه های طبیعی سازگار شده باشد، انتظار می رود که بازنمایی یادگیری شده توسط مدل، خواص

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 متدائل در شکل توزیع تصویر را کدگزاری می کنند. همانند تکه های تصویر، هر یک از آنها مانند یک ویژگی لبه مانند جهتدار و محلی هستند. مجموعه کامل آنها پهنهای یک تکه تصویر را فرش می کند و تمام بازه راستها و فرکانسهای مکانی تصاویر طبیعی را می پیماید 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 قبلًا بازنمایی های مشابهی با استفاده از مدلهای آمارگان خطی که کارایی کدگزاری تصویر را بهینه می کردند بدست آمده بود (Olshausen 96, Bell 97). در مدلی که در اینجا به آن اشاره شد، با وجود اینکه ویژگیها صریحاً برای بازسازی تصویر استفاده نمی شود ولی در عوض به بهبود توزیع کدشده کمک می کند (پیکانها در شکل ۳-۳) بنابراین در حالیکه تفسیرهای سنتی از کدگزاری اولیه حسگری اینگونه است که باید بازسازی مناسبی از محرک را بدست دهد، نتیجه این روش تفسیری به آن اضافه می کند: این کدگزاری باید تنوع در توزیعهای تصویر را نیز منتقل کند و به قسمتهای بینایی که در مسیر داده ها پس از دیگری قرار گرفته اند امکان ارائه بازنمایی مجردتری را می دهد.

دومین مجموعه پارامترها، وزنهای w_{jk} هستند که نقش هر نورون را در شکل دادن به توزیع تصویر کدشده مشخص می کنند. به جز پاسخ دادن به تعداد محدودی از ویژگیهای تصویری برانگیخته کننده و یا فرونشاننده، 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 کدگزاری 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 فعل می شود 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 می شود، پنجره های سینوسی، آزمود. پارامترهای مدل پس از آموزش با تصاویر طبیعی ثابت می شوند و پاسخ نورونی روی مجموعه ای از الگوهایی که مرکز آنها در محدوده بینایی است و پاسخ بیشینه را بدست داده اند محاسبه می شود. این نورون خاص مدل خواص متنوعی از خود نشان می دهد که در نورونهای پیچیده V1 سلولهای V2 وجود دارند؛ خواص نظری نامتغیری به فاز، تنظیم با راستا و ویژگیهای فرونشانی پیچیده (شکل ۳-۷). زیرمجموعه بزرگی از این جمعیت خواص مشابهی را نشان می هند، در حالیکه برخی دیگر از آنها الگوهایی پیچیده تر (مانند الگوهایی که در لایه های بالاتر بینایی مانند V2 و V4 مشاهده شده اند) را کدگزاری می کنند. باید تاکید کرد که ویژگیهای تصویری که در نورونها شکل گرفته است بدون هیچ پیش فرضی درباره ساختار تصویری که توسط نورونها کدگزاری شده اند و نیز بدون تطبیق دادن یک مدل بر داده های آزمایشات فیزیولوژی بدست آمده اند. خصوصاً اینکه این مدل ویژگیهای تصویری کدگزاری شده را محدود به محلی بودن و راستادار بودن نمی کند و از پیش نیز چگونگی ترکیب زیرواحدها در الگوهایی که هر نورون نشان می دهد را نیز تعیین نمی کند.

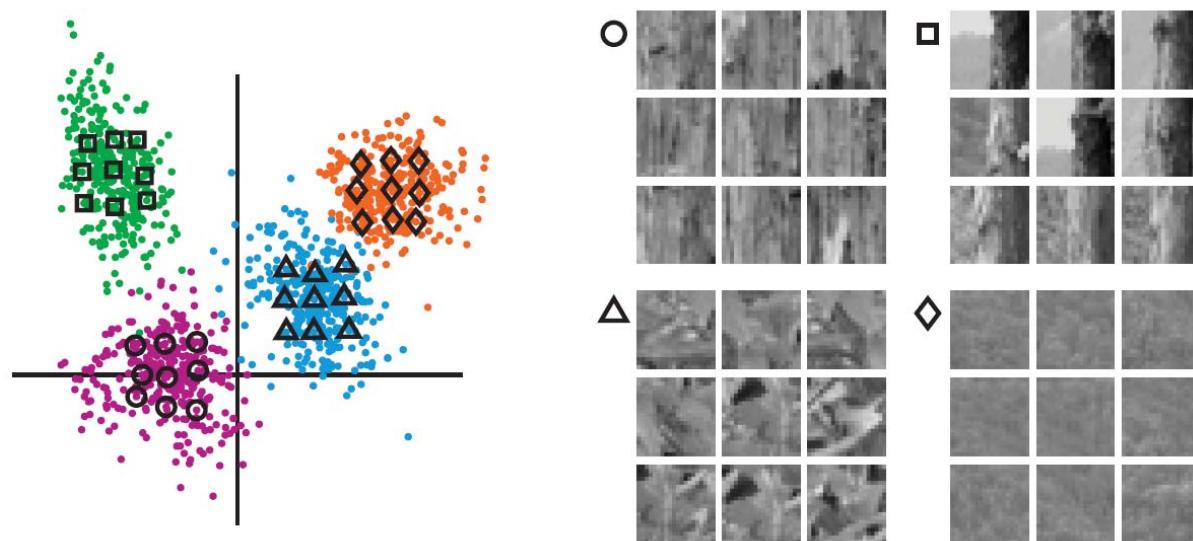


شکل ۳-۳۷: هنگامی که پنجره های سینوسی به نورون مدل داده شود، از خود وجوه مشترک زیادی با سلولهای پیچیده V1 نشان می دهد. این پاسخ با راستای پنجره بخوبی تنظیم می شود ولی نسبت به فاز آن نامتغیر است. افزودن پنجره به نواحی اطراف (سومین نمودار از سمت چپ، ${}^{\circ}$) باعث فرونشانی نسبت به پاسخ مینا به یک پنجره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

حال به روی که مدل از جمعیت نورونها استفاده می کند تا تصویر را بازنمایی کند پرداخته می شود. اگر مدل بتواند روی تنوع زیادی که در تصاویر طبیعی وجود دارد تعیین 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 نگاشت بازنمایی مدل از مجموعه تصاویر روی دو بعد (همانند فضای تصویر در شکل ۳-۳۵) نشان داد که اینکار در شکل ۳-۳۸ انجام شده است. همانگونه که انتظار می رفت، با کدگزاری توزیع تصویر بجای کدگزاری ویژگیهای خاصی درون یک تصویر، نورونهای مدل توانسته اند که تصویری 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 سرراستی برای کدگزاری نورونی بدست نمی دهد. با این وجود می توان استنتاج درون مدل را با یک محاسبه سر راست تخمین زد: یک نورون مربعات پاسخهای تعداد زیادی از ویژگیهای b_k تصویر را جمع می کند و الگو را با وزنهای w_{jk} همبسته می کند. این محاسبات را می توان به چشم یک تعیین 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 نورونهای 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 تجمعی 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 نورونهای مدل الگوهای مکانی پیچیده تری را تولید می کرند که پاسخ نورونی 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) خوش‌هایی را نسان می‌هدد 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 غیرخطی ای که توسط نورونهای مدل بروز می‌کند (شکل ۳-۳۷) قبلاً نیز در مدل‌های سلوهای پیچیده گنجانده شده بودند (Cavanagh 1996, Carandini 1997, Heeger 1992, Cavanagh 2002, Carandini 1997, Heeger 1992, Cadieu 2007) و بنابراین تفسیری کارکرده 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), یا نامتغیری به ازای موقعیت و مقیاس (Riesenhuber 1999) بوده‌اند. با اینحال این مدلها نیز صرحتاً به مسئله تعمیم نبرداخته‌اند که در مقاله (Karklin 2008) توسط استنتاج توزیع آماری که متحمل ترین توصیف تصویر را بدست می‌دهد، این هدف تامین شده است. یک مزیت مهم این روش اینست که بجای فرض کردن نامتغیری (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 معکس کند. بنابراین این مدل یک راه کمی برای تفسیر پاسخ نورونها به محرکهای پیشرفت‌ه است که به قاعده مندی آماری آنها باز می‌گردد.

۲-۳- مروری بر روشهای آشکارسازی چهره

۱-۲-۳- مقدمه

تحقیقات پر شتاب در زمینه پردازش چهره بر مبنای این فرض است که اطلاعات گرانبهایی مانند هویت کاربر، وضعیت و منظور او از تصاویر قبل 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 قبلاً 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. به 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)، 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 چهره‌ای در تصویر وجود دارد یا خیر، و اگر وجود دارد مکان و گستردگی هر چهره را بازگردد (Yang 2002). چالش‌های مرتبط با تشخیص چهره می تواند به دسته‌های زیر تقسیم شود:

وضعیت: تصاویر چهره‌ها بدلیل وضعیت نسبی چهره-دوربین (робو، نیم‌رخ، بالا به پایین) تفاوت می‌کند، و برخی از ویژگی‌های چهره مانند یک چشم یا بینی ممکن است قسمتی یا تماماً پوشانده شوند.

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

حالات چهره: ظاهر چهره‌ها 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 چهره‌ها ممکن است توسط چهره‌ها دیگر مسدود شده باشد.

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

شرایط تصویربرداری: هنگامی که تصویر تشکیل می‌شود، عواملی مانند نور (طیف، توزیع منبع و شدت آن) و شاخص‌های دوربین (پاسخ سنسور، عدسی‌ها) می‌تواند روی ظاهر یک چهره موثر باشد.

²²⁵ occlusion

مسائل متعددی موجود است که قربات زیادی با آشکارسازی چهره دارد. مکانیابی چهره^{۲۲۶} سعی می‌کند 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، حفره‌های بینی، ابرو، دهان، لبها، گوشها، و غیره با فرض وجود تنها یک چهره در تصویر است. شناسایی/تعیین هویت چهره^{۲۲۸} تصویر ورودی (پروب) را در یک پایگاه داده‌ها (گالری) جستجو می‌کند و تطابقهای احتمالی را گزارش می‌کند. هدف تصدیق هویت چهره^{۲۲۹}، تصدیق هویت ادعا شده فردی است که در تصویر حضور دارد. روش‌های ردگیری چهره^{۲۳۰} دائمًاً مکان و احتمالاً راستای یک چهره را در دنباله تصاویر به صورت بلادرنگ تخمین می‌زنند. تشخیص حالت چهره^{۲۳۱} به شناسایی حالت‌های درونی انسان (مثلاً ناراحت، خوشحال، منزجر،...) می‌پردازد.

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 کرده‌اند و نتایج بسیار خوبی را عرضه داشته‌اند.

یک مسئله مربوط و مهم اینست که کارایی یک روش چگونه ارزیابی می‌شود. 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 می‌توان به زمان یادگیری، زمان اجرا، تعداد نمونه‌های 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 می‌شود. در این پژوهش نرخ تشخیص عبارتند از: نسبت بین تعداد چهره‌هایی که به درستی آشکارسازی شده‌اند و تعداد چهره‌هایی که توسط انسان تعیین شده است. یک ناحیه از تصویر که به عنوان چهره شناخته شده است، درست محسوب می‌شود اگر ناحیه تصویر بیش از کسر خاصی از چهره را در تصویر پوشش دهد. به طور کلی تشخیص دهنده‌ها دو دسته خطا دارند: اشتباه‌منفی^{۲۳۲} که در آن چهره‌ها از قلم می‌افتد و در نتیجه نرخ تشخیص تنزل پیدا می‌کند و اشتباه‌ثبت^{۲۳۳} که در آن ناحیه از تصویر چهره در نظر گرفته می‌شود در حالی که اینچنین نیست. یک ارزیابی عدلانه باید این دو عامل را به حساب آورد زیرا با تنظیم پارامترهای روشی می‌توان نرخ تشخیص را افزایش داد در حالی که تشخیص‌های اشتباه نیز افزایش می‌یابد.

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

تشخیص چهره می‌تواند به صورت یک مسئله شناسایی دو کلاسه دیده شود که در آن یک ناحیه تصویر به یکی از دو دسته «چهره» و «غیر چهره» تشخیص می‌یابد. طبعاً، آشکارسازی چهره یکی از محدود تلاشهایی است که سعی می‌کند از خود تصویر و نه بازنمایی تجربیدی آن یک کلاس از اشیا را تشخیص دهد که خود دارای تنوع بین کلاسی بالایی است (که قبلًا اشاره شد). همچنین یکی از محدود کلاسهایی از اشیا است که در آن تنوع توسط مجموعه داده‌های آموزشی بزرگ پوشش داده شده است و بنابراین برخی از روش‌های تشخیص قابل اعمال به طیف گسترده‌تری از مسائل شناسایی هستند. آشکارسازی چهره همچنین چالش‌های جالبی به مسئله شناسایی الگو و تکنیک‌های یادگیری ذیل خود می‌افزاید. هنگامی که یک تصویر خام یا فیلتر شده به عنوان ورودی به دسته بندی کننده اعمال می‌شوند، ابعاد فضای ویژگیها بسیار بزرگ است (یعنی همان تعداد پیکسلها در داده‌های آموزشی نرمال شده). کلاسهای چهره و غیر چهره با توابع توزیع چندقله‌ای بیان می‌شوند و مرزهای تصمیم‌گیری احتمالاً غیرخطی هستند. برای کارایی، هر دسته بندی کننده باید بتواند از تعداد محدودی داده آموزشی یاد بگیرد یا وقتی با این داده‌های آموزشی ابعاد بالا کار می‌کند 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 می‌گیرد. روش‌های یاد شده را می‌توان در چهار دسته قرار داد. برخی از روش‌ها نیز کاملاً از روی مرزهای این دسته‌ها عبور می‌کند 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 می‌دهد،

درون یک پایگاه قوانین کدگذاری می‌کند. عموماً این قوانین روابط بین ویژگیهای چهره را مد نظر قرار می‌دهد. هدف عملده 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 عمدها برای مکانیابی چهره استفاده می‌شوند.

۳. **روش‌های تطبیق کلیše^{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 می‌شود و مورد استفاده روال آشکارسازی قرار می‌گیرد. این روشها هم برای مکانیابی و هم برای تشخیص چهره استفاده می‌شود.

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 کلیشه‌ها) از روی مجموعه تصاویر ورودی یاد گرفته می‌شود که باید تنوع نمایش‌های گوناگون ظاهر چهره را اخذ کند. این مدل‌های یادگیری شده عموماً برای آشکارسازی چهره استفاده می‌شود.

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

۲-۲-۳- روش‌های مبتنی بر دانش

در این گرایش، روش‌های آشکارسازی چهره بر اساس قوانینی انجام می‌پذیرد که از دانش پژوهنده از چهره افراد نشأت می‌گیرد. توصیف ویژگیهای چهره و روابط آنها با قوانین ساده کار آسانی است. به عنوان مثال یک چهره معمولاً^{۲۳۸} با دو چشم متقاضن، یک بینی، و یک دهان در تصاویر ظاهر می‌شود. ارتباط بین این ویژگیها می‌تواند با فاصله و موقعیت نسبی آنها بیان شود. نخست ویژگیهای چهره در تصویر باید استخراج شود، سپس چهره‌های کاندید بر اساس قوانین وضع شده انتخاب می‌شوند. یک فرایند تصدیق^{۲۳۹} نیز معمولاً برای کاهش تشخیص‌های نادرست بکار گرفته می‌شود.

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 حالات ممکن چالشی بزرگ است. لیکن باید در نظر داشت که الگوریتمهای اکتشافی^{۲۴۰} درباره چهره برای تشخیص چهره تمام رخ^{۲۴۱} در پس زمینه‌های ساده (غیر درهم ریخته) بخوبی عمل می‌کند.

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

²³⁷ appearanc-based methods

²³⁸ verification

²³⁹ heuristics

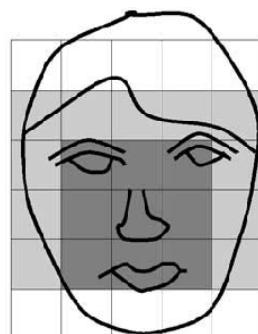
²⁴⁰ frontal face

²⁴¹ uncluttered



شکل ۳-۳: (چپ) $n = 4$ تصویر اصلی (دوم) $n = 8$ (سوم) $n = 16$ (راست) 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 دقتها پایینتر بکار گرفته می‌شوند عبارتند از: (الف) بخش مرکزی صورت (که در تصویر ۳۹-۳ تیره تر سایه خورده است) دارای چهار سلول با شدت روشنایی یکنواخت است. (ب) بخش گرد بالای صورت (که در تصویر ۴۰-۳ 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 ۴-۳: 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 می‌شوند.

جستجو بدنال کاندیدهای چهره بودن در تصویر دارای کمترین دقت(سطح یک) صورت می‌گیرد و آنها برای پردازش بیشتر در دقتها بهتر استفاده می‌شوند. در سطح دوم، یکنواخت سازی هیستوگرام محلی^{۲۴۲} روی کاندیدهای چهره بودنی که به سطح دوم آمده اند صورت می‌گیرد و سپس لبه یابی روی آن انجام می‌شود. نواحی کاندیدهای باقیمانده در سطح سوم با مجموعه قوانین دیگری که به ویژگیهای چهره مانند چشمها و دهان پاسخ می‌دهند آزموده می‌شوند. 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 نمی‌دهد (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

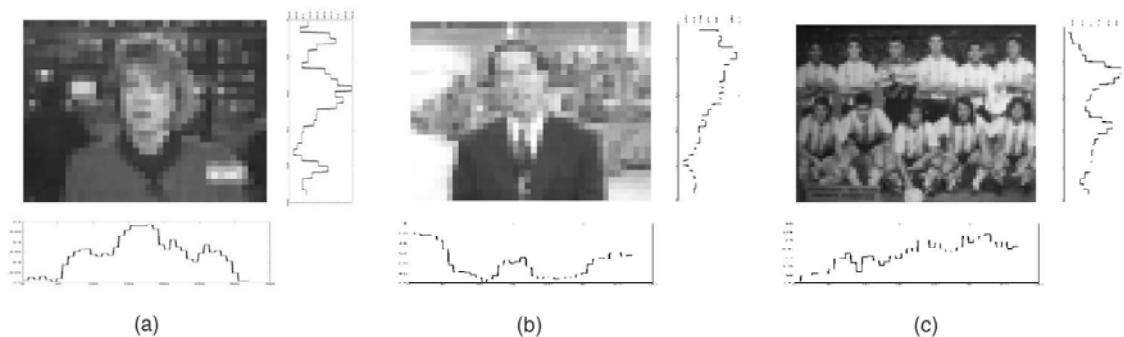
²⁴² local histogram equalization

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 ۳-۴۰-۴۱ چپ یک مثال را نشان می دهد 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 ۳-۴۱-۴۲: (چپ) و (وسط) $n = 8$ (راست) $n = 4$ نماهای افقی و عمودی. تشخیص یک چهره با جستجو بدنیال 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 ۳-۴۱-۴۳ 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 شکل ۳-۴۱-۴۲ 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 ویژگیهای 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 می‌گیرند. محدوده ویژگیها می‌تواند برای چهره‌ها ضعیف شود در حالی که سایه‌ها می‌توانند منجر به لبه‌های قوی‌ای شوند که با وجود این مسائل الگوریتمهای دسته‌بندی مفهومی بی‌استفاده خواهند بود.

۳-۲-۱-۳- ویژگی‌های 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 لبه‌ها (لبه‌یاب کنی) و الگوریتم اکتشافی برای حذف و گروه بندی لبه‌ها استفاده می‌کند تا در نهایت تنها لبه‌های روی کانتور چهره باقی بماند. سپس یک بیضی روی مرز سر و پس زمینه برآش می‌شود. این روش به دقت ۸۰ درصد دست یافته است.

در مقاله‌ای دیگر بجای استفاده از لبه‌ها، از روش ساده برای تشخیص صورت به کمک حبابها و رگه‌ها^{۲۴۵} (دباله ای از لبه‌های همسو) استفاده شده است. این مدل چهره از دو حباب تیره و سه حباب روشن برای بازنمایی چشمها، گونه‌ها، و بینی استفاده می‌کند. مدل همچنین از رگه‌ها برای نمایش پیرامون چهره، ابروها، و لبه‌ها استفاده می‌کند. دو تنظیم مثبتی برای کد کردن رابطه مکانی بین حبابها استفاده شده است. یک تصویر لاپلاسین دقت پایین برای تسهیل یافتن حبابها تولید می‌شود. 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 می‌شود.

در مقاله دیگری یک راه برای مکانیابی ویژگی‌های چهره و چهره‌ها در تصاویر خاکستری فام تهیه شده است. پس از فیلتر باند گذر، عملگرهای مورفلوژی برای بسط نواحی دارای شدت روشنایی بالا و اشکال معین استفاده می‌شوند (مثلاً چشمها). هیستوگرام تصویر پردازش شده معمولاً یک قله برجسته را نشان می‌دهد. بر اساس این قله و پهنه‌ی آن، مقادیر آستانه تطبیقی انتخاب می‌شوند تا دو تصویر دوستخی (باينری) بسازند. در هر دوی این تصاویر مؤلفه‌های همبند تشخیص داده می‌شوند تا محلهای کاندید برای ویژگیهای چهره را نشان دهند. پس از آن ترکیب این محلها توسط یک دسته بندی کننده ارزیابی می‌شود تا تعیین کند که آیا چهره‌ای موجود هست یا خیر. در این مقاله نحوه اعمال عملگرهای مورفلوژی 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 دهنده‌های 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) 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 می‌شود و مدل آن روی مجموع تصاویر به صورت توزیع گوسی بدست می‌آید. یک کلیشه چهره با میانگین‌گیری از پاسخ فیلترهای مشتق گوسی در

²⁴⁵ blobs and streaks

²⁴⁶ random graph matching

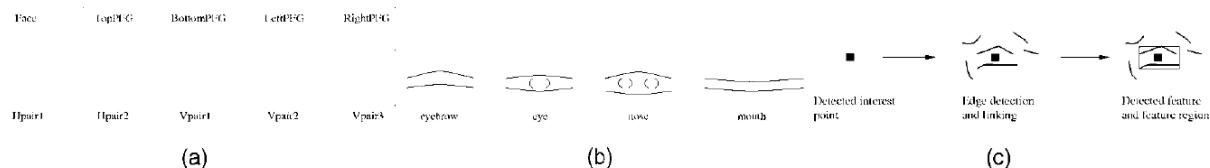
راستاها و مقیاسهای گوناگون (در پیکسلهای درون ویژگی چهره) برای تعدادی از چهره‌های موجود در پایگاه داده تعریف می‌شود. با دادن یک تصویر آزمایشی، ویژگیهای کاندید چهره توسط تصابق پاسخ فیلتر در هر پیکسل با بردار کلیشه شناسایی می‌شود (که ذاتاً مشابه عمل همبستگی است). دو کاندیدای برتر ویژگیها که قویترین پاسخ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

بهای استفاده از فاصله متقابل برای توصیف رابطه بین ویژگیهای چهره در مجموعه ستاره‌ای، می‌توان از مدل دیگری برای چهره نیز استفاده کرد. در یک مقاله، بازنمایی و رتبه بندی مجموعه‌های ستاره‌ای به مدد نظریه احتمالات شکل صورت گرفته است. آمار شکل، یکتابع توزیع احتمال توان روی n نقطه ویژگی است که برای ویژگی نام با (λ_1, x_1) و با فرض اینکه نقاط ویژگی اصلی روی صفحه‌ای با توزیع n^2 بعدی توزیع نرمال است، نشان داده می‌شود. این مقاله از همان روش بیشینه درستمایی برای تعیین کاندید چهره استفاده کرده است. یک مزیت این روشها اینست که چهره‌های که بخشی از آنها پوشانده شده است را می‌تواند تشخیص دهد. با اینحال روش نیست که آیا این روشها می‌توانند 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 می‌کند. در نخستین گام یک فیلتر مشتق دوم گوسی به نسبت سه بر یک بر تصویر خام اعمال می‌شود. نقاط مهم در بیشینه‌های محلی پاسخ فیلتر تشخیص داده می‌شود که نمایانگر مکانهای احتمالی ویژگیهای چهره است. در گام دوم لبه در اطراف نقاط مهم آزموده می‌شود و نقاط را به نواحی جداگانه ای گروه بندی می‌کند. این گروه بندی ادراکی لبه‌ها 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 می‌شود. میانگین و ماتریس کوواریانس هر بردار ویژگی چهره از روی داده‌های آموزشی ویژگیهای چهره محاسبه می‌شود. یک ناحیه تصویر می‌تواند یک کاندید قبل قبول ویژگی چهره باشد اگر فاصله مابالاتوبیس بین بردار ویژگیهای متناظر آن از آستانه ای کمتر باشد. ویژگی‌های برچسب خورده براساس دانش مدل -که می‌داند که آنها نسبت به هم در کجاها باید دیده شوند- بازهم گروه بندی می‌شوند. هر ویژگی چهره و گروه سپس با استفاده از شبکه بیزی ارزیابی می‌شود. یک وجه جالب این روش اینست که که این روش می‌تواند چهره‌ها را در راستاها و مکانهای تشخیص دهد کارایی آن برای تصاویر چهره‌هایی با ابعاد، راستاها و دیدگاههای مختلف به^{۸۵} درصد

²⁴⁷ constellation

هم رسیده است. با این وجود نرخ تشخیص اشتباه^{۲۸} 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 ۴۲-۳ شمایی از این روش تشخیص چهره مبتنی بر ویژگی را نشان می دهد.



شکل ۴۲-۳: (چپ) مدل چهره به صفحه یک صورت یک صفحه با شش ویژگی چهره هم راستا (ابروها، چشمها، بینی و دهان) (وسط) هر ویژگی چهره به صورت یک جفت لبه هم راستا 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 پویش 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 پویش جزئی تر تصویر پالایش می شود تا چهره پیدا شود.

نویسندهای 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 مقطعی^{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 اشتباهًا 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.

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 بافت‌ها. این روش از شبکه های عصبی همبستگی سری^{۲۵۳} برای دسته بندی با ناظر و شبکه های خود سازمانده ویژگی کوهون^{۲۵۴} برای تشکیل خوشه های کلاسیهای مختلف بافت استفاده می کند. برای استنتاج از برچسب بافت‌ها، 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 پویش 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.

۳-۲-۳-۳ - 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 عمدها در شدت روشنایی آنهاست و نه در خلوص رنگی آنهاست.^{۲۵۵} چند فضای رنگی مختلف از قبیل RGB، CIE LUV، CIE XYZ، YIQ، YCrCb، HSI^{2۵۶} یا HSV شده، استفاده شده است. روش‌های زیادی پیشنهاد شده اند تا یک مدل رنگ پوست را بسازند. ساده ترین روش، تعریف یک ناحیه از پیکسلهای همفام پوست با استفاده از مقادیر $R(Cr,Cb)$ است. یعنی $R(Cr,Cb)$ از نمونه های پیکسلهای رنگ پوست. با انتخاب

²⁵³ cascade correlation neural networks

²⁵⁴ Kohonen's self organizing feature map

²⁵⁵ chrominance

مناسب آستانه ها (Cr, Cb)، یک پیکسل دارای تون پوست دسته بندی می شود اگر مقادیر (Cr_1, Cr_2, Cb_1, Cb_2) 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 کنترل 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 بیشنه کردن امید ریاضی تخمین زده می شود.

در یکی از تحقیقات گسترده، که در آن نزدیک یک میلیارد پیکسل تون رنگ پوست برچسب گزاری 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 دقیق 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. بهر حال 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 رنگ^{۲۵۶} 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 شی 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. بهر حال این روش نمی تواند برای پیدا کدن چهره در تنها یک تصویر ثابت استفاده شود.

²⁵⁶ color constancy

معمولاً رنگ پوست به تنها بیان 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.

۳-۲-۴- 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 نظری 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 خوش 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 روشها 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 دو بعدی 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 بودن 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 روشنی 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 خوش در تصویر دو سطحی مشخص شود، یک مجموعه از ۱۱ گشتاور هندسی مرتبه پایین با استفاده از

تبديل فوريه و ملين شعاعي^{۲۵۷} استفاده می شود. برای تشخيص، 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 يك تابع توزيع شرطی کلاسي در فضای YES صورت می شود و پس از آن برای بدست آوردن نواحي متصل هموارسازی می شوند. سپس يك كليشه بيضي چهره برای تعين شباht نواحی رنگ پوست بر اساس فاصله هاوسدورف^{۲۵۸} استفاده می شود. در نهايـت نقاط ميانـي چشم بر اساس چند تابع هـزيـنهـ کـهـ اـزـ تـقارـنـ ذاتـيـ چـشمـ وـ صـورـتـ بهـرهـ مـیـ روـنـدـ،ـ پـيـداـ مـیـ شـونـدـ.ـ نـوكـ بيـنـيـ وـ مرـكـزـ دـهـانـ نـيـزـ پـسـ اـزـ آـنـ باـ بهـرهـ گـيرـيـ اـزـ مرـكـزـ چـشمـهاـ پـيـداـ مـیـ شـونـدـ.ـ يـكـ ضـعـفـ اـيـنـ روـشـ اـيـنـستـ کـهـ اـيـنـ روـشـ تـنـهاـ برـايـ حـالـتـيـ اـزـ چـهـرـهـ کـهـ اـزـ روـبـرـوـ وـ درـ حـالـتـيـ کـهـ دـوـ چـشمـ دـيـدـهـ شـوـدـ بـهـ درـدـ مـیـ خـورـدـ.ـ روـشـ دـيـگـرـيـ نـيـزـ اـبـداعـ شـدـهـ اـسـتـ کـهـ اـزـ رـنـگـ وـ تـقارـنـ محلـيـ بـهـرهـ مـیـ جـوـيدـ.

در مقابل روشـهـاـيـ مـبـتنـيـ بـرـ پـيـكـسلـ يـكـ روـشـ مـبـتنـيـ بـرـ سـاخـتـارـ،ـ رـنـگـ،ـ وـ هـنـدـسـهـ تـولـيدـ شـدـهـ اـسـتـ.ـ اـبـتـداـ تـقطـيعـ چـندـ مقـيـاسـهـ بـرـايـ استـخـراـجـ نـواـحـيـ هـمـگـنـ درـ يـكـ تصـوـيرـ اـسـتـفـادـهـ مـیـ شـودـ.ـ باـ استـفـادـهـ اـزـ يـكـ مـدلـ رـنـگـ پـوـسـتـ گـوسـيـ،ـ نـواـحـيـ دـارـايـ توـنـ پـوـسـتـ اـسـتـخـراـجـ مـیـ شـودـ وـ بـيـضـيـ وـارـ گـروـهـ بـندـيـ مـیـ شـودـ.ـ يـكـ چـهـرـهـ درـ صـورـتـ تـشـخـصـ دـادـهـ مـیـ شـودـ اـگـرـ وـيـژـيـهـاـيـ چـهـرـهـ مـانـدـ دـهـانـ وـ چـشـمـهاـ درـ اـيـنـ بـيـضـيـ وـجـودـ دـاشـتـهـ باـشـدـ.ـ آـزـماـيـشـاتـ نـشـانـ مـیـ دـهـدـ کـهـ اـيـنـ روـشـ قـابـلـيـتـ تـشـخـصـ چـهـرـهـ درـ رـاستـاهـاـيـ مـخـتـلـفـ وـ حتـىـ درـ حـضـورـ وـيـژـيـهـاـيـ مـانـدـ رـيشـ وـ عـيـنـكـ رـاـ نـيـزـ دـاردـ.

در روـشـ دـيـگـرـيـ يـكـ اـرـائـهـ حـبـابـيـ بـرـايـ استـخـراـجـ تـوصـيـفـيـ فـشـرـدهـ وـ دـارـايـ سـاخـتـارـ معـنـادـارـ اـزـ تصـاوـيرـ چـنـدـطـيـفـيـ مـاهـوارـهـ اـيـ پـيـشـنهـادـ شـدـهـ اـسـتـ.ـ درـ هـرـ پـيـكـسلـ يـكـ بـرـدارـ وـيـژـگـيـ تـشكـيلـ مـیـ شـودـ کـهـ حـاـصـلـ الـحـاقـ مـخـتـصـاتـ تصـوـيرـ پـيـكـسلـ بـهـ مؤـلفـهـ هـاـيـ طـيـفـيـ (ـيـاـ مـتـنـيـ)ـ پـيـكـسلـ مـیـ باـشـدـ.ـ سـپـسـ پـيـكـسلـهاـ باـ استـفـادـهـ اـزـ اـيـنـ بـرـدارـ وـيـژـگـيـ خـوشـهـ بـندـيـ مـیـ شـونـدـ تـاـ مؤـلفـهـ هـاـيـ چـسـبـيـدـهـ وـ هـمـبـنـدـ،ـ يـاـ حـبـابـهاـ رـاـ تـشكـيلـ دـهـنـدـ.ـ بـرـايـ تـشـخـصـ چـهـرـهـ هـاـ،ـ هـرـ بـرـدارـ وـيـژـگـيـ اـزـ مـخـتـصـاتـ تصـوـيرـ وـ خـلوـصـ رـنـگـيـ نـرـمـالـ شـدـهـ کـهـ رـابـطـهـ آـنـ بـهـ صـورـتـ زـيـرـ اـسـتـ،ـ استـفـادـهـ مـیـ کـنـدـ.ـ سـپـسـ يـكـ الـگـورـيـتمـ اـتـصالـ بـرـايـ رـشـدـ حـبـابـهاـ استـفـادـهـ مـیـ شـودـ وـ حـبـابـ پـوـسـتـ حـاـصـلـهـ کـهـ اـنـداـزـهـ وـ شـكـلـ آـنـ بـهـ يـكـ چـهـرـهـ مـعيـارـنـزـديـكـترـ اـسـتـ بـهـ عنـوانـ چـهـرـهـ تـشـخـصـ دـادـهـ مـیـ شـودـ.

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

فاصلـهـ وـ رـنـگـ نـيـزـ درـ روـشـيـ دـيـگـرـ بـرـايـ تـشـخـصـ چـهـرـهـ استـفـادـهـ مـیـ شـودـ.ـ نقـشـهـ عدمـ تـوـافـقـ^{۲۵۹}ـ مـحـاـسـبـهـ مـیـ شـودـ وـ اـشـياـ بهـ کـمـکـ يـكـ هيـسـتوـگـرامـ عدمـ تـوـافـقـ اـزـ پـسـ زـمـينـهـ تـقطـيعـ مـیـ شـودـ.ـ بـرـايـ اـيـنـكـارـ فـرـضـ مـیـ شـودـ کـهـ پـيـكـسلـهاـيـ پـسـ زـمـينـهـ دـارـايـ عـمقـ يـكـسانـ هـسـتـنـدـ وـ اـزـ پـيـكـسلـهاـيـ اـشـيـاـيـ پـيـشـ زـمـينـهـ بـسـيـارـ بـيـشـتـرـنـدـ.ـ باـ استـفـادـهـ اـزـ تـوزـعـ گـوسـيـ درـ فـضـاـيـ رـنـگـيـ RGBـ نـرـمـالـ شـدـهـ،ـ نـواـحـيـ تـقطـيعـ شـدـهـ باـ رـنـگـ شـبـيهـ چـهـرـهـ بـهـ عنـوانـ چـهـرـهـ طـبـقـهـ بـندـيـ مـیـ شـونـدـ.

²⁵⁷ radial Mellin transform

²⁵⁸ Hausdorff distance

²⁵⁹ disparity map

۴-۲-۳- روش‌های مبتنی بر تطبیق کلیشه

در تطبیق کلیشه، یک الگوی استاندارد چهره (معمولًاً روبرو) از قبل بطور دستی تعریف می‌شود یا با یک تابع پارامتری می‌شود. با داشتن یک تصویر ورودی، مقادیر همبستگی با الگوی استاندارد برای پیرامون چهره، چشمها، بینی، و دهان بطور جداگانه محاسبه می‌شود. وجود یک چهره بر مبنای مقادیر همبستگی اعلام می‌شود. این روش دارای مزیت سادگی پیاده سازی است. به حال ثابت شده است که این روشها برای تشخیص چهره کافی نیستند زیرا نمی‌توانند با تغییرهایی در مقیاس، وضعیت، و شکل بطور کارایی برخورد کنند. برای کسب نامتغیر بودن با مقیاس و شکل، زیرکلیشه‌های دارای ابعاد و دقتهای مختلف و نیز کلیشه‌های انعطاف‌پذیر پیشنهاد شده‌اند.

۴-۲-۳- کلیشه‌های 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 می‌شود تا چهره را مدل کند. هر زیرکلیشه با تکه خطهای مشخص می‌شود. خطوط در تصویر ورودی بر مبنای بیشترین تغییرات گرادیان تعیین شده و سپس با زیرکلیشه تطبیق داده می‌شود. همبستگی بین زیرتصاویر و کلیشه‌های پیرامون ابتدا کاندیدهای مکان چهره را تعیین می‌کند. سپس در این مکانهای کاندید تطبیق دادن با سایر زیرکلیشه‌ها صورت می‌گیرد. به عبارت دیگر، فاز نخست تمرکز و توجه یا ناحیه مورد علاقه را مشخص می‌کند و فاز دوم جزئیات را برای تعیین حضور چهره می‌سنجد. ایده تمرکز توجه و زیرکلیشه‌ها 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 لبه‌ها استفاده می‌شود. این لبه‌ها برای جستجوی کلیشه چهره بر مبنای چند قید گروه بندی می‌شوند. پس از آنکه کانتور سر پیدا شد، همین فرایند با ابعاد مختلف برای پیدا کردن ویژگیهای نظیر چشمها، ابروها، و لبها تکرار می‌شود. بعدها مقاله دیگری یک روش مکانیابی با استفاده از ۴۰ کلیشه برای جستجوی ویژگیهای چهره و یک استراتژی کنترلی برای هدایت و ارزیابی تشخیص دهنده‌های ویژگی مبتنی بر کلیشه پیشنهاد کرد.

مقاله‌ای دیگر روشی دو مرحله‌ای برای تشخیص چهره ارائه کرده است که در آن فرضیه‌هایی در مورد وجود چهره بوجود می‌آید و آزموده می‌شود. یک مدل چهره به صورت ویژگیها و با لبه‌ها تعریف می‌شود. این ویژگیها اتحناهای سمت چپ چهره تمام رخ، سمت راست آن و خط رویش مو را توصیف می‌کند. تشخیص دهنده لبه مار-هیلدرت^{۲۶۰} برای تهیه نقشه لبه‌های تصویر ورودی استفاده می‌شود. سپس یک فیلتر برای حذف اشیایی که کانتور آنها احتمالاً جزو چهره نیست استفاده می‌شود. زوج تکه کانتورهای بر اساس نزدیکی و راستای نسبی بهم متصل می‌شوند. گوشه‌ها برای تکه سازی کانتور به ویژگیهای منحنی پیدا می‌شود. این ویژگیهای منحنی سپس با بررسی ویژگیهای هندسی و مقعیت نسبی در همسایگی خود برچسب گزاری می‌شوند. اگر مشخصات زوج ویژگیهای منحنی سازگار باشد (یعنی هر دو بتوانند متعلق به یک چهره باشند) با هم ادغام می‌شوند. نسبت زوج ویژگیهایی که یک لبه را تشکیل می‌دهند با نسبت طلایی مقایسه می‌شود و یک هزینه به

²⁶⁰ Marr-Hildert

لبه تخصیص داده می شود. اگر هزینه گروه سه تایی از ویژگیهای منحنی (با برجستهای 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، روشی 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.

^{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.

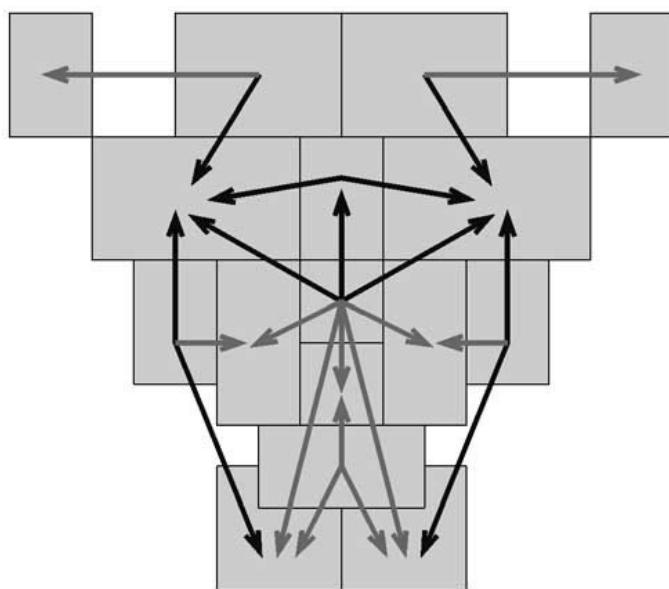
یک روش مکانیابی بر مبنای چند کلیشه برای مؤلفه های چهره نیز پیشنهاد شده است. این روش شامل تعداد زیادی فرضیه برای ظاهرهای احتمالی ویژگیهای چهره است. سپس یک مجموعه از فرضیه ها برای وجود چهره بر اساس فرضیه های ویژگیهای چهره به روش دمپستر-شفر تعریف می شود. با داشتن یک تصویر، تشخیص دهنده های ویژگی فاکتور اطمینان 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 ۴۳-۳ کلیشه بهبود یافته با ۲۳ رابطه را نشان می دهد. این رابطه های تعریف شده به ۱۱ رابطه اساسی (فلشها) تخت) و ۱۲ رابطه تایید کننده (فلشها خط چین) تقسیم می شوند. هر پیکان در این شکل یک رابطه را نشان می دهد، که

²⁶¹ silhouette

پایان نامه کارشناسی ارشد کورش مشگی - پاییز ۱۳۸۹
دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر

نوك 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 16×16 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 دقق تصویر تشکیل می شود(ر.ک. شکل ۳-۳۸) و لبه ها با استفاده از عملگر لاپلاسین 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- 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 ارجاعی^{۲۶۳} مدل می شوند (مانند چشمها) استفاده کرده است. در این روش، ویژگیهای چهره با یک کلیشه پارامتری توصیف می شوند. یکتابع انرژی برای تعریف لبه های متصل، قله ها و دره های تصویر که متناظر با پارامترهای کلیشه هستند، استفاده شد. بهترین تطبیق مدل ارجاعی 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 شئ 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 کانولو می شود و سپس یک عملگر مورفولوژیک برای تقویت لبه ها بکار گرفته می شود. یک مار تغییر یافته n پیکسلی (n 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 کردنتابع 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/جانه بطور دستی برچسب خورده اند و یک بردار از نقاط نمونه برای معرفی شکل بکار برد 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 فعل شکل^{۲۶۶} برای تخمین مکان شئ 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 کالمون توسعه داده شده است تا پارامترهای مستقل از شکل شدت نور را تخمین بزنند و چهره را در دنباله تصاویر ردگیری کند.

۳-۵-۴- 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 می‌شود. در حال حاضر به دلیل محدود کردن حجم محاسبات و کارایی تشخیص، کاهش بعد انجام می‌گیرد.

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 می‌شود و این متغیر تصادفی برای کلاسهای چهره و غیرچهره با تابع توزیع احتمال کلاسی ($p(\vec{x} | \text{face})$) و ($p(\vec{x} | \text{nonface})$) معروفی می‌شود. دسته بندی بیزی یا بیشینه احتمال وقوع برای دسته بندی یک مکان کاندید تصویر برای چهره بودن یا نبودن می‌تواند استفاده شود. متاسفانه 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 عمدی یک روش مبتنی بر ظاهر به معابر ساختن تخمین پارامتری و غیرپارامتری این دو تابع اختصاص می‌یابد.

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 (معمولًاً 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 شده‌اند. این روشها به طور ضمنی الگوها را به بعد بالاتر فضایی می‌برند تا بین دو کلاس الگوی چهره و غیرچهره که در آن فضا بازتاب شده‌اند، صفحه ای جداکننده تشکیل دهند.

۲-۳-۱-۵- چهره‌های یکه

یکی از نخستین مثالهای استفاده از چهره‌های 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 می‌کند. این بردارهای یکه بعدها به چهره‌های یکه معروف شدند.

در مقاله‌ای ثابت شده است که چهره می‌تواند به صورت خطی توسط تعداد محدودی از تصاویر پایه کد شوند. این اثبات بر اساس تبدیل کاراون-لووه^{۲۶۷} 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 می‌شوند، بردارهای پایه که زیرفضای بهینه را پوشش می‌دهند طوری تعریف می‌شوند که خطای میانگین مربعی بین بازتاب تصویر آموزشی به این زیرفضا و تصویر اصلی کمینه شود. این مجموعه بردارهای پایه بهینه تصویر یکه خوانده می‌شود چرا که در حقیقت همان بردارهای یکه ماتریس کواریانسی هستند که از چهره بردار شده در مرحله آموزش بدست می‌آید.

²⁶⁷ eigenfaces

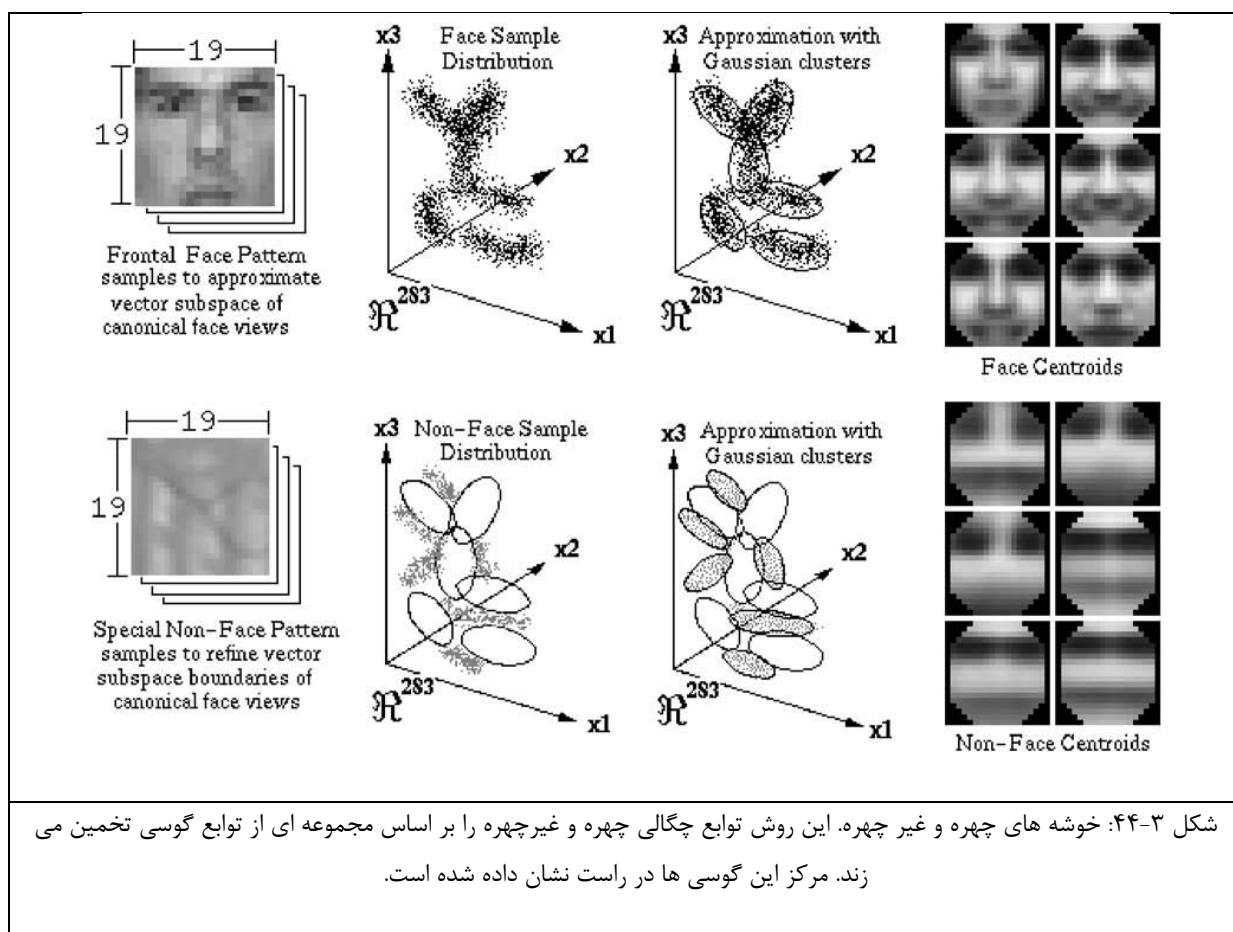
²⁶⁸ Karhunen-Loeve Transform

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 یکه (در اینجا چهره های یکه) را تولید کند که یک زیر فضای فضای تصویر که فضای چهره خوانده می شود جدا کند. تصاویر چهره بر روی این زیر فضا بازتاب شده و خوش بندی می شوند. به طور مشابه تصاویر آموزشی غیر چهره نیز بر روی همان زیر فضا بازتاب شده و خوش 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 فضای 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 یکه و خوش بندی استفاده کرده اند.

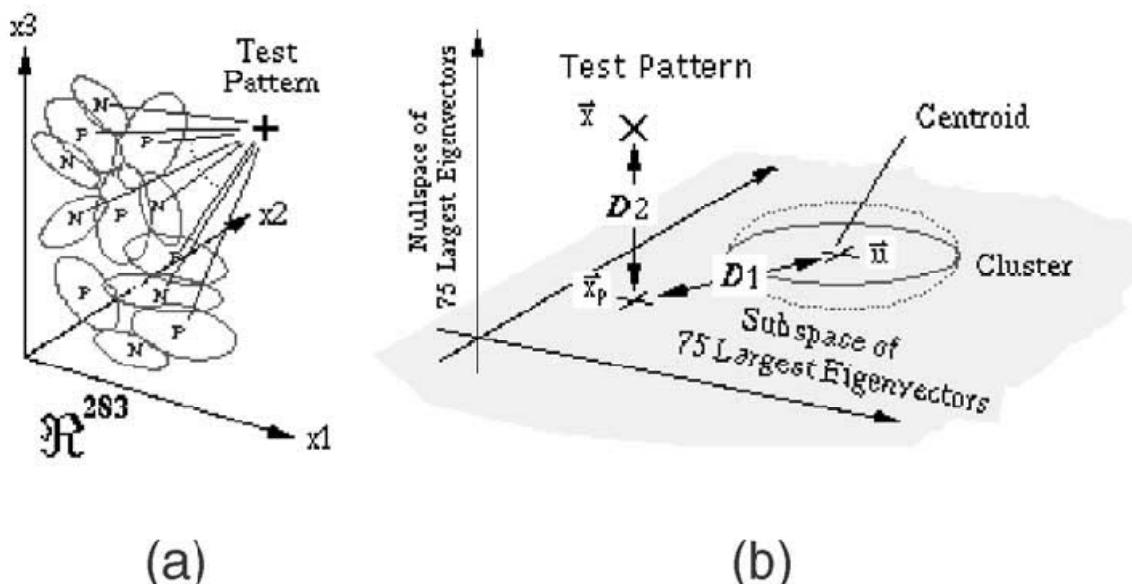
۲-۵-۲-۳- روش‌های مبتنی بر توزیع

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 شی 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 ^{۲۷۰} بُعدی در نظر گرفته می‌شود. سپس الگوها به شش خوش صورت و غیرصورت بر اساس یک الگوریتم k -میانگین تغییر یافته تبدیل می‌شوند که در شکل ۴۴-۳ به تصویر کشیده شده است.

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



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

اولین مؤلفه فاصله، فاصله ماهالانوبیس^{۲۷۱} نرمال شده بین الگوی تست و مرکز خوش است که در زیرفضایی با ابعاد پایینتر که با ۷۵ بردار ویژه بزرگتر خوش متعین شده است، اندازه گیری می‌شود. دومین مؤلفه فاصله، فاصله اقلیدسی بین الگوی تست و بازتاب آن روی زیر فضای ۷۵ بعدی 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 دسته بندی الگوهای پنجره چهره از الگوهای غیرچهره است که با استفاده از دوازده جفت فاصله که بین هر خوش چهره و غیرچهره حساب شده است انجام می‌شود. این دسته بندی 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 غیرچهره ها دشوارتر است. این مسئله با یک روش بوت استرپ^{۲۷۲} تخفیف داده شده است که در آن تصاویر را با انتخاب خود در طول آموزش به داده های آموزشی اضافه می‌کند. با شروع از یک مجموعه کوچک نمونه های غیرچهره در مجموعه آموزشی، دسته بندی کننده با این پایگاه داده از مثالها آموزش می‌بیند. سپس تشخیص دهنده چهره روی یک دنباله تصادفی از تصاویر اجرا

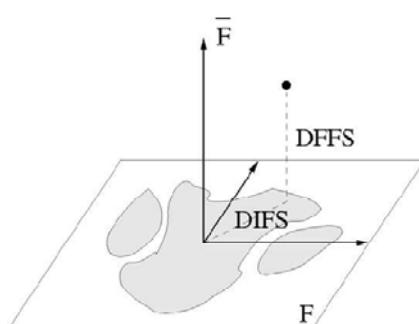
²⁷⁰ modified k-means algorithm

²⁷¹ Mahalanobis distant

²⁷² bootstrap

می شود و تمامی الگوهایی که سیستم فعلی اشتباهًا چهره اعلام کرده است را جمع آوردی می کند. داده های اشتباهًا 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 یکه 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 عمدی بین داده ها را حفظ می کند و قسمتهای خردتر آنرا دور میریزد. این روش فضای بردار را بین دو فضای منحصرًا متقابل و مکمل تجزیه می کند(شکل ۴۶-۳)، 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 شئ استفاده می شوند. روش پیشنهادی برای مکانیابی، کد کردن و شناسایی چهره استفاده می شود. در مقایسه با روش چهره یکه استاندارد، این روش پیشنهادی کارایی بهتری در شناسایی چهره نشان می دهد. در زمینه آشکارسازی چهره، این روش تنها به مکانیابی می پردازد.



شکل ۴۶-۳: تجزیه زیرفضای چهره به دو زیرفضای 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 واریانس های ناخواسته نیز جزو آنهاست.

۳-۵-۲-۳- شبکه های عصبی

شبکه های عصبی در مسئله های شناسایی الگو مانند شناسایی نوری حروف، شناسایی اشیا، 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 متشكل از دو زیرشبکه موازی بود که در آن ورودیها مقادیر شدت روشنایی تصویر اصلی و شدت روشنایی تصویر فیلتر شده با فیلتر 3×3 ^{۲۷۴} 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 دوسری 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 ۱۰۲۴ ورودی، ۲۵۶ واحد در لایه پنهان اول، هشت واحد در لایه پنهان دوم، و دو واحد خروجی است. این پیکر بندی بعدها نیز مجدد استفاده شده است.

یک مقاله به پویش تصویر ورودی با یک شبکه عصبی دارای تاخیر زمانی (با میدان گیرندگی 20×25 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 پیچشی^{۳۷۵} 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.



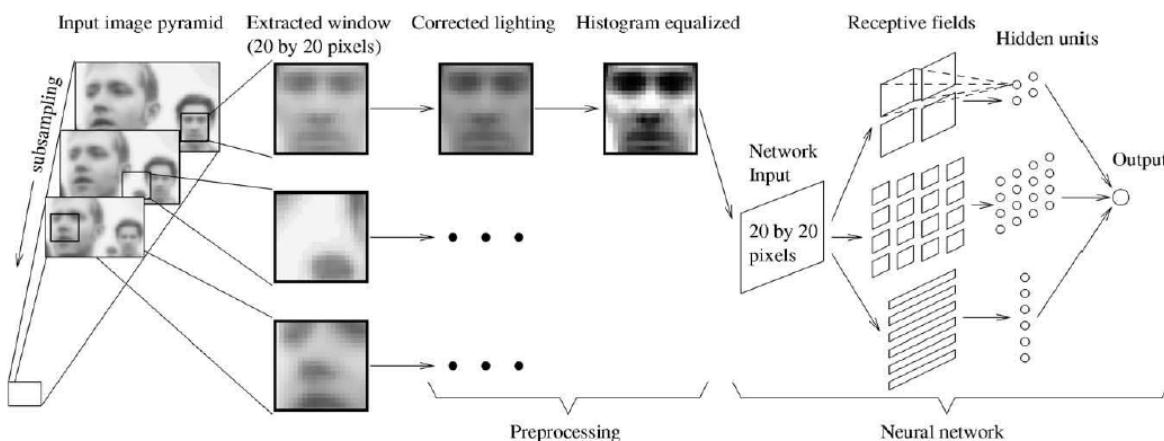
شکل ۳-۴۷: نمونه های هر کلاس چهره برای SOM کوهونن که هر کدام مرکز یک خوشه هستند.

^{۲۷۴} Sobel filters

^{۲۷۵} convolutional neural network

روش دیگری از تشخیص چهره با شبکه عصبی وجود دارد که در آن تعداد زیادی نمونه از چهره ها و غیر چهره ها توسط الگوریتم SOM کوہونن 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 نور پردازی نرمالسازی می شوند. سپس هر پنجره نرمالیزه با یک MLP دسته بندی می شود (شکل ۴۷-۳).^{۴۷-۳}

در مقاله ای دیگر که بی شک یکی از موقترين روشاهای 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 ۱-۱ 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 الگوی غیر چهره دارد (نزدیک به -۱). 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 زیر نمونه گیری شود و شبکه به هر مقیاس اعمال شود. نزدیک به ۱۰۵۰ نمونه چهره از اندازه ها، راستاهای، موقعیتها و نور پردازیها برای آموزش شبکه استفاده شده اند. در هر تصویر ورودی، چشمها، نوک بینی، و گوشها و سطح 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. روشاهای قضاوت^{۲۷۷} ساده ای مانند عملگرهای منطقی (AND و OR) و روشاهای رأی گیری برای افزایش کارایی استفاده شده اند. چندین سیستم دیگر نیز با روشاهای قضاوت تولید شده اند.



شکل ۴۸-۳: دیاگرام سیستم روش رولی. هر چهره پیش از خورانده 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 arbiterate scheme

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

۳-۲-۴-۵- ماشینهای بردار پشتیبان

ماشینهای بردار پشتیبان^{۲۷۸} می‌توانند برای آموزش دسته بندی کننده‌های تابع چند جمله‌ای، شبکه‌های عصبی، و تابعهای شعاعی استفاده شوند. در حالی که خیلی از روشها آموزش یک دسته بندی کننده (مثلًاً بیزی، شبکه عصبی، و تابعهای شعاعی) بر مبنای کمینه کردن خطای آموزش و ریسک تجربی است، SVM‌ها بر مبنای قاعده استنتاجی دیگری کار می‌کنند که کمینه کردن ریسک ساختاری نام دارد و هدف آن اینست که کران بالای خطای تعیین را حداقل کند. یک دسته بندی کننده SVM یک دسته بندی کننده خطی است که در آن ابرصفحه جدا کننده طوری انتخاب می‌شود 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 می‌شوند تعریف می‌شود. تخمین این ابرصفحه بهینه برابر حل کردن یک مسئله برنامه نویسی درجه دو با قیود خطی است. بهر حال محاسبات از هر دو لحظه زمان و حافظه حجمی است.

در مقاله‌ای روشی کارا برای آموزش SVM در مسائل بزرگ پیشنهاد شده و برای تشخیص چهره اعمال شده است. بر اساس دو مجموعه داده حاوی ۱۰۰۰۰ رنگی آزمون 19×19 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 چهره‌ها و عابران پیاده در دامنه موجک نیز استفاده شده اند.

۳-۲-۵-۵- شبکه تنک غربالها

مقاله‌ای یک روش با استفاده از معماری یادگیری SNoW برای تشخیص چهره پیشنهاد داده است تا چهره‌هایی با ویژگیها و حالات متفاوت، در موقعیتهای گوناگون، و تحت شرایط متفاوت نورپردازی را تشخیص دهد. این مقاله همچنین به بررسی اثرات یادگیری با مؤلفه‌های اساسی و نیز با ویژگیهای چندمقیاسه نیز پرداخته است. در واقع SNoW یک شبکه تنک از توابع خطی است که از قوانین بروزرسانی غربال بهره می‌جویند. این روش بخصوص برای یادگیری در زمینه‌هایی 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، تشخیص ویژگیها و روابط بصورت داده گرای، مکانیسم تصمیم‌گیری، و استفاده از یک قانون بروزرسانی کارا. این روش در قیاس با اکثر روشهای دیگر آشکارسازی چهره از لحاظ محاسبات برتری دارد در حالی که دقیقی در سطح آنها را فراهم می‌کند.

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 می‌دهد. این روش بر ظاهر محلی تاکید دارد زیرا برخی الگوهای محلی یک شکل، از دیگر الگوها یکتاً نیستند. الگوهای شدت روشنایی در اطراف چشم بسیار متمایزتر از اطراف گونه‌ها 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 زیرناحیه‌ها می‌زند. دوم اینکه یک دسته بندی کننده ساده بیزی فرم عملیاتی از احتمال ثانویه را فراهم می‌آورد که می‌تواند آمار توأم ظاهر محلی و موقعیت شئ را ذخیره می‌کند. در هر مقیاس یک تصویر چهره به چهار زیرناحیه مستطیلی تجزیه می‌شود. این زیرناحیه‌ها سپس با کمک PCA به فضایی با ابعاد بازتاب می‌شوند و آمارگان هر زیرناحیه بازتاب شده از نمونه‌های بازتاب شده برای ظاهر محلی تخمین زده می‌شود. با این فرمولاسیون چهره زمانی حاضر است که نسبت احتمال وجود چهره از احتمال ثانویه بیشتر باشد. این روش کارایی نسبتاً خوبی دارد و می‌تواند چهره‌های چرخیده و نیمرخها را نیز تشخیص دهد. نویسنگان این مقاله، بعدها آنرا با ارائه موجک ارتقا دادند.

۳-۲-۵-۷- مدل مخفی مارکوف

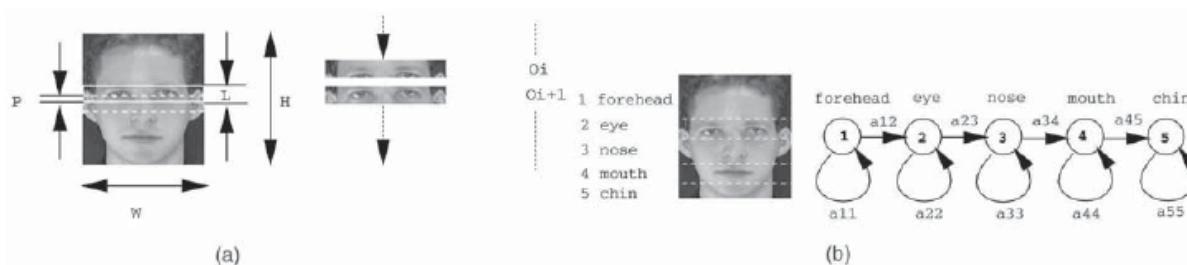
در مدل مخفی مارکوف^{۲۸۰} (HMM) فرض می‌شود که الگوهای داده‌ها را می‌توان با یک فرایند تصادفی به صورت پارامتری مدل‌سازی کرد، به طوری که پارامترهای این مدل با یک روش دقیق قابل محاسبه باشند. در طراحی HMM برای مسائل بازنگاری الگو برای تعیین ساختار مدل، ابتدا باید تعدادی حالت مخفی در نظر گرفت. سپس می‌توان HMM را آموزش داد تا احتمال انتقال میان حالات را از روی نمونه‌های آموزشی که رشته‌ای از مشاهدات هستند، فرآگیری کند. هدف از آموزش یک HMM، بیشینه سازی احتمال مشاهده داده‌های آموزشی از طریق تغییر و اصلاح پارامترهای موجود در مدل HMM با استفاده از الگوریتم‌های استاندارد تقطیع ویتری^{۲۸۱} و الگوریتم بام-ولش^{۲۸۲} است. پس از آموزش HMM، احتمال خروجی یک مشاهده، کلاس داده موردنظر را تعیین می‌کند.

به صورت شهودی می‌توان الگوی چهره را به چند ناحیه از جمله پیشانی، چشمها، بینی، دهان و چانه تقسیم کرد. یک الگوی چهره را می‌توان با فرایندی که طی آن این ناحیه‌ها در یک ترتیب مناسب (از بالا به پایین و از چپ به راست) مشاهده می‌شوند، بازنگاری کرد. به جای تکیه کردن بر روی ترتیب صحیح، مانند تطبیق الگو یا روش‌های مبتنی بر ظاهر (که در آنها ویژگی‌های چهره مانند چشمها و بینی باید با توجه به یک نقطه مرکزی مرتب باشند)، در این روش، هدف، ایجاد تناظر میان ناحیه‌های مختلف چهره با حالاتی از یک مدل مخفی مارکوف با چگالی پیوسته است. روش‌های مبتنی بر HMM عموماً با یک الگوی چهره به صورت رشته‌ای از بردارهای مشاهده برخور德 می‌کنند. هر یک از این بردارها، همانگونه که در شکل ۳-۴۸ چپ نشان داده شده، مجموعه‌ای از پیکسل‌ها می‌باشد.

²⁸⁰ hidden Markov model

²⁸¹ standard Viterbi segmentation method

²⁸² Baum-Welch algorithm



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

برای الگوهای چهره، همانگونه که در شکل ۴۹-۳ 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 بازنگشایی چهره و مکان‌یابی آن مورد استفاده قرار گرفته است. یک مقاله نشان داد که حالت‌های HMM متناظر با ناحیه‌های چهره هستند (شکل ۴۸-۳ راست). به عبارت دیگر، یک حالت مسئول مدل‌کردن 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 بازنگشایی 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 تک بعدی به یک حالت نسبت داده می‌شود. شکل ۴۸-۳ 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 ویتری جایگزین شده و پارامترهای HMM با روش بام-ولش مجددًا تخمین زده می‌شود. همانطور که در شکل ۴۸-۳ 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 مقاله‌ای دیگر از HMM و تبدیل کارهونن-لووه (KLT) برای مکان‌یابی و بازنگشایی 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 بازنگشایی 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 دو بعدی، ۹۰ درصد دقیق بوده 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 مرتبه 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 پیشینه برای دسته‌بندی دودویی باز هم پردازش می‌شود. نتایج آزمایشات نشان داده است که هر دو روش نرخ شناسایی بالاتری نسبت به روش پایه داشتند، اما اعلامهای اشتباہ آنها نیز بالاتر بود.

۳-۵-۸- 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 همرزلی-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.

مقاله‌ای 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 (x) و (q) 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 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 مقاله‌ای 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 دوتابع 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 درستنمایی محاسبه گردیده و برای شناسایی مکان چهره‌ها به کار گرفته می‌شود.

مقاله‌ای 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 یک 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.

۳-۵-۹- 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 8×8 پیکسلی است که با یک بردار حاوی 30 ویژگی بازنمایی می‌شود. این ویژگی‌ها عبارتند از: آنتربوی، 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 2340 تصویر چهره تمام رخ از دادگان FERET، مکان‌یابی چهره‌ها با دقت 96 درصد انجام گرفته است.

مقاله‌ای دیگر روشی را ارائه کرده است که در آن از الگوریتم Find-S می‌چل برای یادگیری مفهوم چهره استفاده می‌شود. نویسنده‌گان این مقاله آنها تصور می‌کرده اند که توزیع الگوهای چهره ($p(x | face)$) را می‌توان با مجموعه‌ای از

²⁸⁶ watershed

خوشه‌های گوسی تخمین زد و همچنین فاصله یک نمونه چهره تا یکی از مراکز خوش باید کمتر از یک نسبتی از بیشینه فاصله نقاط درون آن خوش تا مرکز آن خوش باشد. سپس الگوریتم Find-S اعمال می‌گردد تا فاصله حد آستانه را فراغیری 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.

۴-۵-۳- 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.

۶-۳- جمع بندی

در این پژوهش سعی بر این بوده است که مروی بر ایده‌های آشکارسازی چهره انجام پذیرد و دسته‌های ساختاری برای روش‌های معرفی شده در آنها ارائه شود. این اثر به عنوان تکمیل یک مقاله مروی(Yang 2002) با الگوریتم هایی از جدیدترین روش‌های تشخیص چهره ارائه شده است که در طبقه‌بندی ارائه شده توسط این مقاله گنجانده شده اند. آنچه که ادبیات نیاز زیادی به آن احساس نمی‌کند، داده‌های آموزشی و آزمایشی استاندارد و رویه‌های ارزیابی کارایی نظام مندتر است. این امر کمک می‌کند تا کاربران الگوریتم‌های آشکارسازی چهره بفهمند که کدام روشها در کدام زمینه رقبای جدی تری به شمار می‌آیند. همچنین محققان را به سمت تولید الگوریتم‌هایی که واقعاً کارا باشند سوق می‌دهد.

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.

²⁸⁷ holistic

آشکارسازی چهره به تنها یک مسئله دشوار و جالب است. با این وجود می تواند 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 گاهًا در پس زمینه های درهم ریخته گرفته می شوند. بنابراین آشکارسازی چهره با جمیع مشکلاتی که برای کاربرد عام آن، تشخیص اشیا، نیز وجود دارد دست و پنجه نرم می کند. بهر حال کلاس چهره دارای قاعده مندی بسیار آشکاری است که توسط بسیاری از الگوریتمهای اکتشافی و روشهای مبتنی بر مدل به خدمت گرفته می شود و یا توسط روشهای داده گرا، یاد گرفته می شود. 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 غیرچهره) است.

در پایان مهمترین مدلهای بکار گرفته شده در آشکارسازی چهره مرور می شود:

تحلیل مؤلفه اصلی: از تبدیل کارهونن-لووه مشتق شده است. با داشتن یک بازنمایی یکبعدی از هر چهره در مجموعه آموزش تصویر، PCA به دنبالیک زیرفضای یکبعدی می گردد که بردارهای پایه آن متناظر با جهتهای بیشترین واریانس در فضای اصلی تصویر باشند. این یزرفضای جدید دارای ابعاد کمتری است ($S \ll t$). اگر عناصر تصویر به عنوان متغیرهای تصادفی در نظر گرفته شوند، بردارهای پایه PCA به عنوان بردارهای یکه نمودار پراکندگی آنها خواهند بود و تصویر حاصل از آنها را چهره یکه می نامند.

تحلیل مؤلفه مستقل: این روش وابستگیهای مرتبه دوم و مراتب بالاتر در تصویر ورودی را کمینه می کند و تلاش می کند که بردارهای پایه خود را که داده ها روی آنها نگاشت داده می شوند، مستقل آماری انتخاب کند. دو معماری برای آشکارسازی چهره از این روش استفاده می کند: ۱- تصاویر پایه مستقل آماری - ۲- بازنمایی کد جدایی پذیر

تحلیل جداکننده خطی: این روش بردارهایی را در فضا می یابد که کلاسها را به بهترین نحو متمایز کنند. برای تمام کلاسها، ماتریس پراکندگی درون کلاسی (S_B) و بین کلاسی (S_W) تعریف می شوند. هدف این روش بیشینه سازی S_B در عین کمینه سازی S_W است که می توان 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 یکه $S_B^{-1} \times S_W^{-1}$ باشند.

تعقیب تکاملی^{۲۸۸}: یک روش تطبیقی مبتنی بر چهره های یکه که بدنبال بهترین محورهایی می گردد که افکنش روی آنها یک تابع برازش را بیشینه کند و در همان حال دقت دسته بندی و قدرت تعییم 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 می شود.

تطابق گراف دسته انعطاف پذیر^{۲۸۹}: تمام چهره های انسانی دارای ساختار توپولوژیک مشترکی هستند. چهره را می توان با یک گراف بازنمایی کرد که رأسهای آن روی نقاط مطمئنی قرار دارد (بینی، ...) و یالهای آن با بردارهای فاطله بویعده

²⁸⁸ evolutionary pursuit (EP)

پایان نامه کارشناسی ارشد کورش مشگی - پاییز ۱۳۸۹
دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر

برچسب خورده اند. هر رأس حاوی 40^{290} 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 رأسهای که با یالها متصلند که رأسها با جت ها برچسب گزاری می شود و یالها با فاصله برچسب گزاری 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.

تبديل ردگیری²⁹¹: این تبدیل تعمیمی بر تبدیل رادون²⁹² 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 شونده²⁹³: 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 بیزی²⁹⁴ 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 بیز اندازه گیری می شود.

ماشینهای بردار پشتیبان: با داشتن مجموعه نقاطی که به دو کلاس مربوط است، این ماشینهای 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. ایندا 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: مجموعه‌ای از مدل‌های آماری هستند که برای تشخیص خصوصیات آماری یک سیگنال استفاده می‌شوند. این مدلها از دو فرایند مداخل تشکیل شده‌اند: ۱- یک زنجیره مارکوفی غیرقابل مشاهده با حالتی محدود، یک ماتریس انتقال حالت احتمالی و یک توزیع احتمال حالت اولیه و ۲- یک مجموعه از تابعهای توزیع مناسب به هر حالت.

بوستینگ و راه حل‌های دسته جمعی: ایده پشت سر بوستینگ استفاده مداوم از یک یادگیرنده ضعیف روی نسخه وزن گزری شده ای از زیرمجموعه نمونه آموزشی است تا بتواند مجموعه ای از دسته بندی کننده‌های همنوع خود را تعیین 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

فصل چهارم

[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

۱-۴- 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 کدگزاری 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 چشممان، بینی، لبها، و موها باشند ولی اجراءً 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 کدگزاری یک تصویر با کدگزاری 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 کدگزاری 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 مفصلأً به آنها پرداخته شد را مرور می کنیم. هنگامی که تصویری روی شبکیه 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 رساند به بخش اولیه بینایی و لایه V1 هدایت می شوند. در این لایه داده ها شروع به کدگزاری می شوند. این کدگزاری در لایه های ابتدایی ساده است و با پیشروی در طول مسیر بطنی، که به دلیل ذات شناختی مسئله در اینجا دنبالی شود، کدگزاری 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 V1 به تغییر روشنایی در میدان گیرنده‌ی آن حساس هستند. مدلهایی نظری فیلترهای گابور برای توجیه عملکرد این سلولها پیشنهاد شده است. سلولهای پیچیده لایه V1 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، کدگزاری پراکنده، مخلوط مقیاس گوسی، مؤلفه های گوسی و مؤلفه های کواریانسی برای توجیه عملکرد این لایه پیشنهاد شده اند. بر طبق نظریه اطلاعات کدھایی که از نظر آماری کاراتر هستند، همانهایی هستند که قاعده مندیهای 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 مندیها موثر است و با اصول بیولوژیکی مطابقت دارد، مدلهای خود را ارائه داده اند که از جمله این معیارها می توان به همبستگی، پراکنده‌ی اطلاعات متقابل نورونها، استقلال آماری، و تطابق با کواریانس اشاره کرد. در کل سیستمهای بیولوژیکی زیر فشار تکاملی شدیدی هستند، می توان انگاشت که کدھایی که آنها استفاده می کنند بسیار کاراست. در اکثر مدلها، 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 گیرنده‌ی نورونهای واقعی، این فرض آنچنان که در نگاه اول به نظر می آید محدود کننده نیست. پس از این لایه V4، دارد که حساسیت آن به زوایا، بافتها، گوشه ها و کانتورها در نظر گرفته شده است. پس از آن هم لایه V2 قرار دارد.

که نقش آن تشخیص شکلهای ساده و کانتورهای بسته است. در تمامی این لایه‌ها اتصالات جانبی وجود دارد که وظایف مختلفی برای آنها فرض شده است که از جمله آنها می‌توان به تقویت و تضعیف نورونهای همسایگی، هماهنگی واریانس، و نرمال سازی با تقسیم (بر اساس پدیده فرونشانی اطرافیان) اشاره کرد. از سوی دیگر برای شبکه‌های عصبی چند لایه الگوریتمهای یادگیری، مسیرهای پسخور و الگوریتمهای رقابتی هم طراحی شده اند که همگی از خصوصیت‌های کلی مغز (که قشر بینایی نیز از آنها مستثنی نیست) الگوبرداری شده اند. ایده‌هایی مانند کدگزاری جمعیتی نیز در سطح یک لایه بسیار الهام بخش و سازگار با مشاهدات بیولوژیکی بوده است.

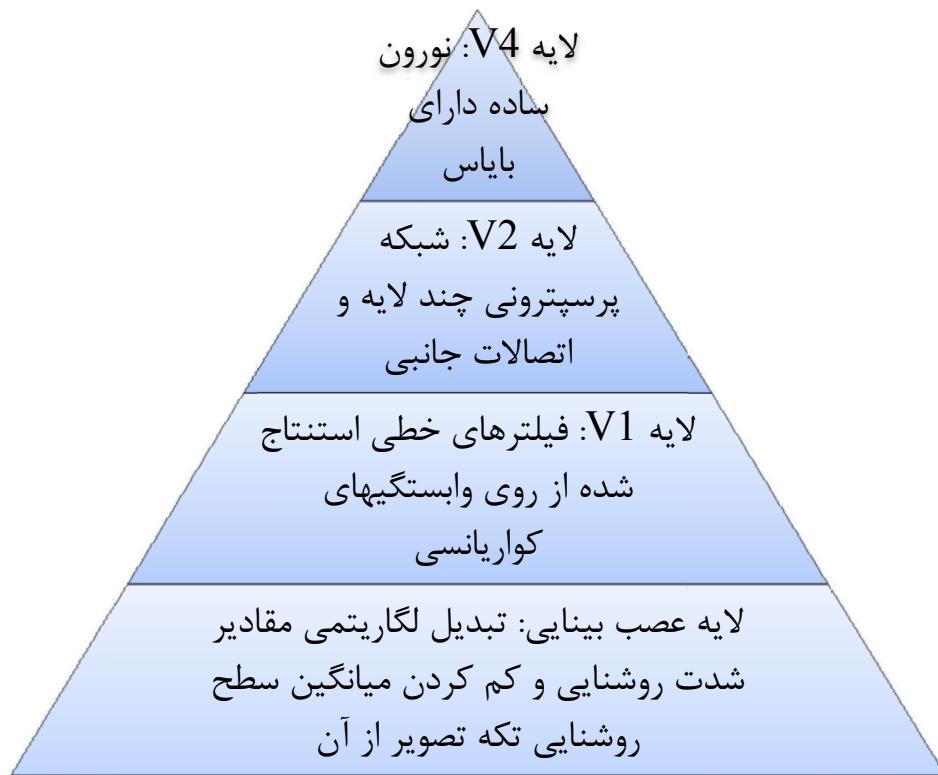
در این پژوهش سعی شده است تا با الهام از مدل‌های مشاهده شده در مغز، یک چارچوب برای مسئله آشکارسازی چهره معرفی شود. این مدل در قالب یک ساختار شبکه عصبی معرفی می‌شود و هر کدام از لایه‌های کلی آن جدآگانه بررسی می‌شود. شایان ذکر است که یادگیری هر کدام از لایه‌های پایینتر مستقل از لایه‌های بالاتر صورت می‌گیرد و تنها در فاز آشکارسازی چهره، داده‌های تصویری به صورت پیشخور از لایه‌ای به لایه بعدی منتقل می‌شود. سازمان این نورونها طوری است که هر چه از لایه‌های پایین به بالا پیش برویم، به گزینندگی نورونها افزوده می‌شود، و هر نورون ورودی خود را از تعداد زیادی نورون دریافت می‌کند. با حفظ نقشه تپوگرافیک نورونها، همانند آن چیزی که در مغز انسان صورت می‌پذیرد می‌توان مکان دیده شدن چهره را در طول لایه‌ها حفظ کرد.

در لایه نخست این مدل که مشابه لایه V1 است، هدف بدست آوردن یک بازنمایی جمعیتی است که بتواند ویژگیهای مشابه لایه V1 از خود نشان دهد. همانگونه که در فصل ۳ درباره کاهش افزونگی بحث شد، لایه V1 به گونه‌ای تطبیق یافته است که قاعده مندیهای آماری موجود در تصاویری که از طبیعت به چشم می‌رسد را کشف کرده، و یک کدگزاری کارا برای آن ارائه دهد. لذا برای مدل‌سازی این لایه به ادبیات فهم صحنه‌های طبیعی مراجعه شود و مدل‌های 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 چهره‌های انسانی را به طور خودکار بیاموزد، و در مقابل همبستگیهای نایستای تصویر (که از غیرصلب بودن چهره، تغییر دیدگاه و نورپردازی که موجب پدیده‌های کواریانسی (Karklin 2008) بهره جسته می‌شود، از ... مقاوم باشد. برای این مدل‌سازی از ایده مؤلفه‌های کواریانسی (Hyvarinen 2000) بخوبی از آن استفاده شده است بهره گرفته فرمولاسیون متغیرهای نهفته که در تحلیل مؤلفه مستقل (Hyvarinen 2000) بهره جسته می‌شود (Hyvarinen 2000, ۹۲)، استفاده از درستنمایی لگاریتمی که در این روش استفاده شده است مناسب به نظر می‌رسد (Pham et al 2003a)، در محاسبه درستنمایی از مدل مؤلفه‌های واریانسی (Karklin 2008) استفاده می‌شود و مطابق روشهای استفاده شده در (Hyvarinen 2000, Karklin 2008) از تابع اولیه زبرگوسی لابلسین استفاده می‌شود، با توجه به نیاز به سرعت و نیز تشابه به روش یادگیری هبی (R.K. فصل ۲) که در شبکه‌های عصبی طبیعی و مصنوعی کاملاً شناخته شده است، از روش شبی نوردی برای تقریب مقدار پارامترها و متغیرهای نهفته استفاده می‌شود و به صورت قانون یادگیری هبی به شبکه دیکته می‌شود، و سپس آزمایشاتی برای اثبات مناسب بودن کد بدست آمده برای تصاویر طبیعی ارائه می‌شود.

در لایه دوم این مدل که مشابه لایه V2 است، از بازنمایی بدست آمده در لایه V1 استفاده می‌شود و برای تمایز بازنمایی‌هایی که احتمالاً از تکه‌ای از تصویر چهره بدست آمده اند و بازنمایی‌هایی که چنین نیستند استفاده می‌شود. 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 تفاوت های ساختاری). حاصل نورونهای موجود در این لایه یک احتمال خواهد بود که احتمال اینکه تکه 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 تعییه 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 ۱-۴ آورده شده است.



شکل ۱-۴: نمایی کلی مدل پیشنهاد شده

مقدار پارامترهای این مدل در جدول ۱-۴ آورده شده است:

پارامتر	شرح	پیش فرض
I	ابعاد ورودی، تعداد پیکسلهای موجود در میدان گیرندگی نورون ساده V1، پیکسلهای تکه تصویر	۴۰۰

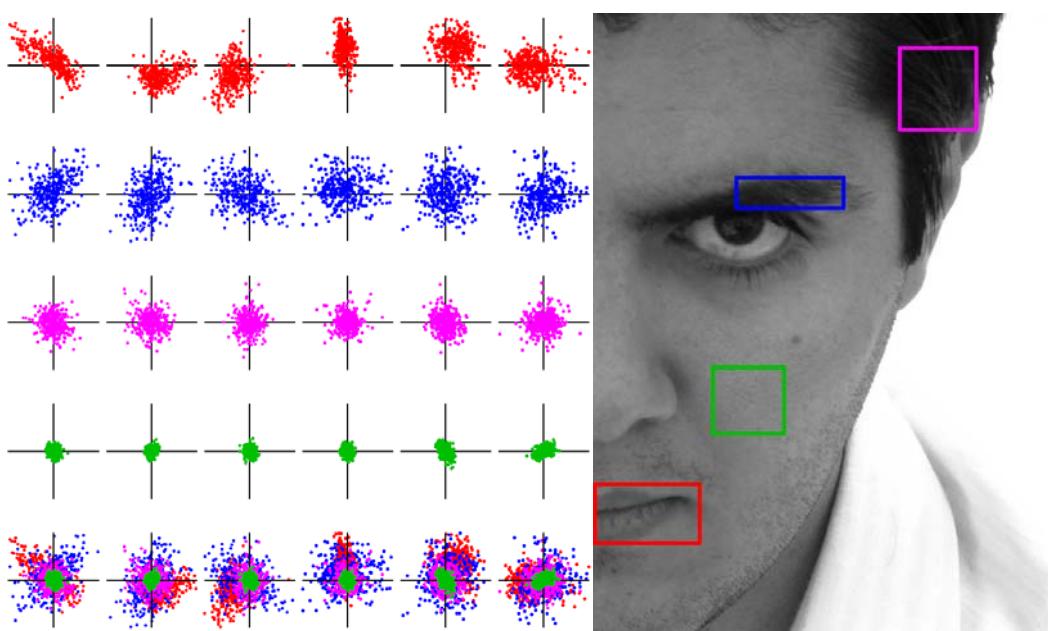
۱۵۰	تعداد نورونهای پیچیده V1 ، تعداد خروجی های واحد بازنمایی، طول کد حاصل از V1	<i>J</i>
۱۰۰۰ (فراکامل)	تعداد جهتهایی که انقباض و انبساط توزیع تکه تصویر در آنها سنجیده می شود	<i>K</i>
۱۰۰	تعداد نمونه های آموزشی در هر بسته ورودی به استنتاج	<i>N</i>
۷۰۵۰۰ تکه چهره	تعداد کل نمونه های آموزشی سیستم	<i>N_{tot}</i>
۷۰۵۰۰ غیرچهره		
۵۰	تعداد گامهای مرحله استنتاج درستنمایی بیشینه	<i>iter_y</i>
۰.۰۰۵	ضریب تغییر مقدار فعالیت نورون در هر گام آموزش	<i>ε_y</i>
۰.۰۰۱	ضریب تغییر مقدار فعالیت جهت کدگزاری توزیع در هر گام آموزش	<i>ε_b</i>
۰.۰۱	ضریب تغییر مقدار وزن بردارهای کدگزاری در هر گام آموزش	<i>ε_w</i>
۰.۰۱	نرخ تخفیف وزنهای بردارهای کدگزاری برای جلوگیری از زیاد نشدن آنها	<i>decay_w</i>
۱ (توزیع لاپلاسین)	واریانستابع احتمال اولیه گوسی که یک توزیع گوسی تعییم یافته فرض شده است	<i>var_y</i>
۱	واریانس نمودار گوسی وزن تأثیر نورونهای همسایه بر هم در لایه V2	<i>σ_{lat}</i>
۳	شعاع همسایگی	<i>R</i>
۲۰	فاصله پیکسل آغازین میدان گیرندگی دو نورون ساده لایه V1	<i>d_p</i>
۴	آستانه آتش کردن نورون در لایه V4	<i>θ</i>
۳	درشتی میدان گیرندگی نورونهای V4	<i>C</i>

جدول ۱-۴: پارامترهای مدل پیشنهاد شده و مقادیر استفاده شده در پیاده سازی نمونه

۴-۲-۲-۴- 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 بدهنند. 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 ۲-۴) و مدل غنی تری برای اخذ این همبستگی های وابسته به زمینه و در عین حال واریانس متغیر مورد نیاز است. در اینجا تعییمی 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 ۴-۴: توزیع محلی تصاویر 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 واریانسی 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 تعییه شده در هیچکدام از مدلهای مولده، مدلهای مبتنی بر انرژی^{۲۹۶}، مخلوط مقیاس گوسی، و مدلهای وابستگی موقتی^{۲۹۷} به طور صریح ساختار وابستگی نایستا داده ها را مدل نکرده است. می توان نشان داد که توزیع همبسته در بین یک مجموعه از محورها بسادگی یک توزیع ناهمریخت 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) \mathbf{D} هنگامی که توسط یک تبدیل خطی \mathbf{A} به فضای جدیدی وارد می شوند دارای کواریانس \mathbf{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. بهر حال 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 \bar{z} در نظر بگیرد:

²⁹⁶ energy-based models

²⁹⁷ temporal coherence models

$$\begin{aligned} \bar{x} | \vec{y} &\sim N(0, C) \\ C &= f(\vec{y}) \end{aligned} \quad (1-4)$$

از آنجاییکه ماتریس کواریانس مقید نیست، می تواند هر نوع تبدیل خطی را در خود جذب کند 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 بک 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 کواریانسها 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.

۲-۴-۲- کارهای پیشین

علاوه بر مدل مؤلفه های واریانسی (Karklin 2003a) که به آن اشاره شد، مدلهای سلسله مراتبی دیگری نیز واریانسها نایستا 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 نایستای درون داده های تصویری پرداخته است (Karklin 2008). با این وجود مدل کردن کواریانس و تخمین آن به زمینه های پژوهشی متعدد دیگری نظری 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 کواریانس^{۲۹۸} (Boik, 2002) و یا معکوس آن (Dempster, 1972)، فاکتور گیری از معکوس در واریانسها جزئی و همبستگیهای جزئی معکوس (Wong et al., 2003)، عاملهای Yang and Berger, 1994; Daniels (Smith and Kohn, 2002)، ماتریسهای دوران گیونز^{۲۹۹} (Daniels and Kass, 1999)، تجزیه چولسکی^{۳۰۰} (Leonard and Hsu, 1992; Chiu et al., 1996)، تبدیلات لگاریتمی ماتریسها^{۳۰۱} (and Kass, 1999)، 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 دشوارهای پارامتری سازی ماتریس کواریانس، شرط مثبت- معین بودن ماتریس است. این شرط غالباً به شرط‌های ناجوری بر روی عناصر ماتریس منتهی می‌شود. مسئله مهم دیگری در این زمینه قابلیت تفسیر آماری^{۳۰۲} است. خصوصاً برای این پژوهش که در نهایت از متغیرهای نهفته و بازنمایی مدل برای تحلیل راهکار احتمالی نورونها برای کدگزاری تصویر استفاده می‌شود، این نکته اهمیتی مضاعف می‌یابد.

یکی از مرتبط ترین کارهای پیشین، تخمین بیزی پارامترهای ماتریس کواریانس (Leonard and Hsu, 1992; Chiu et al., 1996) است. این روشها می‌کوشیدند که یکتابع احتمال اولیه معقول برای تخمین ماتریس کواریانس بدست آورند. روشی خسیسانه برای تعییه دانش اولیه برای تخمین کارای این پارامترها لازم است. قرار دادن مستقیم احتمال اولیه روی عناصر ماتریس کواریانس دردس ساز است زیرا ماتریس حاصل باید ماتریس مثبت- معین باشد. یک روش مستقیم استفاده از خانواده توزیعهای ویشارت^{۳۰۳} است که معمولاً در مدلسازی بیزی کواریانس به عنوان احتمال اولیه استفاده می‌شود (Evans, 1965)، ولی این توصیف بسیار محدودی است که تنها یک پارامتر (درجه آزادی) دارد و اجازه تعییه دانش بیشتری را نمی‌دهد.

لنونارد و هسو^{۳۰۴} 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 می‌دهد.

$$A = V[diag(d_1, d_2, \dots)]V^T \quad (2-4)$$

این به توان رساندن ماتریس توسط سری زیر تعریف می‌شود 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:

$$\exp(A) = V[diag(e^{d_1}, e^{d_2}, \dots)]V^T \quad (4-4)$$

که در آن V بردارهای یکه را جمع می‌کند و d_i مقادیر یکه A می‌باشند.

³⁰² statistical interpretability

³⁰³ Wishart family of distributions

³⁰⁴ Leonard & Hsu

تخمین پارامترها در فضای کواریانس لگاریتمی، بردارهای یکه فضای اصلی را حفظ می‌کند. اگر ماتریس به توان صفر برسد یک ماتریس همانی بدهست می‌آید که بدین معنی است که احتمالهای اولیه به مرکز صفر در فضای لگاریتمی کواریانس، یک ماتریس کواریانس همانی را ترجیح می‌دهند. این نوع از احتمالهای اولیه توسط لئونارد و هسو در سال ۱۹۹۲ پیشنهاد شده است: عناصر بالای قطر اصلی ماتریس لگاریتمی کواریانس خطی می‌شوند، و یک احتمال اولیه گوسی چندمتغیره روی عناصر قرار می‌گیرد. چیو و همکاران^{۳۰۵} این روش را با مدل کردن عناصر توسط ترکیب خطی ماتریسهای طراحی توسعه داد(Chiu et al., 1996)

$$A = \sum_j y_j A^j \quad (5-4)$$

اگر تعداد ماتریسهای طراحی اندک باشد، می‌تواند تعداد پارامترها (و متغیرهای نهفته اگر y_j ها برای نقاط داده ای مختلف متفاوت باشند) را محدود کند ولی یک توصیف قابل تفسیر از ساختار داده را مهیا نمی‌کند. در ادامه مدلی بر مبنای این کار معرفی خواهد شد ولی برای اخذ الگوی همبستگی نایستا با استفاده از پارامترهایی که کاملاً قابل تفسیر باشند توسعه داده خواهد شد.

۳-۲-۴- تعاریف

درستنمایی داده ها: ابتدا به مدل کردن داده های پیکسلی با یکتابع توزیع احتمالی گوسی چندمتغیره و یک مجموعه از توابع پایه خطی در ماتریس لگاریتمی کواریانس پرداخته می‌شود.

$$\vec{x} | \vec{y} \sim N(0, C) \quad (6-4)$$

$$C = \sum_j y_j A^j$$

هر A^j یک ماتریس مربعی با اندازه ای برابر با ماتریس کواریانس است. همانند مدلهای گذشته فرض می‌شود که میانگین توزیع این درستنمایی $p(y|x)$ صفر است. این ساده سازی برای راحت کردن محاسبات بسیار مطلوب است و می‌توان در کارهای آینده به آن پرداخت ولی این روش محاسباتی (با استفاده از شبکه نوری صعودی و تابع احتمال ثانویه متغیر نهفته) نسبت به مینیمم های محلی که از دو فاکتور میانگین و واریانس توأم 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 نایستا 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 می‌کنند).

³⁰⁵ Chiu et al

هر ماتریس متقارن را می‌توان به صورت حاصل جمع ماتریس‌هایی با رتبه یک نمایش داد. مثلاً می‌توان اینکار را با انجام یک تجزیه مقدار مفرد^{۳۰۶} و سپس محاسبه یک جمع وزن‌دار حاصل ضرب خارجی بردارهای یکه انجام داد که بردارهای یکه دارای نرم یک بوده و ضرایب آنها چیزی جز مقادیر یکه متناظر نیست.

$$A^j = \sum_k w_{jk} \vec{b}_{jk} \vec{b}_{jk}^T \quad (7-4)$$

انتخاب ضرب خارجی ماتریسها منجر به تفسیر پذیری قابل توجهی برای پارامترها می‌شود. 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 یکه یک ماتریس تأثیر می‌گذارد، تغییری در راستای مشخص شده توسط بردارها نمی‌دهد. 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 $\left\{ e^{w_{jk}} \right\}$ داده می‌شود. با این وجود اگر مجموعه بردارها متعامد نباشند یا اگر ماتریس‌های چند مؤلفه‌ای $\log(C) = \sum_j A^j = \sum_{jk} w_{jk} \vec{b}_{jk} \vec{b}_{jk}^T$ اضافه شدند که مؤلفه‌های آنها متقابلاً متعامد نبود، برهمنهی و تعامل بین بردارهای \vec{b}_k نیز باید در نظر گرفته شود. تعامل اینها با تغییر شکل بردارهای یکه دخیل در جمع نهایی است. بردارهای تقریباً متعامد جملات بینابینی^{۳۰۷} بزرگی را تولید نمی‌کند و ساختار طیفی ماتریس حاصل را چندان تحت تأثیر قرار نمی‌دهد. البته بردارهایی که کمتر متعامد هستند ساختار طیفی را طوری تغییر می‌دهد 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، عملآ فضای پارامترها کوچکتر نشده است (و همان $O(D^2)$ پارامتر را داریم). در اینجا فرض می‌شود که اگر تعداد محدودی از جهتها را به احجام متفاوت با یکدیگر ترکیب کنیم می‌تواند خیلی از الگوهای همبستگی متداولی که در صحنه‌های طبیعی دیده می‌شود 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 می‌تواند بسیار بیشتر از ابعاد داده‌ها باشد (یعنی تعداد جهتها که K است در این بازه است $\ll D^2 \ll K$). تمام مؤلفه‌های کواریانس از این گروه جهتهای ممکن برای دستکاری توزیع به طور متحددالشكل بهره می‌برند ولی هر کدام آزادند که ترکیب خطی وزن‌دار متفاوتی از این بردارها را استفاده کنند.

$$A^j = \sum_k w_{jk} \vec{b}_k \vec{b}_k^T \quad (8-4)$$

از آنجاییکه ضرایبی که A^j را ترکیب می‌کنند مستقل فرض شده اند، وزن w_{jk} بطور کارامدی تغییرات همبستگی در شکل توزیع در جهتهای \vec{b}_k را اخذ می‌کند. 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 زاویه مربوط می‌شوند و وزنهای آنها مختلف العلامه خواهد بود.

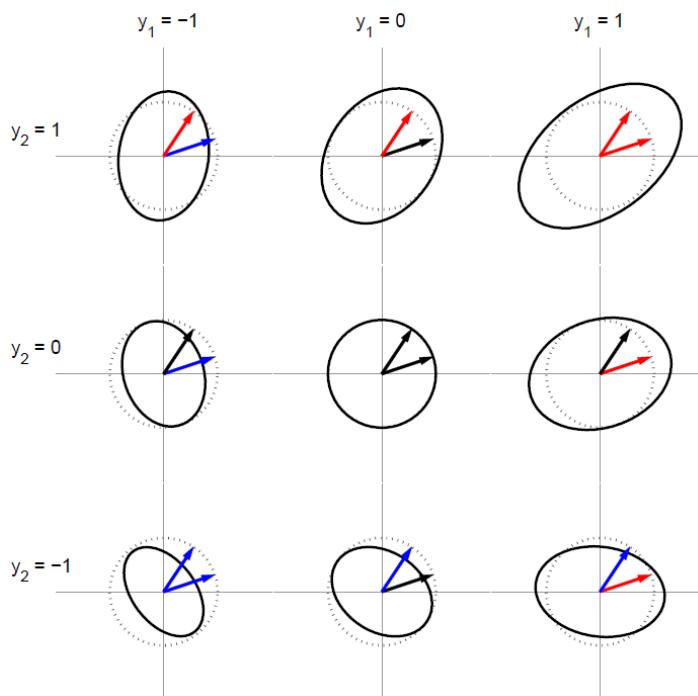
³⁰⁶ singular value decomposition

³⁰⁷ cross-terms

بنابراین ماتریس لگاریتمی کواریانس به صورت زیر تعریف می‌شود:

$$\log(C) = \sum_j y_j A^j = \sum_{jk} y_j w_{jk} \vec{b}_{jk} \vec{b}_{jk}^T \quad (9-4)$$

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 کدگزاری کند. بردارهای \vec{b}_k توزیع را در طول یک راستای خاص دستکاری می‌کند 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 وزنها جهت تأثیرگذاری را منعکس می‌کند و به یک y_j اجازه می‌دهد 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{z} ضرایب کواریانس گفته می‌شود. اثر ضرایب کواریانس بر روی توزیع توأم در شکل ۳-۴ به نمایش در آمده است. هنگامی که متغیر نهفته برابر صفر تنظیم شده یاشد، ماتریس کواریانس برابر ماتریس همانی I است که با توزیع کانونی، گوسی دایره‌ای شکل متناظر است.



شکل ۳-۴: تصویری نمادین از اثر ضرایب کواریانس بر روی ساختار کواریانس داده‌ها. هر بیضی یک کانتور دو بعدی از توزیعی واحد است که این توزیع همان گوسی دو بعدی $p(x|y, \theta)$ می‌باشد 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 تنها یکی از رو بردار b_1 یا b_2 مرتبط است. یعنی ماتریس وزنها برابر ماتریس همانی است $\mathbf{W}=\mathbf{I}_2$. ضرایب کواریانسی مثبت (قرمز) در راستای متناظر خود در فضای تصویر منبسط می‌کنند و ضرایب منفی (آبی) منقبض می‌کنند. 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 وایتنینگ 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 همانی نیست مگر اینکه تابعهای پایه خطی متعامد باشند و اگر A ماتریس فراکامل باشد اصلاً نمی تواند یک ماتریس همانی 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 \bar{b}_k استفاده شود و بازهم کواریانس پیش فرض را همان ماتریس همانی 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 کدگزاری خطی و تناظر به نورونهای نسبتاً خطی مانند سلولهای ساده در محاسبات مدل مستقر باشند. درباره این موضوع بیشتر بحث خواهد شد.

درستنمایی لگاریتمی به صورت زیر تعریف می شود:

$$\begin{aligned} L &= -\frac{1}{2} \log |\det C| - \frac{1}{2} \vec{x}^T C^{-1} \vec{x} \\ L &= -\frac{1}{2} \text{Tr} \left(\sum_j y_j A^j \right) - \frac{1}{2} \vec{x}^T \exp \left(-\sum_j y_j A^j \right) \vec{x} \\ L &= -\frac{1}{2} \sum_{jk} y_j w_{jk} - \frac{1}{2} \vec{x}^T \exp \left(-\sum_{jk} y_{jk} w_{jk} \bar{b}_k \bar{b}_k^T \right) \vec{x} \end{aligned} \quad (10-4)$$

و از آنجاییکه بردارهای \bar{b}_k دارای نرم واحد هستند، در جمله دترمینان لگاریتمی جایی ندارند.

احتمال اولیه متغیر نهفته: در این مدل متغیرهای نهفته همان ضرایب \bar{y} هستند در حالی که \bar{b}_k و وزنهای w_{jk} پارامترهایی هستند که باید تخمین زده شوند. همانند قبل می توان توزیع مستقل لاپلاسین را به عنوان توزیع اولیه \bar{y} در نظر گرفت:

$$\log p(\bar{y}) = \sum_j \log p(y_j) \propto -\sum_j |y_j| \quad (11-4)$$

یک چنین احتمال اولیه متقارنی به طور ضمنی فرض می کند که برای یک الگوی همبستگی داده شده، چنانکه با A^T تعریف می شود، ضرایب مثبت و منفی دارای احتمال یکسان هستند؛ و آشکارتر از آن فرض می کند که هردوی این مجموعه ضریب های مثبت و منفی یک مجموعه بهینه از مؤلفه های کواریانس را می سازد. چنین فرضی درباره تقارن در مدلهای 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 صورن 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 پیشرفتی برای تخمین پارامترهای مدل در آینده رفع شود. ولی در این پژوهش نتایج با همین احتمال اولیه لابلسین (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

$$\nabla_Q \exp(A) = V \left[(V^{-1} Q V) \odot \Phi(A) \right] V^{-1} \quad (12-4)$$

که در آن $A = V D V^T$ همان تجزیه طیفی است، عملگر ' \odot ' همان ضرب عنصر به عنصر (هادامارد^{۳۰۸}) دو ماتریس است و $\Phi(A)$ مشکل از

$$[\Phi(A)]_{ij} = \begin{cases} \frac{e^{d_i} - e^{d_j}}{d_i - d_j} & i \neq j \\ e^{d_i} & i = j \end{cases} \quad (13-4)$$

است که در آن d_i همان i امین عنصر ماتریس قطری D است (ر.ک. به رابطه ۱۴۴ از مقاله Najfeld and Havel, 1995) هر متغیر y 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:

$$\frac{dL}{dy_j} = \frac{1}{2} \left(\vec{x}^T \left(V \left[(V^{-1} A^j V) \odot \Phi(A) \right] V^{-1} \right) \vec{x} - \text{Tr}(A^j) \right) \quad (14-4)$$

توجه شود که این عملیات مستلزم تجزیه فرانسی $A = \sum_j y_j A^j$ به ازای هر بروزرسانی \vec{y} است و اینکار برای ماتریسها بزرگ بسیار کند است.

تقریب زنی با استفاده از بسط سریها: به منظور تسريع 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:

³⁰⁸ Hadamard

$$-\frac{1}{2} \vec{x}^T C^{-1} \vec{x} = -\frac{1}{2} \vec{x}^T \exp(-A) \vec{x} = -\frac{1}{2} \left(\vec{x}^T \vec{x} - \vec{x}^T A \vec{x} + \frac{1}{2} \vec{x}^T A A \vec{x} - \frac{1}{6} \vec{x}^T A A A \vec{x} + \dots \right) \quad (15-4)$$

اگر \mathbf{B} را ماتریسی حاوی بردارهای \vec{b}_k و \vec{w}_j را بردار ستوانی وزنهای $w_{j1}, w_{j2}, \dots, w_{jK}$ در نظر بگیریم، می‌توان گرادیان تقریب درستنمایی لگاریتمی (تعریف شده توسط سری) را به صورت زیر داشت:

$$\frac{d\hat{L}}{dy_j} = \frac{1}{2} \vec{w}_j^T \left(-1 + (B^T \vec{x})^2 - (B^T \vec{x}) \odot (B^T A \vec{x}) + \frac{1}{3} (B^T \vec{x}) \odot (B^T A A \vec{x}) - \dots \right) \quad (16-4)$$

که \vec{A} 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 $B\vec{x}$ ، به توان دو رساندن حاصل، و مقایسه آن با یک، یعنی علماً رابطه $-1 - (B\vec{x})^2$ هم ارز است. سپس حاصل مقدار \hat{z} را بر حسب وزن متناظر \vec{w}_j بالا یا پایین می‌برد. جملات مرتبه بالاتری که شامل جمع کامل درون A می‌شوند ($B A \vec{x}$ وغیره) کواریانسی که باعث پدید آمدن وضعیت کنونی \hat{z} 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 وزنهای 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 شبکه نوردهی سعودی 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{z} را نمی‌توان بدست آورد اما چند روش دیگر برای اینکار موجود است.

کواریانس لگاریتمی تک نمونه^{۳۰۹}، یک روش استفاده از تخمین تک-نمونه‌ای کواریانس

³⁰⁹ single sample log-covariance

$$\hat{C} = \vec{x}\vec{x}^T \quad (17-4)$$

و محاسبه مجموعه ای از ضرایب است که این تخمین را بخوبی تقریب می‌زند، به طوری که $\hat{y} \approx \log(\hat{C})$. به حال کواریانس تک نمونه یک ماتریس مفرد است و تنها یک بردار یکه دارد با مقدار یکه غیرصفر دارد (در طول بردار \vec{x}). تمام ابعاد دیگر دارای حجمی نیستند و یک کدگزاری 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 جرمیمه کردن مقادیر بزرگ \bar{y} از واگرایی جلوگیری می‌کند.

اگر به ماتریس کواریانس تک-نمونه، یک ماتریس کواریانس یکریخت با واریانس کوچک (ε) بیفزاییم^{۳۱۰}،

$$\hat{C} = \vec{x}\vec{x}^T + \varepsilon I \quad (18-4)$$

مقدارهای یکه این ماتریس برابر $(\|\vec{x}\|^2 + \varepsilon, \varepsilon, \dots, \varepsilon)$ خواهند بود و بردارهای یکه آن شامل $\|\vec{x}\|/\varepsilon$ و یک مجموعه از بردارهای متعامد خواهند بود. لگاریتم ماتریس دارای بردارهای یکه مشابهی خواهند بود و مقدارهای یکه آنها عبارتست از $(\log(\|\vec{x}\|^2 + \varepsilon), \log \varepsilon, \dots, \log \varepsilon)$. از آنجایی که مدل یک مجموعه از بردارهای پایه خطی در فضای ماتریس‌های لگاریتم کواریانسی را به خدمت می‌گیرد، می‌توان مجموعه ای از ضرایب را محاسبه کرد که این ماتریس لگاریتم کواریانسی غیر مفرد کواریانسی را بهترین تقریب از آنرا بدست می‌دهد تخمین زده می‌شود: را تقریب بزند:

$$\sum_j \hat{y}_j A^j \approx \log(\hat{C}) \quad (19-4)$$

ابتدا توابع پایه برداری می‌شود ($\vec{a}_j = \text{vec}(A^j)$ ، سپس ماتریس لگاریتمی کواریانس برداری می‌شود و در نهایت بردار \bar{y} که بهترین تقریب از آنرا بدست می‌دهد تخمین زده می‌شود:

$$\begin{aligned} [a_1; a_2; \dots; a_J] \vec{y} &= \text{vec}(\log \hat{C}) \\ \hat{y} &= [a_1; a_2; \dots; a_J]^\dagger \text{vec}(\log \hat{C}) \end{aligned} \quad (20-4)$$

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 فضای کاملاً درست نباشد چون تغییر در برخی از عناصر کواریانس لگاریتمی، درستنایی لگاریتمی را بیش از بقیه تحت تأثیر قرار می‌دهد.

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 می‌شوند)، جملات بیتابیینی را نادیده گرفته، و با جملات درستنایی لگاریتمی جاهایی کار داشت که توان ماتریس به صورت عنصر به عنصر ارزیابی می‌شود که اینکارها بجای تکیه بر محاسبه توان ماتریس انجام می‌شود. ماتریس قطری Z با عناصر $z_{kk} = \sum_j y_j w_{jk}$ و ماتریس B بردارهای ستونی \vec{b}_k را جمع می‌کند. آنگاه:

³¹⁰ augment

$$L = \log p(\vec{x} | \vec{y}) = -\frac{1}{2} \text{Tr}(Z) - \frac{1}{2} \vec{x}^T \exp(-B Z B^T) \vec{x} \quad (21-4)$$

اگر ماتریس B متعامد یک^{۳۱۱} باشد (که در واقع مؤلفه های برداری آن نرمال شده هستند ولی دقیقاً متعامد نیستند) آنگاه می توان رابطه را اینگونه بازنویسی کرد:

$$\begin{aligned} L &= -\frac{1}{2} \text{Tr}(Z) - \frac{1}{2} \vec{x}^T B \left[\text{diag}(e^{-z_{11}}, e^{-z_{22}}, \dots) \right] B^T \vec{x} \\ L &= -\frac{1}{2} \sum_j y_j \sum_k w_{jk} - \frac{1}{2} \sum_k \left(e^{-z_{kk}} [B^T \vec{x}]_k [B^T \vec{x}]_k \right) \end{aligned} \quad (22-4)$$

مشتق نسبت به y_j برابر است با

$$\begin{aligned} \frac{\partial L}{\partial y_j} &\propto -\sum_k w_{jk} + \sum_k \frac{\partial z_{kk}}{\partial y_j} e^{-z_{kk}} [B^T \vec{x}]_k^2 \\ \frac{\partial L}{\partial y_j} &\propto -\sum_k w_{jk} + \sum_k w_{jk} e^{-z_{kk}} [B^T \vec{x}]_k^2 \end{aligned} \quad (23-4)$$

با مساوی قرار دادن این رابطه با صفر رابطه زیر برای مقدار بهینه \hat{y}_j بدست می آید:

$$\sum_k w_{jk} = \sum_k w_{jk} e^{-z_{kk}} [B^T \vec{x}]_k^2 \quad (24-4)$$

یکی از جوابها وقتی بدست می آید که $e^{-z_{kk}} [B^T \vec{x}]_k^2 = 1$. این منجر به تخمین زیر می شود:

$$\begin{aligned} e^{z_{kk}} &= [B^T \vec{x}]_k^2 \\ z_{kk} &= \log([B^T \vec{x}]_k^2) \\ [W\vec{y}]_k &= \log([B^T \vec{x}]_k^2) \\ \hat{y} &= W^\dagger \log([B^T \vec{x}]_k^2) \end{aligned} \quad (25-4)$$

که در آن W ماتریس تمام وزنهای w_{jk} است و W^\dagger شبیه معکوس این ماتریس است. این تقریب مدل مؤلفه های کواریانسی را به فرم مدل مؤلفه های واریانسی (Karklin 2003a) فرموله می کند که در آن واریانس های لگاریتمی در راستاهای \vec{b}_k توسط یک تبدیل خطی، در اینجا W ، تعریف می شوند.

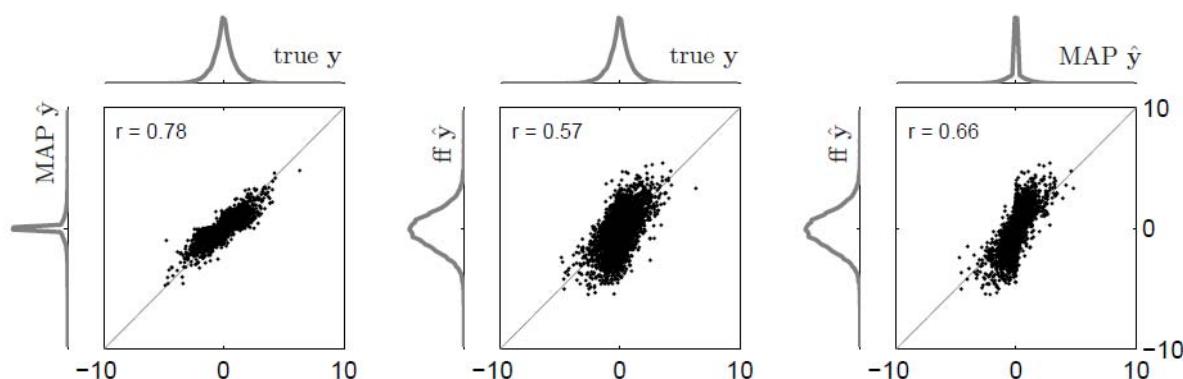
هیچ یک از این دو کدگزاری تقریبی پیشخور احتمال اولیه پراکنده روی \vec{y} را پشتیبانی نمی کند. برای محاسبه تقریبی پیشخوری که پراکنده هم باشد، می توان روش های خطی با قیدهای پراکنده ای استفاده می شوند را

³¹¹ orthonormal

به خدمت گرفت؛ روش‌هایی مانند روش‌های تعقیب پایه^{۳۱۲} (Chen et al., 2001) یا تقریب زنای کوچک کننده^{۳۱۳} (Hyvärinen, 1999). این روشها با در نظر گرفتن یک جریمه درتابع هدف برای راه حل‌های ناپراکنده الگوریتم را متمایل به یافتن پاسخهای پراکنده تر می‌نمایند. نمونه‌ای از این جریمه‌ها در زیر آمده است:

$$L = \left\| \log([B^T \vec{x}]^2) - W \vec{y} \right\|^2 + \phi(\vec{y}) \quad (26-4)$$

تقریب زنای بالا (بدون قید پراکنده‌گی) روی داده‌های 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 نگاشت مستقل و محاسبه واریانس لگاریتمی (رابطه ۲۶-۴)، مقادیر MAP را بسیار بهتر از تخمین کوارینس لگاریتمی تک نمونه‌ای (رابطه ۲۰-۴) تقریب می‌زند (اثبات نظری این نتایج به کارهای آینده موکول می‌شود. احتمالاً تقریبهای کم خطأ در فضای عناصر ماتریس کواریانس لگاریتمی در ساختار مسئله منعکس نمی‌شود). همبستگی‌های بین تخمینهای MAP و مقادیر محاسبه شده MAP در تکرار در شکلهای ۴-۴ و ۵-۴ آورده شده است.



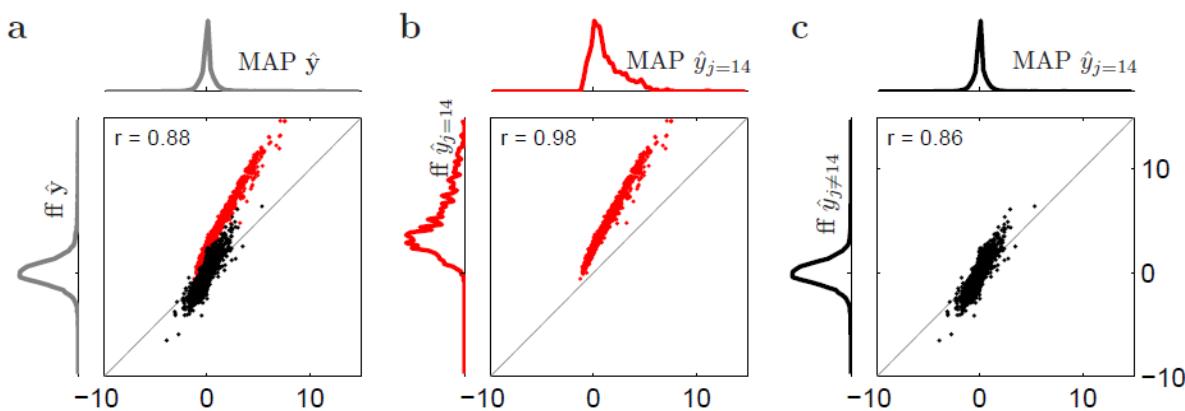
شکل ۴-۴: مقادیر متغیرهای نهفته واقعی در مقایسه با تخمینهای MAP (\hat{y}) و تخمینهای پیشخور آنها (ff \hat{y}) در یک مجموعه داده ساختگی (که مستقیماً از مدل نمونه گیری شده است) ابعاد داده‌ها ۹۹، ابعاد y برابر ۲۰ و پارامترهای مدل از یک توزیع نرمال استاندارد استخراج شده‌اند. در شکل مقایسه تخمین MAP با مقادیر واقعی (چپ)، تقریب پیشخور با مقادیر واقعی (وسط)، و تقریب پیشخور با تخمین MAP (راست) دیده می‌شود. ضرایب همبستگی نیز در شکل آورده شده است. مقادیر MAP با مقادیر واقعی بسیار همبسته‌اند و چنانکه انتظار می‌رفت بسیار پراکنده‌اند (توزیع MAP نباید با تابع احتمال اولیه مطابق باشد زیرا مقدار بیشینه احتمال ثانویه را اخذ می‌کند. تخمین پیشخور با مقادیر MAP همبسته است (راست) ولی از تخمین MAP و مقادیر واقعی بسیار کمتر است).

برای داده‌های ساختگی حتی می‌توان 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 ۷۸٪ است) در حالی که تناظر بین تخمین MAP و تقریب پیشخور آن کمی معتلتر است و برابر ۶۶٪ می‌باشد (شکل ۴-۴ راست). همانگونه که انتظار می‌رفت، توزیع مقادیر

^{۳۱۲} basis pursuit methods

^{۳۱۳} shrinkage estimator

MAP از یکتابع احتمال اولیه لاپلاسین 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 آموزش داده شده مشاهده گردید ولی ساختار پارامترهای یاد گرفته شده به مقید کردن تقریبهای تخمین پیشخور کمک می کند و در عین حال بایاس را معین می کند. به طور متوسط، تخمین پیشخور بسیار با مقادیر MAP همبسته است (شکل ۴-۵-چپ). بهر حال 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 نورده صعودی روی تکه تصویرهای 10×10 بدست آمده است. از آنجایی که مؤلفه کواریانس متناظر با تغییرات عمومی (DC، واحد واریانس) 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 اربیت تر) است (شکل راست) زیرا بازنمایی مدل برای بازنمایی الگوهای کواریانسی 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 قیدهای پراکندهی و محسوب کردن وابستگی بین مؤلفه های نامتعامد کواریانس بهبود می یابد. توزیع نامتنازن ضرایب DC کواریانس نشان می دهد که احتمال اولیه لاپلاسین برای واحد کنتراست DC مناسب نیست. در مدلها خطی مانند ICA، اثر مشابهی برای ضرایب روشنایی DC^{۳۱۴} 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 یکتابع اولیه انعطاف پذیر برای این ضرایب، و یا با اصلاح مدل برای مستثنی کردن این راستا از فضای ماتریسیهای کواریانس بهبود داد.

^{۳۱۴} 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 تصاویری که در فوتوای می افتد- ماهیتی سریع دارد و برای سیگنالها زمانی اندکی برای انتشار با 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 تجمعی کنند که این اطلاعات ممکن است توسط اتصالات بازدارنده محلی دروازه بندی شوند. بر اساس رابطه ۲۵-۴ این محاسبات عبارت است از نگاشت بر روی یک مجموعه از ویژگیهای خطی $\vec{x}^T B$ که تقریباً با فعالسازی سلولهای ساده در V1 متناظر است (و ویژگیهای \vec{b}_k نماینده میدان گیرنده 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 $\log([B^T \vec{x}]^2)$ نیز بزرگتر خواهد بود و بر عکس هر چه کوچکتر باشد، منفی تر خواهد بود. این اطلاعات توسط نورونهای بعدی، آر، به تناسب قدرت اتصال که با ماتریس W^* نشان داده می شود وزن گذاری شده و سپس جمع آوری می شود. این وزنهای شبیه چه هستند؟ در این پیاده سازی، ماتریس W در تمام فضای ویژگیهای خطی گستردۀ نشده است (تعداد ابعاد کواریانس معمولاً بسیار کمتر از فضای توسعه یافته \vec{b}_k هاست) و عملگر شبه معکوس ماتریسی را می سازد که ورنها آن به طور کیفی مشابه ماتریس اصلی است.

هر پراکنده سازی که جزوی از مدل باشد ولی در مکانیسم استنتاج آورده نشده باشد، از بازدارنده‌ی جانی بین نورونهای مرحله دوم نشأت می گیرد.

۴-۵-۲-۴- تخمین پارامترها

روش تخمین بیشینه درستنمایی: پارامترهای مدل با بیشینه کردن درستنمایی داده ها بیشینه می شود. با در نظر گرفتن :

$$\theta = \left\{ w_{jk}, \vec{b}_k \right\}$$

(Karklin 2003a)، درستنمایی بیشینه (ML) اینگونه حساب می شود

$$\hat{\theta} = \arg \max \langle p(\vec{x}_n | \theta) \rangle_n \quad (27-4)$$

(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. ایندا باید روش متغیرهای نهفته آر توزیع حاشیه را بدست آورد:

$$p(\vec{x} | \theta) = \int p(\vec{x} | \theta, \vec{y}) p(\vec{y}) d\vec{y} \quad (28-4)$$

به حال به خاطر شکل بسیار غیرخطی درستنمایی و نیز احتمال اولیه غیرگوسی ($\bar{y}|p$ ، 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

$$p(\vec{x} | \theta) \approx p(\vec{x} | \theta, \hat{y}) p(\hat{y}) \quad (29-4)$$

$$\hat{y} = \arg \max p(\vec{y} | \vec{x}, \theta)$$

همانند بالا، می‌توان بجای استفاده از تقریب از روش‌های نمونه‌گیری 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 می‌سپریم.

برای بدست آوردن تخمین‌های درستنمایی بیشینه از شبیه نوردی صعودی استفاده شده است. همانند تخمین متغیرهای نهفته، می‌توان از تعریف رسمی گرادیان 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. گرادیانهای تقریبی که در شبیه سازیها استفاده شده است و بر بسط سریها استوار است به شرح زیر است (اثبات در (Karklin 2007):

$$\frac{\partial \hat{L}}{\partial b_k} = \left(\sum_j w_{jk} y_j \right) \left(I + \vec{x} \vec{x}^T - \frac{1}{2} (\vec{x} \vec{x}^T A + A \vec{x} \vec{x}^T) + \frac{1}{6} (\vec{x} \vec{x}^T A A + A A \vec{x} \vec{x}^T + A \vec{x} \vec{x}^T A) + \dots \right) b_k \quad (30-4)$$

$$\frac{\partial \hat{L}}{\partial w_{jk}} = \frac{1}{2} \left(-1 + (b_k^T \vec{x})^2 - (b_k^T \vec{x})(b_k^T A \vec{x}) + \frac{1}{3} (b_k^T \vec{x})(b_k^T A A \vec{x}) + \frac{1}{6} (b_k^T A \vec{x})^2 - \dots \right) y_j$$

برای مجموعه تصاویر، گرادیانها برای هر تکه تصویر ارزیابی می‌شوند و میانگین آنها گرفته می‌شود. اینکار درستنمایی لگاریتمی مورد انتظار روی تمام مجموعه داده‌ها را بیشینه می‌کند. پیش از هر گام شبیه نوردی، بردارهای \bar{b}_k نرمال‌سازی می‌شوند تا از برقراری $\|\bar{b}_k\| = 1, \forall k$ اطمینان حاصل گردد. نرم بردار وزنهای هر متغیر نهفته، $\|w_j\|$ ، به طور تدریجی تنظیم می‌شود تا واریانس مطلوب زلرا حفظ کند. اگر به وزنهای اجازه تغییر آزادانه داده شود، تخمین MAP برای ساخت توزیع حاشیه‌ای متغیر نهفته به یک وضعیت منحصراً شود که در آن پارامترها بزرگ و بزرگتر می‌شوند در حالی که واریانس متغیر نهفته کاهش می‌یابد.

۴-۲-۶- رابطه با مدل مؤلفه‌های کواریانسی

با وجود اینکه این مدل به بازنمایی ماتریسهای کواریانس می‌پردازد، رابطه نزدیکی با مدل مؤلفه‌های کواریانسی (Karklin 2003a) دارد و در حقیقت ایده اولیه تولید این مدل از تعمیم مدل قبل بدست آمده است. تغییر واریانس در یک بعد اگر فضا دوران داده شده باشد، به مثابه معرفی جملات کواریانس بین هر دو بعد است. این مدلها در فضای وايتنینگ شده اعمال می‌شوند (بنابراین توزیع کانونی دارای آمارگان f/I است)، لذا واریانس متغیر در این مدل روی مقدار پیکسلها تأثیری نمی‌گذارد و روی جهت‌هایی کار می‌کند که ترکیبات پیکسلها دیکته می‌کنند.

اگر مدل واریانسی دارای توزیع گوسی شرطی باشد یعنی

³¹⁵ Markov chain Monte Carlo

$$p(s_i | \vec{v}) = N(0, \exp([B\vec{v}]_i)) \quad (31-4)$$

آنگاه توزیع در فضای داده ها پس از نگاشت روی توابع پایه A اینگونه خواهد بود:

$$p(\vec{x} | \vec{v}) = N(0, A \exp([B\vec{v}] A^T)) \quad (32-4)$$

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 $C = \sum_{jk} y_j w_{jk} \vec{b}_k \vec{b}_k^T$ برابر خود \vec{b}_k می باشد و فرمول به توان رساندن ماتریس کواریانس تا حد به توان رساندن عنصر به عنصر مقادیر یکه ساده می شود که $C = B[e^{W\vec{y}}]B^T$. 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 یکریخت 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.

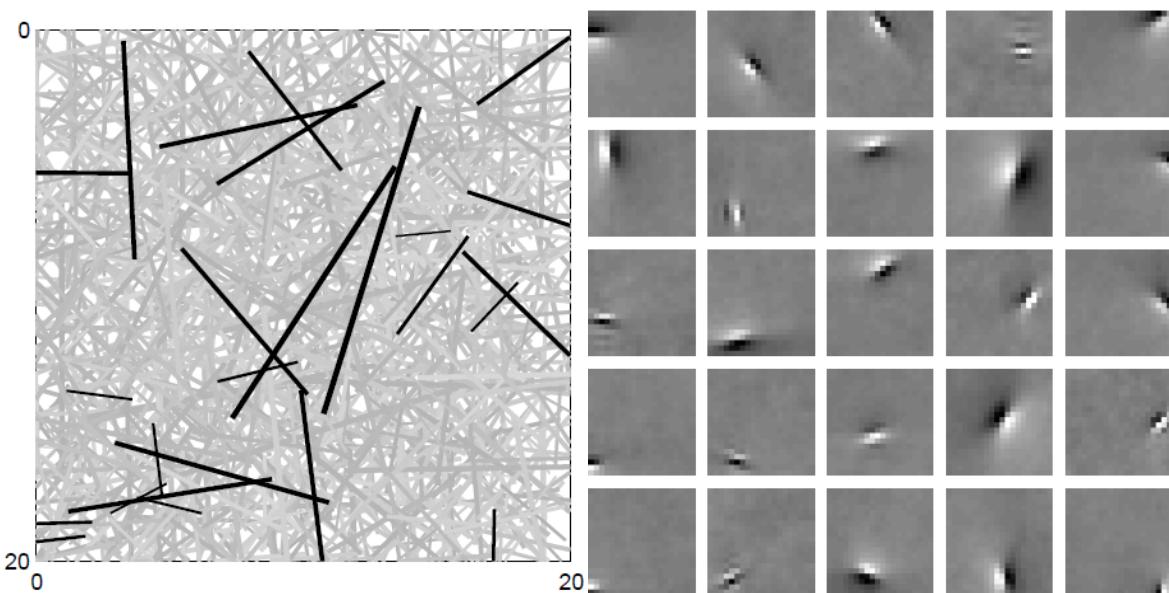
یک مزیت دیگر مدل کواریانسی اینست که به طور ضمنی توزیع حاشیه ای ضرایب خطی \vec{a} را محاسبه می کند. در مدل واریانسی امکان محاسبه توزیع حاشیه ای احتمال ثانویه $p(\vec{v} | \vec{x})$ وجود دارد. البته بجای اینکار از تخمین MAP ضرایب خطی و واریانس ضرایب مؤلفه ها استفاده می شود. یک ضعف احتمالی برای مدل مؤلفه 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.

۴-۳-۱- نتایج روی تصاویر طبیعی

۴-۳-۲- 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 ($(\gamma_0, \gamma_1, \dots, \gamma_L)$) در مدل برابر ۱۵۰ می باشد و تعداد ویژگیهای تصویر که برای توصیف توزیع استفاده می شود (\vec{b}_k) برابر ۱۰۰۰ است. با آزمایشها 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 درستنمایی 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 وزنهای w_{jk} که نورونها را به آنها مرتبط می کنند هم‌زمان بهینه سازی می شوند تا توزیع تصویر را بخوبی بازنمایی کنند.

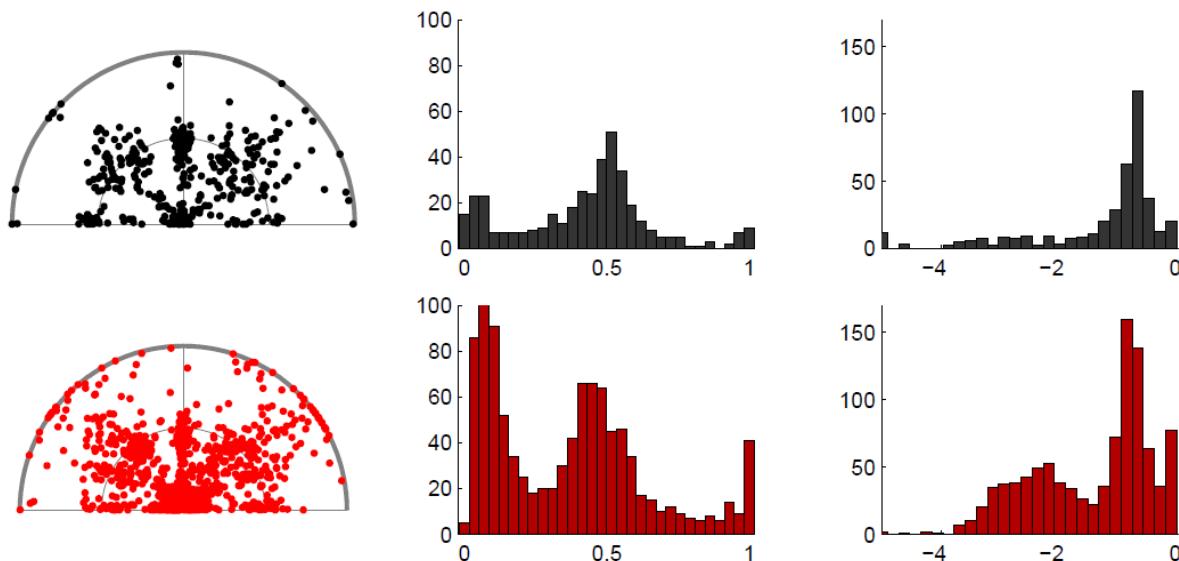
همانگونه که گفته شد مجموعه ۱۱۰ عکس خاکستری فام از صحنه های بیرون از منزل (van Hateren and van der Schaaf, 1998) به عنوان داده های آموزش استفاده می شود، که این انتخاب به طریقه سنتی و بر اساس روال (Hyvarinen 2000, Schwartz 2001, Karklin 2003a, Doi 2005, Karklin 2008) مقاله های معتبری مانند (Karklin 2003a, Doi 2005, Karklin 2008) صورت گرفته است. شدت روشنایی پیکسلها ابتدا تبدیل لگاریتمی می شود (مشابه تبدیلی که در سلولهای مخروطی شبکیه 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 می‌یابد، بردارهای b_k که توسط مدل برای توصیف توزیع تصویر استفاده می‌شوند، 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 می‌دهد: پیکسلهایی که رنگ مشابهی دارند، مثلًاً دو پیکسل سفید همبسته ترند و پیکسلهایی که رنگهایی متفاوت دارند ضد همبستگی بیشتری دارند. (چپ) مجموعه کامل ویژگیهای تصویر که در پهنهای 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 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 \bar{b}_k جهت انبساط و انقباض کلی شکل توزیع تصویر را مشخص می‌کند. شکل ۴-۶-راست یک زیرمجموعه نمونه را پس از آموزش نشان می‌دهد. 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 می‌شوند (شکل ۴-۶-چپ) و تمام راستاها و فرکانسهای مکانی تصاویر طبیعی را پوشش می‌دهند (در شکل نیامده است). این ویژگیهای تصویری راستدار و باندگذر با تصویرهای که باعث بیشترین تحريك سلولهای ساده درون بخش اولیه قشر بینایی مغز می‌شود 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 می‌کرد (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 کدگزاری شده به کار گرفته شده است (پیکانها در شکل ۳-۴). بنابراین با وجود اینکه تفاسیر قدیمی از کدهای حسگری اولیه حکایت از این داشتنند که که این کدها برای بازسازی منصفانه محرک سازگار می‌شوند، این نتایج بدست آمده تفسیر جدیدی را به آنها می‌افزاید: این کدها تغییرات در توزیعهای تصویر را انتقال می‌دهند و به ناحیه‌های تصویری بعدی قشر مخ امکان تشکیل بازنمایی‌های مجردتر را می‌دهند.

بررسی مجموعه ویژگیهای خطی همچنین نشان داده است که همانند مدل سلسله مراتبی واریانس (Karklin 2003a)، خصوصیات جمعیتی این ویژگیها با جمعیتی که از مدلها خطي مانند ICA بدست می آید متفاوت است. مجموعه کامل بردارهای \bar{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 ۷-۴).



شکل ۷-۴: 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 حرک را به صورت خطی کد نمی کند، برای آن فیلترهایی معادل فیلترهای مدلها خطی نمی توان متصور شد.

این مدل یک مزیت احتمالی دیگر را نیز برای بازنمایی فراکامل فراهم می کند. واحدهای مدل توزیع تصویر را کدگزاری می کنند و برای توصیف این توزیعها به ویژگیهای خطی تکیه می کنند (شکل ۳-۴). انواع مختلف کواریانس به مجموعه متفاوتی از جهتهای زیرین احتیاج دارند تا توزیع آنها دستکاری شود. برای مثال یک مؤلفه کواریانسی که تنها محتوای فرکانسی تصویر را تغییر می دهد باید مجموعه های ویژگیهای خطی \bar{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.

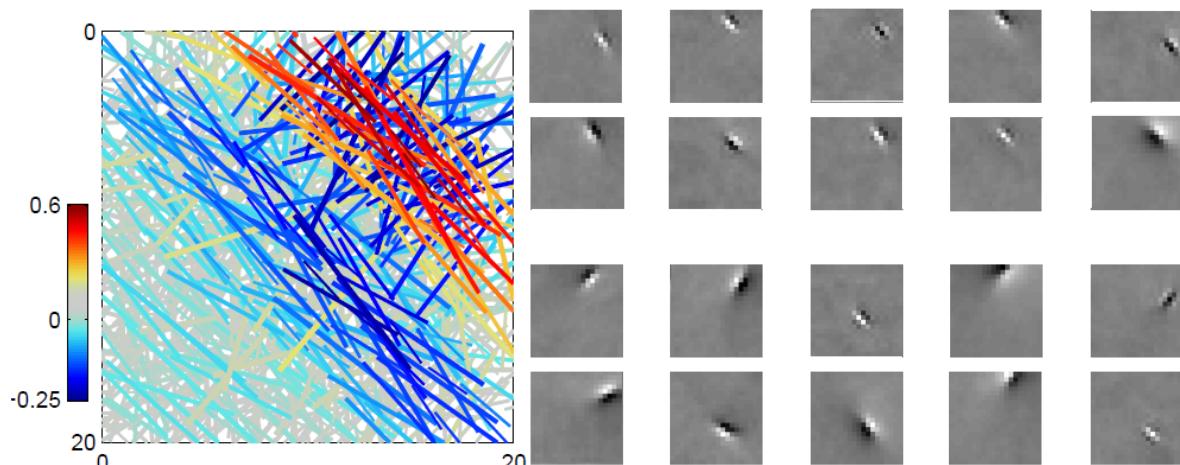
برای اینکه یک واحد دیگر بتواند تغییرات در محتوای راستهای تصویر را مستقلًا کدگزاری کند، باید به مجموعه ای از ویژگیهای خطی که تمام راستهای گوناگون را پوشش می دهند دسترسی داشته باشد. یک پایه کامل، که توسط ICA یا یک مدل سلسله مراتبی با تعداد محدودی ویژگی ساخته می شود، حاوی مجموعه کاملی از ویژگیها نیست. از نگاهی دیگر می

^{۳۱۶} Nyquist limit

توان دید که ویژگیهای راستدار و محلی برای کدگزاری کلارای 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

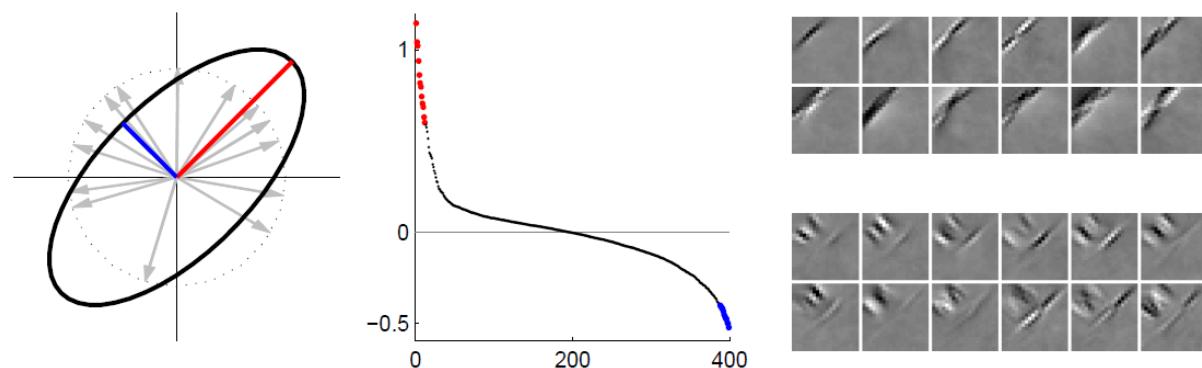
دومین مجموعه پارامترها، وزنهایی است که نورونهای مدل را به مجموعه بردارهای \tilde{b}_k مربوط می کند و نقش هر نورون را در شکل دهی به توزیع کدگزاری شده تصویر معین می کند. یک مجموعه از وزنهای آموزش دیده برای یک نورون مدل معمولی در شکل ۴-۸ راست آمده است. این نورون قویترین اثر خود را برای ویژگیهایی که در سمت راست بالای تکه تصویر قرار دارند نشان می دهد، فعالیت نورون برای آنها 45° - دارند (راستای «مرجح» نورون) افزایش می یابد و برای آنها می که در راستای عمود قرار دارند و یا هم راستا هستند ولی موقعیت آنها متفاوت است فعالیت کمتری دارد. بجای اینکه نورون تنها به یکسری ویژگیهای تصویری فرونشاننده 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 کدگزاری شده است، آنها همچنان ویژگیهایی که برای بیشترین و کمترین تحریک این نورون لازم است نشان می دهند.



شکل ۴-۸: (چپ) وزنهای یک نورون معمولی برای بردارهای b_k که در شکل ۴-۷ ترسیم شده بودند. (به ترتیب) رنگها نمایانگر علامت و بزرگی وزنهای سمت (به نمودار رنگی توجه شود). وزنهای مثبت (رنگهای گرمتر) نشانگر افزایش تغییر در ویژگی متناظر هستند. وزنهای 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 جایجا 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 کدگزاری می کند فعال می شود. بر عکس محرکهایی که در قسمتی از فضای تصویر قرار دارند که که این نورون برای آنها احتمال کمی قائل است از فعال شدن آن جلوگیری می کنند.

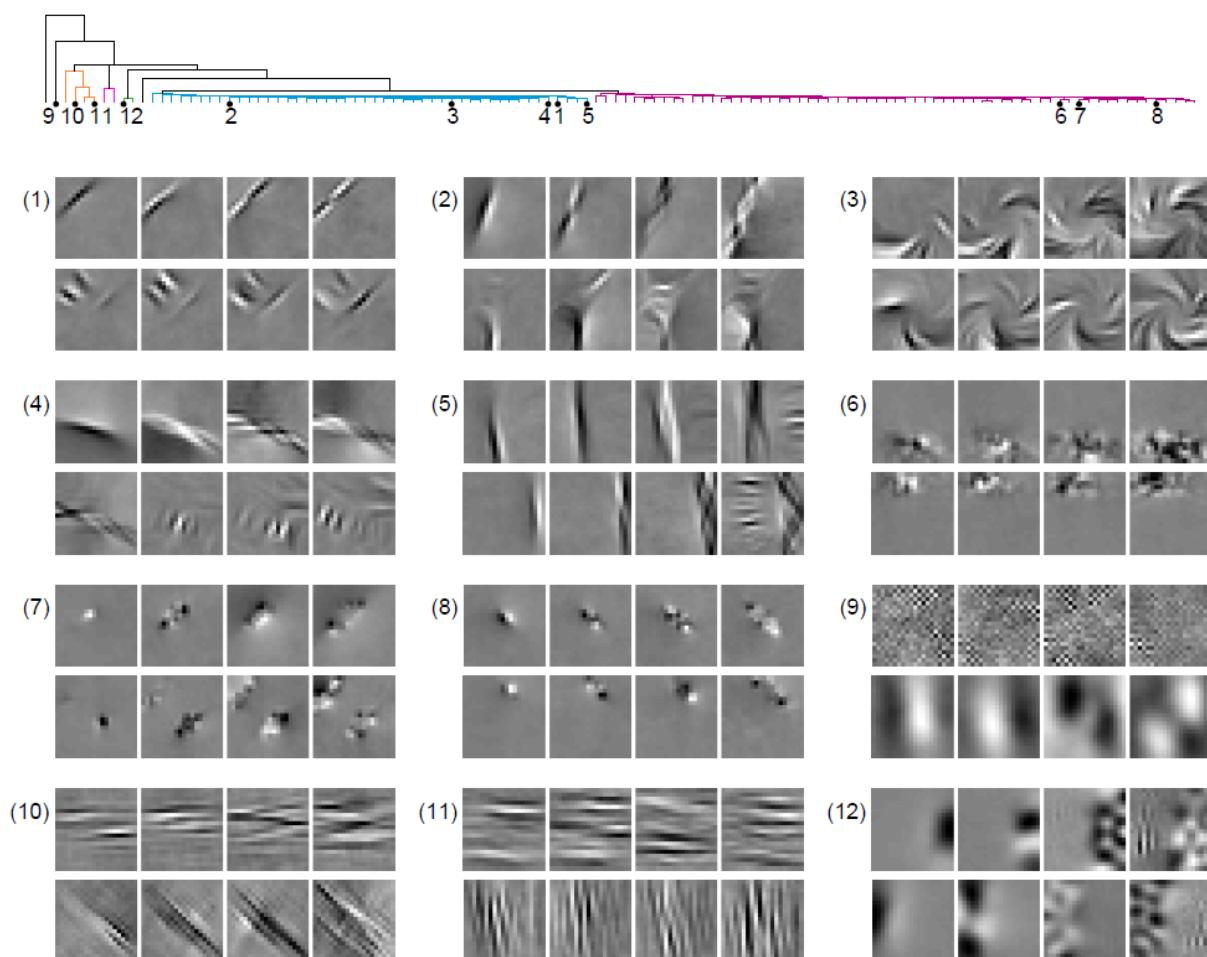
این مدل اجازه محاسبه تحلیلی راستایی که بیشترین انگیختگی و بیشترین فرونشانی برای نورون دارد را می دهد که این انگیزش و فرونشانی با توجه به قدر مطلق مقادیر \mathbb{Z}^d تعریف می شود. با \mathbb{A} قرار دادن فعالیت نورون \mathbb{Z}^m و صفر قرار دادن $C = \exp(A^j)$. تحلیل طیفی این ماتریس نشان می دهد که ابعاد بزرگترین انقباض یا بیشترین انبساط عبارتند از بردارهای یکه متناظر با راستای بزرگترین بردارهای یکه فضای تصویر که در آن واریانس بیشترین افزایش را نسبت به توزیع کلی (وقتی که $0 = \bar{y}$ است) دارد. مقادیر یکه نشانگر اینست که توزیع فشرده شده است. این تحلیل برای یک نورون مدل در شکل ۹-۴ و برای بخشی از جمعیت در شکل ۱۰-۴ پایین آمده است.



شکل ۹-۴: (چپ) شمایی از اثر یک نورون بر توزیع کدگزاری 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 یکه متناظرند (خطوط آبی و قرمز) (Karklin 2008). (وسط) مجموعه کامل ۴۰۰ مقدار یکه مقیاس همه جهتهای در فضای تصویر را توصیف می کند. (راست) بردارهای یکه ای که مثبت ترین و منفی ترین مقدارهای یکه را دارند به شکل تصویری به ترتیب در بالا و پایین تصویر قرار دارند. این مقدارهای یکه کرانی در تصویر میانی مشخص شده اند.

۳-۴-۳- کدگزاری جمعیتی

برای هر تصویر ورودی، فعالیت توازن جمعیتی از نورونهای مدل، توزیع تصویر استنتاج را توصیف می کند. برای فهم این کدگزاری جمعیتی باید از تحلیل خوشها برای شناسایی نورونهایی با عملکرد یکسان استفاده کرد (شکل ۱۰-۴ بالا) و سپس به بازنمایی نورون نماینده آنها نگاه کرد (شکل ۱۰-۴ پایین).



شکل ۱۰-۴: تحلیلی از جمعیت واحدهایی که مدل یاد گرفته است. (بالا) ۱۲۰ تا از فعالترین نورونها (از مجموع ۱۵۰ تا) که بر اساس جنبه‌ای از تصویر که کدگزاری می‌کنند به طور سلسله مراتبی خوش بندی شده‌اند. این خوش بندی دو رسته بزرگ و چند دسته خاص از نورونها را مشخص می‌کند (پایین) برای تهیه یک توصیف آگاهانه از هر نورون، ویژگیهایی از تصاویر که بیشتر از همه آنها را برانگیخته یا فرونشانی می‌کنند تشخیص داده شده‌اند. در این تصویر دوازده نورون نمونه از جمعیت یاد گرفته شده انتخاب شده‌اند. برای هر یک از نورونها چهار ویژگی تصویری برانگیخته کننده در سطر بالا و چهار ویژگی تصویری فرونشاننده در سطر پایین آورده شده‌اند. اعداد مشخص شده جای نورون در دندوگرام شکل بالا را نشان می‌دهد.

برای خوش بندی جمعیت نورونها، بررسی شده است که کدام پارامتر از ویژگیهای تصویری \bar{b}_k بهتر می‌تواند مقدار وزنهای هر نورون، w_{jk} ، را خوش بندی کند. به عنوان مثال نورون شکل ۸-۴ به ساختار راستدار و محلی حساس است و لذا وزنهای آن در برابر ویژگیهای زیرین تصاویر (رنگها در شکل ۸-۴ چپ) در بهترین حالت توسط مکان و راستای ویژگیهای \bar{b}_k توصیف می‌شوند (۹۳٪). واریانس در w_{jk} ها با رگرسیون روی این دو پارامتر تعریف می‌شوند. برای هر نورون، یک بردار تعريف شده است که میزان مشارکت پارامتر ویژگیها (مکان، راستا، فرکانس، و تمام ترکیبات آنها) را در توصیف وزنهای نورونها *loc-or*, *loc-fr*, *or*, *fr*, *loc*, *loc-or-fr* و *fr-or* 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 (معمولًاً ۵ ظرف) تقسیم می‌شوند و از مقدار میانگین هر ظرف (آبرظرف 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 *fr* تنها آنرا 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 نورونها استفاده شده است (Hegd'e and Van Essen, 2003; Gallant et al., 1996). ولی استفاده از این روش وقتی عملکرد نورون از پیش معلوم نیست و محرك غنی و پیچیده‌ای مانند تصاویر طبیعی موجود است، دشوار است. یک روش معمول ساختار میدان گیرنده‌گی نورون را با توجه به راستای غالب و اندازه گیری وسعت تنظیم راستا، گستردگی ویژگیهای برانگیخته کننده و فرون Shanandeh، و موقعیت مکانی مورد ارزیابی قرار می‌دهد. برای مدل فعلی این روش تنها برای برخی از نورونها خوب کار می‌کند. بررسی شهودی این نکته را نشان می‌دهد که تنظیم راستا و گستردگی فرون Shanani برای توصیف این جمعیت ناهمگن کافی نیستند و سایر گزینه‌های تحلیل نیز به انتظاری که از رفتار نورون می‌رود وابستگی زیادی دارند. با این وجود تحلیلی که از پژوهش‌های فیزیولوژیکی برآمده باشد ممکن است به درد اینکار بخورد، مخصوصاً وقتی که تناظر نورونهای قشر مخ و بازنمایی مدل مورد بررسی قرار گیرد.

خوشه بندی جمعیت دو دسته بزرگ و چند دسته خاص از نورونها را مشخص می‌کند و جمعیت نورونها تعدادی از ویژگیهای مشاهده شده در سلولهای بینایی قشر مخ را نیز به نمایش می‌گزارند. یک مجموعه بزرگ از نورونهای حساس به راستا و محلی تشخیص داده شده اند (به عنوان مثال نورون شکل ۴-۸ چپ). بسیاری از آنها پدیده بازداری از فعالیت به ازای راستاهای متقطع 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 نورونها تصاویر متنوعی را کدگزاری می‌کنند، که برخی از آنها انحناء و ویژگیهای پیچیده تری از آنها را شامل می‌شود (الگوهای ۵ الی ۹).

مجموعه بزرگ دیگری از نورونها 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 ارضاء 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, 1998b; Mareschal and Baker, 1994) سازگار است.

در بین نورونهای مدل برخی محتوای فرکانس مکانی تصویر را تحلیل می کند (مثالاً 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 ارضاء می کنند.

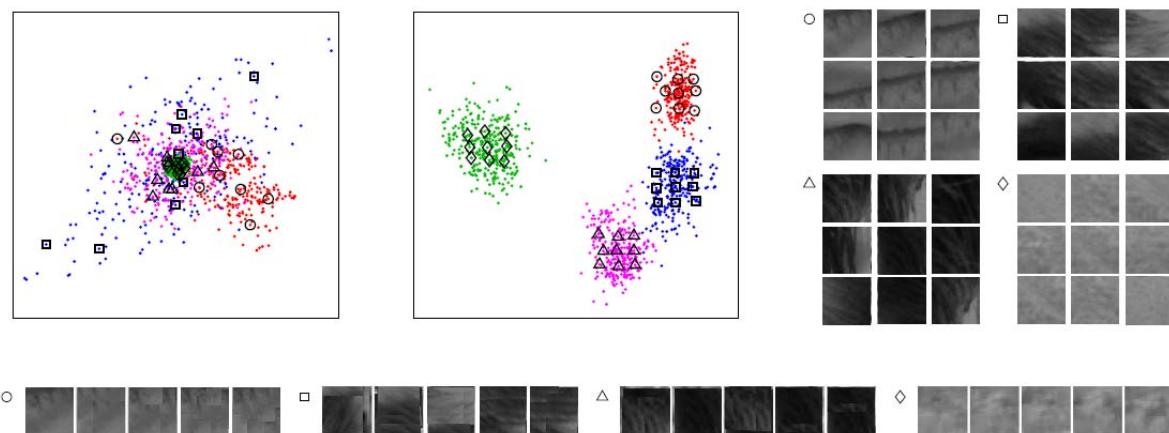
سایر نورونهای جمعیت ساختار عمومی راستا را منتقل می کنند (الگوی ۲ و ۳) ولی به محتوای فرکانس مکانی آن حساس نیستند. این خصوصیات کدگزاری در نورونهای V4 دیده شده اند که برخی از آنها کاملاً برای راستا تنظیم شده اند در حالی که برخی دیگر اطلاعات فرکانسی را کدگزاری می کنند(David 2006). سایر نورونهای مدل نشانگر 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 کدگزاری مرز بافت‌ها در قشر بینایی مخ به محركهای 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-بعدی مدل از مجموعه تصاویر روی دو بعد (چنانکه بری فضای تصویر در شکل ۲-۴ انجام شد) روشن شود. شکل ۱۱-۴ بالا-وسط دوتا از خطی- جدایی پذیرترین ابعاد را در فضای آراها برای چهار نوع تصویر نشان می دهد (مثالها در شکل ۱۱-۴ 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 یکه ($S_w^{-1} S_b$) 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 متوجه این تصاویر دارای لبه های عمودی تاریک-روشن قابل درکی است). نگاشت مجموعه آرایی 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 کدگزاری توزیع تصویر بجای ویژگی‌های دقیقی در هر تصویر، مدل می‌تواند تصاویری که از لحاظ مفهومی به هم شبیه‌ند را با بازنمایی شبیه‌ی کدگزاری کند و نوعهای متمایز تصویر را از هم جدا کند.



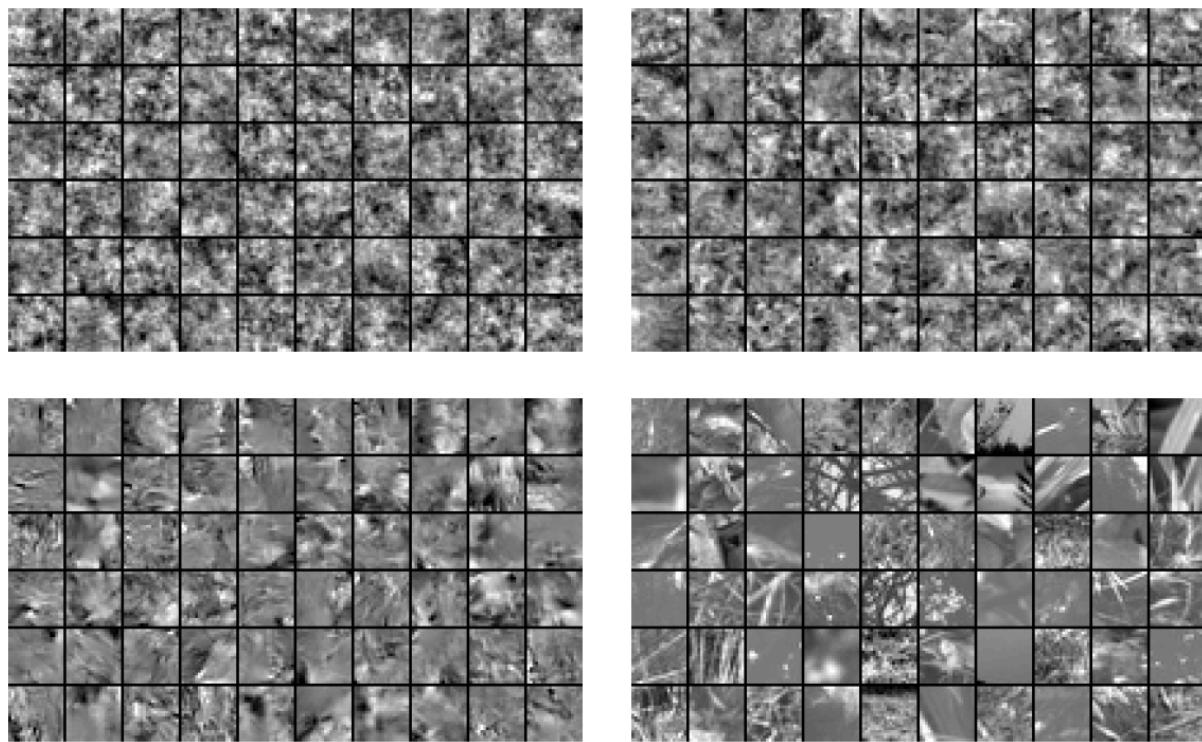
شکل ۱۱-۴: بازنمایی مدل توانسته است روی تنوع نواحی چهره تعمیم یابد و بین انواع مختلفی از انواع تصویری تمایز قائل شود. (بالا چپ) بازنمایی خطی تصاویری که از چهار ناحیه شکل ۴-۲ استخراج شده اند همپوشانی زیادی دارند (رنگها مشابه تصویر ۴-۲) (بالا وسط) یک افکنش دو بعدی از بازنمایی مدل گروه جدا پذیری از خوشه‌ها 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 بافت‌هه است ولی همچنان آنها را تمایز می‌دهد. (پایین) تکه‌های تصویری که از بازنمایی مدل از هریک از چهار نوع نمونه کمیر 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 کدگزاری می‌کند آزمود. میانگین تخمینهای MAP برای هر ناحیه تصویر توزیع «مرکزی» متناظر با هر نوع تصویر را بدست می‌دهد. پس باید مقدار \bar{z} را برابر با این مقدار قرار داد و تکه‌های تصویر را نمونه برداری کرد (شکل ۱۱-۴ 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 تکه تصویرهای 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 ۱۲-۴ تصویری 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 ۱۲-۴: تکه های تصویری نمونه گرفته شده از مدلها مولّد گوناگون. این تکه های 20×20 از مدلها ICA (چپ بالا)، 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)

مدلهای 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 ار 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. به حال 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

۱-۴-۴- شرح مسئله

روش کدگزاری مؤلفه های کواریانسی، که یکی از مدلهاست 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 روشانی پیکسل (آمارگان مرتبه اول) در آن خلی وارد نمی کند. همینطور عدم وابستگی نوع کدگزاری 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 کدگزاری 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 خوشة 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{z}_j ، بازنمایی نهایی تصویر را بر اساس مؤلفه های آموزش دیده w_{jk} و b_k بدست می دهد. این مدل در لایه آخر خود دارای یک نورون است که باید میزان شباهت به چهره را در بازه $(0,1)$ 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 سرتا 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 150×150 نورون به لایه اول شبکه عصبی وارد می کند. شبکه عصبی با استفاده از قانون پس انتشار خطای 20×20 نورون لایه میانی خود، و با توجه به اینکه تکه تصویر از مجموعه آموزشی مثبت یا منفی بوده است، وزنهای شبکه را آموزش می دهد. تکه های تصویر غیرچهره ای که میزان فعالیت نورون لایه آخر برای آنها از آستانه 5.0 بالاتر باشد، الگوهای «شبکه 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 وزنهای شبکه ثابت می شود و خروجی شبکه احتمال اینکه یک تکه تصویر چهره یک انسان اخذ شده باشده را در خروجی خود نشان می دهد.

برای آزمودن برتری کدگزاری حاصل در مرحله قبل ابتدا باید به کاهش ابعاد کدگزاری حاصله از 400×400 بعد به 150×150 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 400×400 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 واینتینگ 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 متقطع 10 بار تا کردن³²²، میانگین درصد تشخیص برای شبکه ها به ترتیب 89.2% ، 89.5% و 56.8% 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 کدگزاری) بوده است. در این آزمایش تشخیص دادن چهره در شبکه معادل بالاتر بودن میزان فعالیت نورون لایه آخر از آستانه ای برابر 5.0 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 گرینه 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 احتمالهای بالا برای تصویر غیرچهره و احتمالهای پایین برای چهره اجتناب ناپذیر است. شکل ۴-۱۳ میزان کارایی این روش را با توجه به آزمون بازیابی نشان می دهد.

³²² 10-fold cross validation



شکل ۱۳-۴: آزمون بازیابی تکه تصویرهای مشابه (راست) تصویر بزرگنمایی شده پروب (چپ) ۵ تکه تصویر دارای کمترین فاصله اقلیدسی به بازنمایی پروب

۴-۳-۴-۴ مدل بازنمایی های همخوان

در ادامه روند اطلاعات در مدل، اکنون داده های کدگزاری شده در لایه دوم هستند. برای روشن شدن مدل باید توجه داشت که میدان گیرندگی سلولهای عقده ای شبکیه 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 کدگزاری می کند و ممکن است این کدگزاری وابسته به ویژگیهایی از تصویر باشد که در میدان گیرندگی نورون مجاور قرار نگیرد (مانند یک لبه عمودی که نورون همسایه افقی ممکن 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 باعث فرونشانی آن نیز بشود. به همین دلیل نورونهای همسایه ممکن است دارای کدگزاریهایی 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 خطأ در یک نورون متصور است که همانگونه که ذکر شد استفاده از داشت موجود در همسایگی هم خالی از دردسر نیست.

برای تصحیح این مشکل، از مکانیسم اتصالات جانبی در یک لایه استفاده می‌شود. در این مکانیسم نورونها در لایه به صورت کاشی کاری فرض می‌شوند و هر نورون با تمام نورونهای دیگر اتصالات مستقیم برقرار می‌کند. این اتصالات در یک شعاع همسایگی محدود هستند (وزن اتصالات در بیرون از این شعاع صفر فرض می‌شود). وزن اتصالات درون شعاع همسایگی از یک توزیع گوسی تبعیت می‌کند که با افزایش فاصله اقلیدسی دو نورون در نقشه توپوگرافیک وزن آنها، w_{ij}^{lat} هم کاهش می‌یابد:

$$w_{ij}^{lat} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\|i-j\|^2}{2\sigma^2}\right) \quad (33-4)$$

پارامترهایی که در این لایه روی نتیجه تأثیرگزار 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:

$$y_i = y_i^{bup} + w_{ij}^{lat} y_j^{bup}, \forall j \in N_i, j \neq i \quad (34-4)$$

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 می‌توان شعاع یا واریانس آنرا صفر در نظر گرفت و عملاً تأثیر اتصالات جانبی را حذف کرد. با توجه به استفاده ازتابع 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 می‌کند و اگر فاصله آنها زیاد باشد (معمولًاً 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 داده‌ها در لایه V2، اکنون داده‌های کدگزاری شده به صورت فعالیت نورونهای این لایه هستند. با توجه به اینکه مقادیر بیشتر در این لایه، نشانگر احتمال بیشتر حضور چهره در آن می‌باشند، کافیست در میدان گیرندگی نورون لایه بالاتر که معادل V4 است، حوزه این مقادیر زیاد مشخص شود.

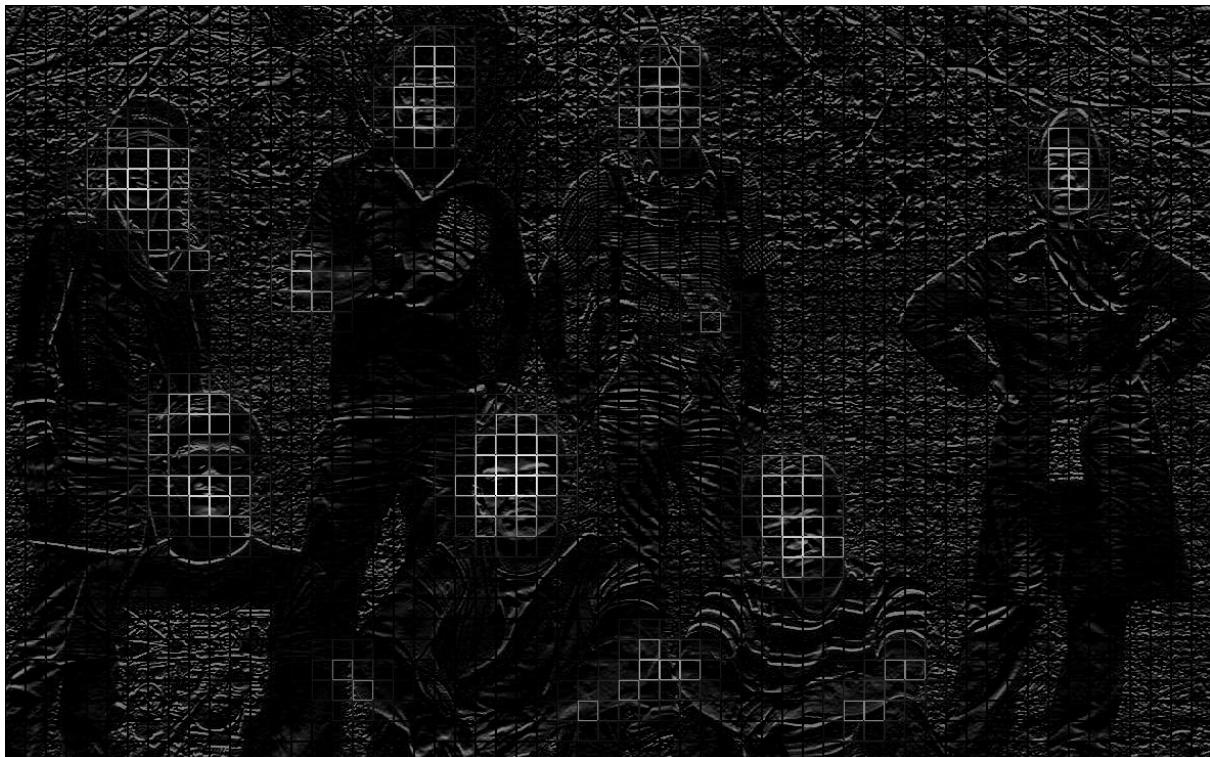
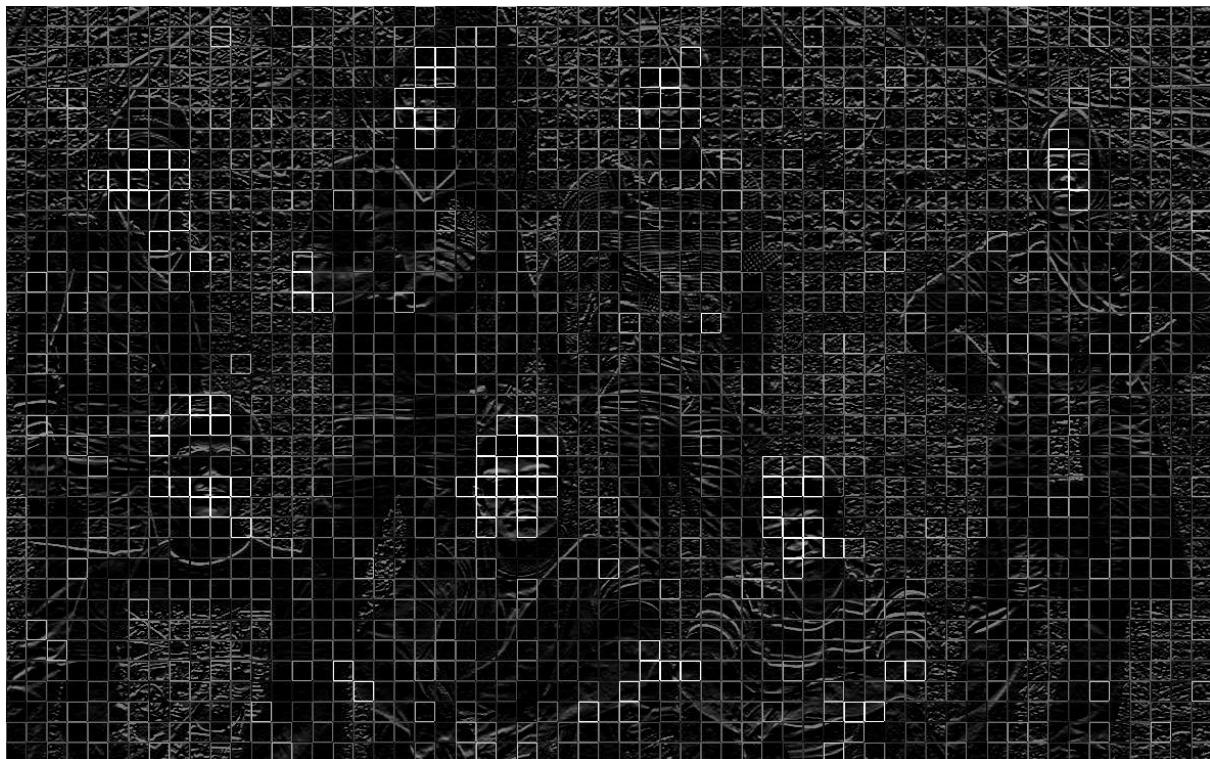
برای اینکار سه راهکار می‌توان در نظر گرفت: در راهکار نخست می‌توان از شبکه عصبی برای بیشینه گیری روی فضای نورونهای لایه پایین استفاده کرد که اینکار خوبترین کاندید برای چهره بودن را نشان می‌دهد. عیب این روش اینست

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 ۷۴ 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 پیتر^{۳۲۳}) با تابع فعالسازی پله، بایاس مشخص، و وزنهای برابر یک است. چنین نورونی در میدان گیرندگی خود که توسط پارامتر درشتی، محدوده آن مشخص می شود، میزان احتمال تمام نورونها 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 نورون آنرا میدان گیرندگی خود می دارد و یک شدن آن برابر احتمال حضور چهره در همان مربع است. نورونهای این لایه دارای همپوشانی است به طوری که تفاوت میدان گیرندگی یک نورون با نورون سمت راست آن تنها در سمت چپترین تکه های تصویری نورون سمت چپ و سمت راستترین تکه های تصویری نورون سمت راست می باشد.

مراحل آشکار سازی چهره در شکل ۱۴-۴ 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 400×400 ورودی شدت روشنایی به لایه نخست وارد می شود. در این لایه با توجه به جهتهای \vec{b}_k و وزنهای 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 نورونهای لایه اول دیده می شود. در لایه دوم این بازنمایی که توسط 150×150 نورون نشان داده می شود، با استفاده از شبکه پرسپترون چند لایه که وزنهای 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 نورون به ازای هر تکه تصویر یافت می شود. سپس در این لایه اتصالات جانبی بر فعالیت نورونها اثر می گذارد 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 نورون فعل است.

³²³ Mc Culloch & Pitts







شکل ۱۴-۴: 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 رویرو مشخص شده است. تصویر دوم پس از پیش پردازش توسط مدل سلوهای 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 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 می‌توان تمام محاسبات درون یک لایه را بطور موازی انجام داد که سرعت را بطور چشمگیری افزایش می‌دهد، پیاده سازی موازی این سیستم به کارهای بعده موکول می‌شود.

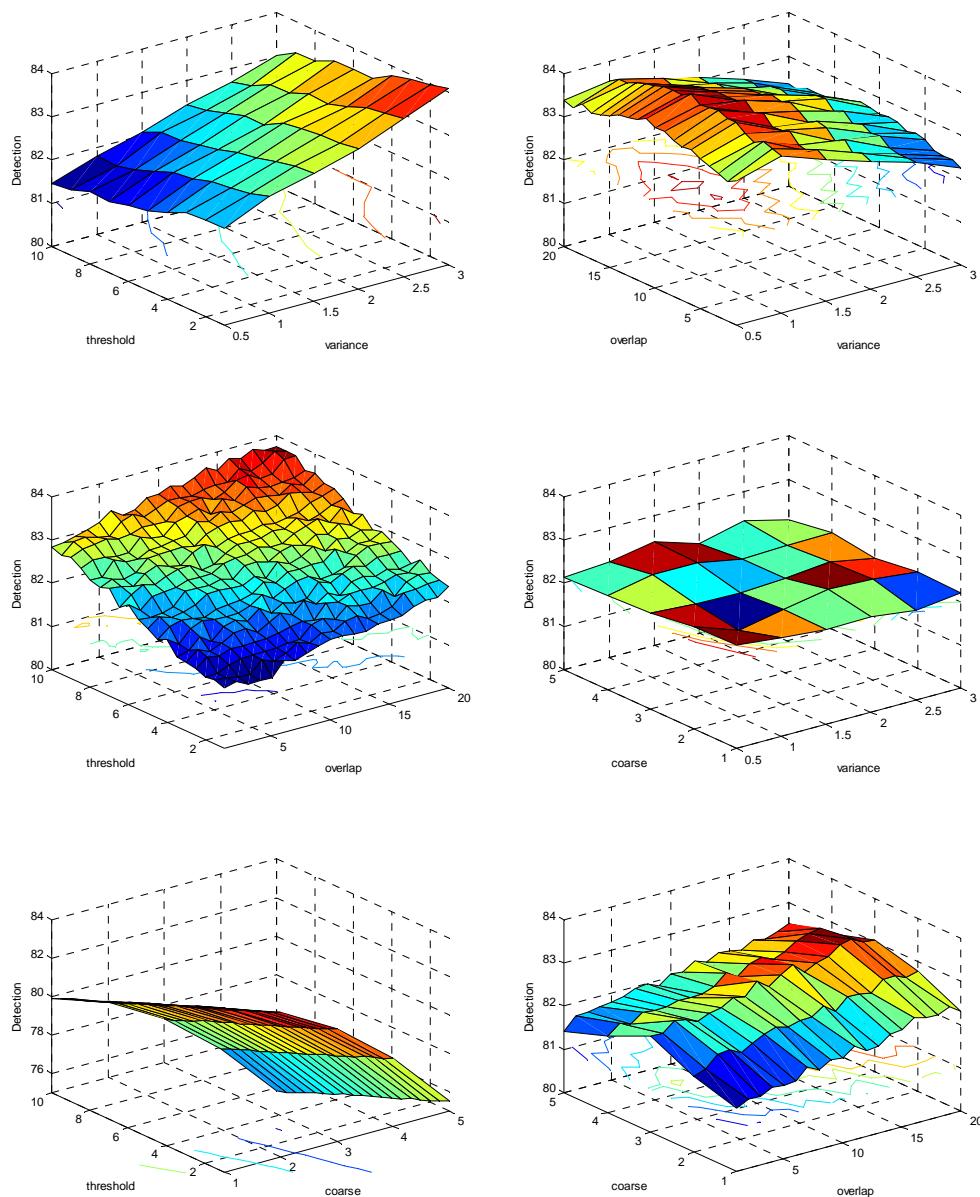
۴-۵-۱- تأثیر پارامترها روی نرخ آشکارسازی چهره ها

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 واوه شده است تا صحت آشکارسازی چهره تایید شود (شکل ۱۵-۴). 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 ۱۵-۴: نمونه ای از مجموعه داده های BioID Face Database

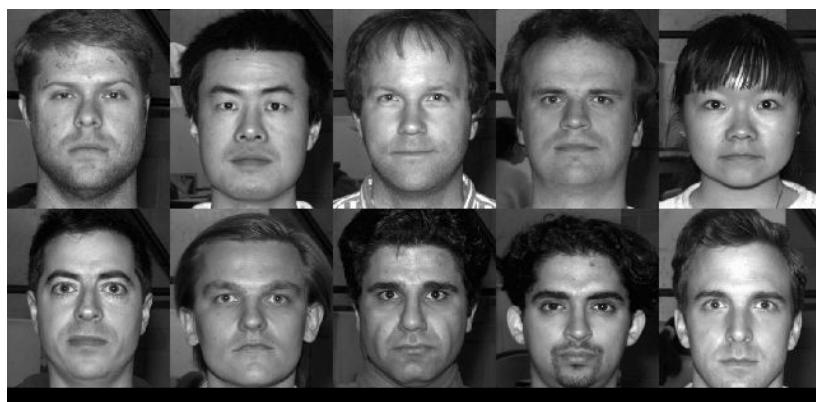
در شکل ۱۶-۴ تأثیر تغییرات پارامترها (واریانس اتصالات جانبی، همپوشانی میدانهای گیرندگی اولیه، آستانه تایید کاندید چهره، و درشتی میدان گیرندگی نورونهای سطح بالا) روی عملکرد مدل برای تشخیص چهره در تصاویر ساده چهره/غیر چهره نشان داده شده است.



شکل ۱۶-۴: 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 تأثیر 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.

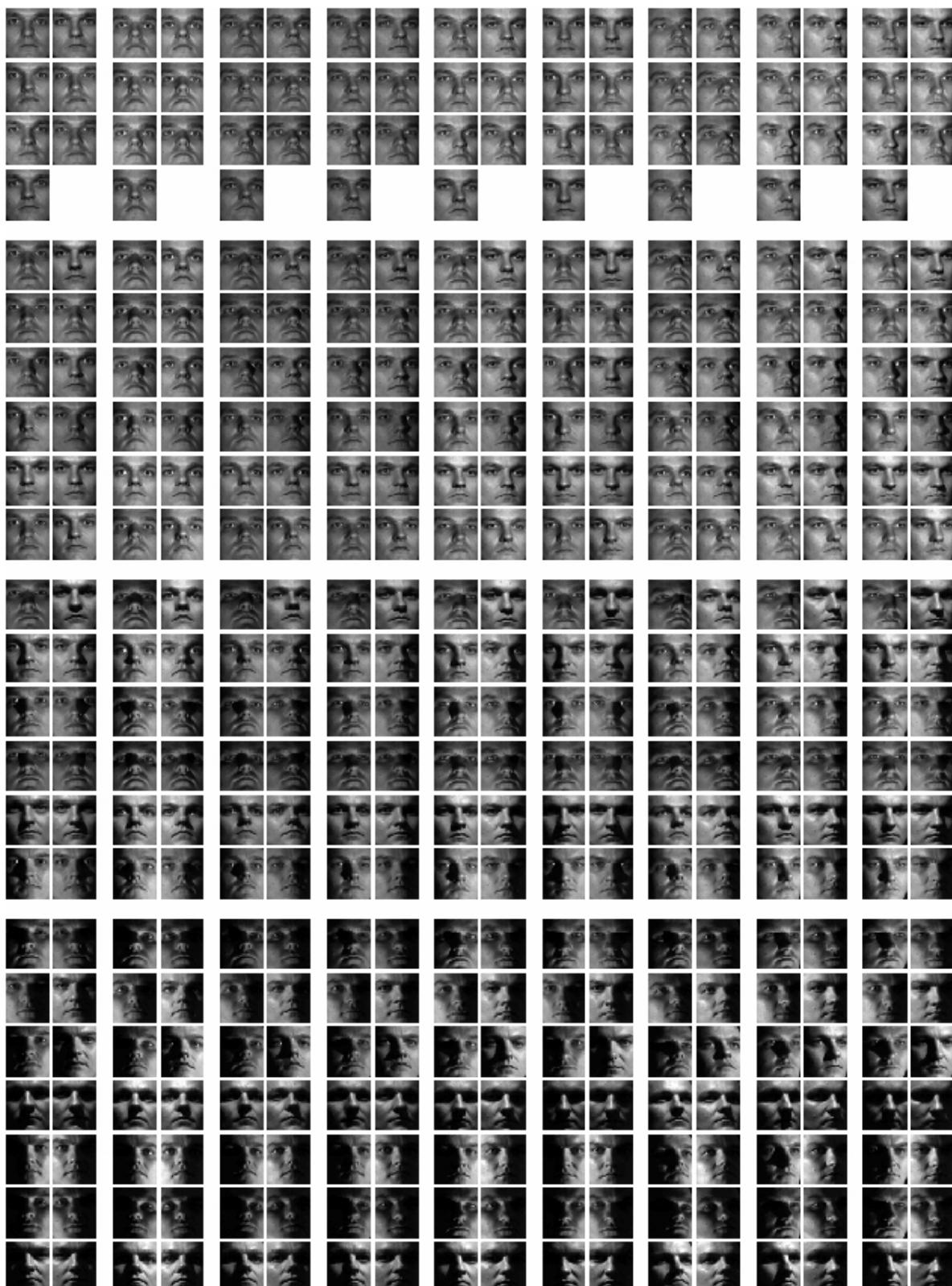
۴-۵-۲- آشکارسازی چهره در شرایط ویژه نورپردازی

در آزمایش روی تصاویر چهره دارای نورپردازیهای متفاوت، نرخ آشکارسازی افت چندان زیادی 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 میماند. 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 ($10 \times 64 \times 9 = 5760$ ، 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 بدليل 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 دقق 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 (JAFFE) Database (شکل ۱۹-۴). 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، باز هم 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 ۱۹-۴: نمونه ای از مجموعه داده های JAFFE

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 اشتباه مثبت (حد آستانه پایین) و یا اشتباه ممنفی (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 اشتباه 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.

فصل پنجم

[جمع بندی]

۱. جمع بندی
۲. مسئله ها و راه حلها
۳. مقایسه با سایر روشها
۴. کارهای آتی

۱-۵- جمع بندی

در این پایان نامه چندین موضوع مستقلًا مورد بررسی قرار گرفت و سپس از ترکیب آنها و تلفیق ایده های برخاسته از هر یک آنها روشنی 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 گردند.
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 بخش های 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 یزگی از ادبیات عصب شناسی به معرفی این پدیده ها و معرفی مدل هایی برای آنها پرداخته است که تا حدودی بتوانند این پدیده ها را به همراه سایر پدیده ها توجیه کنند که تا کنون مدل های یکپارچه با پیدا شدن مثال های نقض و رد شدن پیش بینی های آنها راجع به نحوه مواجهه دستگاه عصبی شکست خورده اند و تنها مدل هایی در گوش و کثار 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 سلول های ساده ناحیه V1 به فیلتر های گابور اشاره کرد که اولی با کشف تعاملات و مسیر های پسخور بین نواحی قشر بینایی کم اعتبار شده است و دومی از توجیه رفتارهای غیر خطی نورون عاجز مانده است. البته بد لیل ذات فرایند مدل سازی که ماهیت آن ساده سازی پدیده های پیچیده است این مدلها همچنان مورد استفاده قرار می گیرند و بر اساس آنها کشف های جدید و پیشرفت های علمی زیادی صورت می گیرد، ولی باید در نظر داشت که به دلیل ناقص بودن پیش فرض های چنین دستاوردهایی، آنها نیز به نوعی خود دارای کاستی هایی هستند و تنها به صرف توجیه پدیده های بیشتر از مشاهدات نمی توان یک مدل را بر دیگری برتری داد.
۳. بررسی قاعده مندی های 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 سازگاری هایی 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 همین رو باید این محیط نیز مورد بررسی قرار بگیرد. این محیط دارای قاعده مندی هایی درون خود است و از نظم خاصی تبعیت می کند. ادبیات زمینه تحقیقاتی فهم تصاویر طبیعی، به بررسی

این قاعده مندیها می‌پردازد و از آنها در جهت ساخت سیستم‌های بینایی بهره می‌برد، چنانکه انتقال تصاویر تلویزیونی به عنوان یکی از ارکان ارتباطی روزمره انسانها نتیجه چنین بررسی هایی است. طبق یکی از معروفترین نظریه‌ها درباره سازگاری سیستم بینایی با محیط اطراف، سیستم بینایی در بخش اولیه خود فرایندی را آغاز می‌کند 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 می‌کنند (نظیر تنظیم ساعت بیولوژیکی، تخمین سرعت یک جسم نزدیک شونده، استخراج اطلاعات مهم از میان سیل عظیم دادگان بینایی ورودی، شناسایی چهره‌های آشنا برای تنظیم روابط اجتماعی و مرتفع کردن نیازهای روانی و ...)، اطلاعات را طوری کدگزاری می‌کنند که با ساختار نورونی مغز تطابق داشته باشد و نحوه این کدگزاری بهترین روش موجود برای کدگزاری محرکهای موجود است. این نظریه که به کدگزاری بهینه معروف شده است مبین اینست که از میان تمام کدگزاریهای موجود برای تصاویر طبیعی، بهترین آنها در مغز وجود دارد. با توجه به این نکته پژوهشگران به تعریف معیار بهتر بودن یک کدگزاری نسبت به دیگری پرداخته‌اند. از این دسته معیارها می‌توان به کاهش افزونگی، افزایش پراکندگی کد، کاهش فاصله سیم کشی در مغز، تقلیل میزان مصرف انرژی، وغیره اشاره کرد. کاهش افزونگی یکی از معیارهای 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 می‌گیرد و می‌توان پس از این استخراج دانش و برای کدگزاری، این افزونگیها را از بین برداشت. 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 می‌کند. با استفاده از علوم مرتبط مانند نظریه احتمالات، نظریه اطلاعات، پردازش سیگنال و سیستم‌های خطی و غیره می‌توان به کشف قاعده مندیهای تصاویر طبیعی پرداخت. تعریف چارچوب آماری بررسی این قاعده مندیها و افزونگی‌ها را تسهیل می‌کند 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 می‌شود که روش‌هایی آماری برای حل مشکل کاهش افزونگی در اختیار ما قرار می‌دهند و از راه حل‌های موجود ایده‌هایی برداشته شده است.

۴. شبکه‌های عصبی: مسئله بازنمایی و یادگیری شبکه‌های عصبی به عنوان چارچوب اجرایی مغز، بخش دیگری از پژوهش را تشکیل می‌دهد. شبکه‌های عصبی مصنوعی به عنوان مدل ساده سازی شده ای از معادل طبیعی خود، ساختمان داده ای قوی برای انواع کارها فراهم می‌آورند. 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 داده‌هایی 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 ایده‌هایی 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 آنرا توجیه می‌کند، نیار به مدلی که بتواند رفتار آنرا در درک صحنه‌های طبیعی توجیه کند احساس می‌شود. این صحنه‌ها به عنوان محرکهای پیچیده راتر از آن هستند که مدل‌های ساده که

بخشی از عملکرد قسمتی از مغز را بیان می کند بتواند کامل‌آنرا 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 دیگر 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، کدگزاری 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 پدیده هایی که در مدل دیده می شود پیش بینی هایی صورت دهنند 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 صور 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 رشتہ تحریر درآمده است و ابزارهای متعددی از ادبیات هوش مصنوعی در آن بکار گرفته شده است. در این پایان نامه به برخی از ایده های بنیادین استفاده شده در این ادبیات اشاره شده و از آنها ایده هایی 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 V1 را بخوبی مدل کند. برای اینکار ابتدا سیر ذهنی طراحان مدل تحلیل مؤلفه های 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 داده ها 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 V1 بودند وسعی 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 L1 نیز توسط این مدل یاد گرفته می شوند و به کمک آنها می توان کواریانس ساختار همبستگی داده ها را نیز تخمین زد و در کدگزاری حذف کرد. این مدل نیاز به محاسبه توزیع حاشیه ای ضرایب خطی برای محاسبه درستنمایی MAP برای متغیرهای مراتب بالاتر را نیز مرتفع کرده است و توصیف روشتری از توزیع کانونی داده ها 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 تکه های تصویر بدست می دهد که با توجه به ساختار همبستگی و الگوی موجود در آنها (الگوهای

راستایی، فرکانسی و موقعیتی) یک کدگزاری 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 کدگزاری 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 کدگزاری با قابلیت خوش 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 مدلها 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 کدگزاری شده تصاویر توسط یک شبکه پرسپترونی 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 برخورده است، راه حل آنها و ایده های برگرفته شده از منابع دیگر اشاره می شود.

۲-۵- 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 کاراتر هستند، همانهایی هستند که قاعده مندیهای 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 مندیهای 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 عبارتند 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 روشهای 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 دورانهای زیادی است، می‌توان 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 می‌تواند با یک توزیع گوسی و کواریانس آن توصیف شود. تفاوت در توزیعهای توأم پاسخ‌فیلترهای خطی به تصویر، مشخصات الگوهای درون همبستگی‌ها را آشکار می‌کند که وقتی ویژگی‌های بیشتری مد نظر قرار بگیرند و احتمال نیز می‌شود. این الگوها می‌توانند 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 می‌تواند 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 می‌شوند. 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 می‌کند. این تعریف ساده آماری از سویی انعطاف‌پذیری در اخذ زنجیره توزیعهای تصاویر طبیعی و از سوی دیگر سادگی محاسباتی را برای تخمین راحت پارامترها به ارمغان می‌آورد.

یک مزیت مهم این روش اینست که بجای فرض کردن نامتفاوت (ویا حساسیت) به تعداد اندکی از پارامترهای محرک مانند موقعیت و مقیاس، مدل مجموعه ای به مراتب عمومی تر از ویژگیها را یاد می‌گیرد که توسط ساختارهای آماری تصاویر طبیعی بدست می‌آیند. اگر نورونهای لایه های بالاتر بینایی بر اساس این آمارگان تعمیم یافته باشند، باید در مقابل برخی از ابعاد ورودی نامتفاوت 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.

مدل واریانسی، و مجموعه فراکامل توابع پایه \mathbf{A} یک ماتریس کواریانس کاتونی ($\mathbf{A} = \bar{\mathbf{v}}$) برابر \mathbf{AA}^T خواهد داشت در حالیکه در مدل کواریانسی توزیع پیش فرض همان گوسی یکریخت با $\mathbf{C} = \mathbf{I}$ است. با افزایش فراکامل بودن ویژگی‌های خطی این اثر محسوس‌تر می‌شود. هنگامی که یک عضو فروکامل ویژگی‌های خطی استفاده می‌شود، مدل واریانسی تمام فضای داده را پوشش نمی‌دهد (ابعادی وجود خواهد 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 هر چند این مدل نیز نمی‌تواند 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 می‌کند و سیستم یادگیری سریعی خواهد داشت. همچنین روش پیشنهادی شراودولف که مجموعه ای از واحدهای رقابت کنند برای کشف کدگزاری جمعیتی را معرفی کرده است، ایده استفاده از جنبه کدگزاری جمعیتی شبکه عصبی برای تصاویر را تقویت کرد. این ایده در مقابل سایر روش‌های کاهش افزونگی قرار می‌گیرد که سعی می‌کنند افزونگی بین دو

کمیت وابسته را کاهش دهنده و تصویر را پس از کاهش افزونگی بازسازی، و بیان می دارد که تنها کدگزاری برای منظور این پایان نامه کافیست 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 معددی از ویژگیهای تصویری برانگیخته کننده و یا فرونشاننده، 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 زیروحدتها اصلاح ساختار آماری توزیع کدگزاری شده است، این نورونها همچنین ویژگیهای محركی 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 محرك خود خبر بدده، توسط تمام تصاویر این توزیع فعال می شود ولی بسته به محركی 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 مدلها خطی مانند ICA بدست می آید متفاوت است. مجموعه کامل بردارهای \bar{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 مدلها 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 بیشینه 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 بیشینه اشاره شده بود که گزینه جالبی به نظر می رسید زیرا با استفاده از آنبردارهای \bar{b}_k حاصل هر یک برای مجموعه خاصی از ویژگیهای تصویر (ترکیبی از راست، 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 \bar{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 راستادار و محلی برای کدگزاری 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 جرمیه کردن دارن و تأثیر بیشتری می گذارند. همچنین برای پیش پردازش، واپتینیگ روی آنها اعمال می شود تا ساختار همبستگی عمومی آنها حذف شود. در مدل کردن وابستگیهای کواریانسی 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 وتابع احتمال اولیه استفاده شده است.

در انتخاب تابع احتمال اولیه معمولاً به دنبال یک تابع زیرگوسی می گردند که متقارن نیز باشد. توزیع لاپلاسین 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 لاپلاسین برای واحد کنتراست DC مناسب نیست. در مدلها خطی مانند ICA، اثر مشابهی برای ضرایب روشناهی 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 روشایی که برای مدلها خطی با قیدهای پراکنده استفاده می شوند را به خدمت گرفت. این روشها با در نظر گرفتن یک جرمیه در تابع هدف برای راه حلها 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 مدلها بالا به پایین مشاهده شد می توان جملات دیگری را نیز به آن افزود تا جواب بدست آمده خصوصیات دیگری مانند پراکندهگی 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 پراکنده‌گی 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 مجددًا 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 تحلیلهای، 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 گیرنده‌گی 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 استرب^{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 بنده 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 اشتباهاً چهره اعلام کرده است را جمع آورده می کند. داده های اشتباهاً مثبت سپس به پایگاه داده به عنوان نمونه های جدید غیرچهره اضافه می شوند. این روش بوت استرب 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 ۲۰×۲۰ 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، راستاهای، موقعیتها و نورپردازیها برای آموزش شبکه استفاده شده اند.

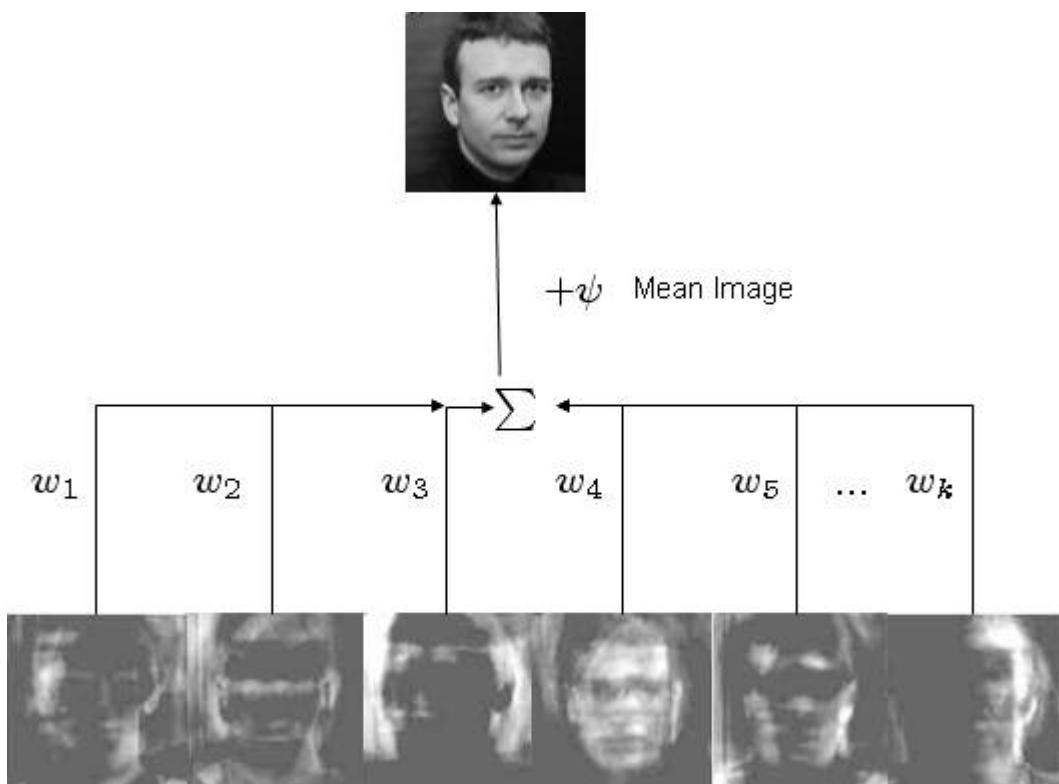
۳-۵- مقایسه با سایر روشها

۳-۱-۳-۵- 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 یکه نام دارد(Turk & Pentland 91) سعی دارد که مؤلفه های اصلی اوزیع چهره ها را بباید که در واقع بردارهای یکه ماتریس کواریانس مجموعه تصاویر چهره می باشد. این بردارهای یکه 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 یکه را به عنوان چهره هایی شیج مانند دانست که به آنها چهره های یکه اطلاق می شود.

³²⁴bootstrap

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 توانند توسط بهترین چهره های یکه تقریب زده شوند. منظور از بهترین، آنهاستی 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 m^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.

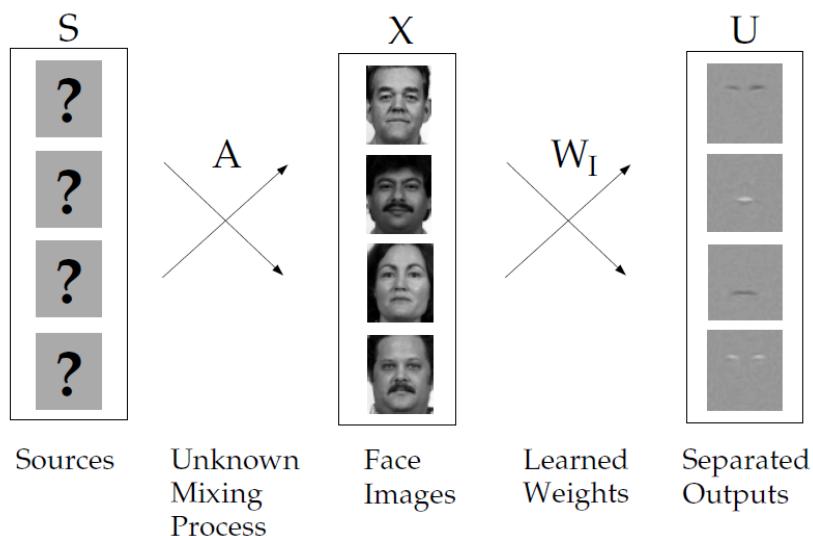
می توان از دانش موجود از فضای چهره استفاده کرد و چهره را در یک تصویر آشکارسازی کرد. ایجاد بردار وزنهای برای یک چهره معادل افکنش آنها روی فضای کم بعدتر چهره است. فاصله بین تصویر و افکنش آن روی فضای تصویر بسیار راحت محاسبه شده و ϵ نامیده می شود. طبق بررسیهای مقاله (Turk & Pentland 91) تصاویر چهره وقتی که بر روی فضای تشوییر 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 ϵ 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 ϵ 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 تسوییر مشخص می شود. متأسفانه محاسبه مستقیم معیار فاصله هزینه محاسباتی زیادی دارد و روشهای متعددی برای محاسبه کارای آن پیشنهاد شده است.

۵-۳-۵- 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)

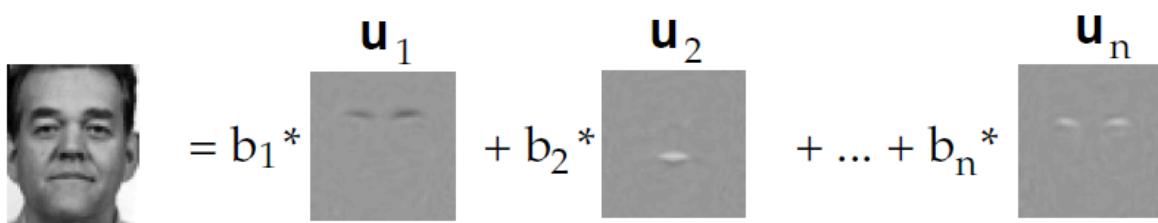
جهت پیدا کردن تصاویر پایه مستقل آماری برای مجموعه‌ی چهره‌ها، بر اساس مدل شکل ۳-۵، اجزای مستقل نمایش تصویر چهره پراکنده می‌شود. تصویر چهره در X به عنوان یک ترکیب خطی از مجموعه‌ای مجھول از منبع تصاویر مستقل آماری S فرض شده است، در جایی که A نیز یک ماتریس مخلوط مجھول می‌باشد. منابع با کمک ماتریس فیلترهای خطی W_I بازیابی می‌شوند، که خروجی‌های مستقل آماری U را تولید می‌کند.



شکل ۳-۵: 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 مجھول 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 تشکیلمی‌دهند. 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 ۴-۵). 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 ۴-۵ 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 خاطرنشان کرد که مدل ساخت تصویر، تصاویر X را به عنوان ترکیبی خطی از مجموعه‌ای مجھول 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 ۵-۵: بازنمایی تصاویر پایه مستقل از ضرایب b برای ترکیب خطی تصاویر پایه مستقل u تشکیل شده است که هر تصویر چهره X را می‌سازد. 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 یکه 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 می‌شود، ترکیبی خطی از پارامترها (تصاویر) بدست می‌دهد که مسئول لیشترين تنوع در مشاهده‌ها (پیکسلها) می‌باشد. استفاده از بردار PCA در ورودی رابطه‌های مرتبه بالا را از بین نمی‌برد. این رابطه‌ها همچنان در میان داده‌ها باقی مانده است ولی جدا نمی‌شود.

فرض کنید P_m ماتریسی است که اولین m محور مولفه‌های اصلی ستون‌هایش را تشکیل دهد. با اجرای ICA برروی U , 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 می‌شوند:

بازنمایی مولفه‌های اصلی از مجموعه‌ای از تصاویر دارای میانگین صفر در X بر اساس P_m به صورت $R_m = X_{rec} \times P_m^T$ تعریف شده است. تقریب کمینه خطای مربعی از رابطه X برای P_m بدست می‌آید.

الگوریتم ICA ماتریس $W_I = W \times W_z$ را چنان تولید می‌کند که

$$W_I \times P_m^T = U \Rightarrow P_m^T = W_I^{-1} U \quad (1-5)$$

بنابراین

$$X_{rec} = R_m \times P_m^T \Rightarrow X_{rec} = R_m \times W_I^{-1} U \quad (2-5)$$

که در آن W_z ماتریس وايتنیتگ است. بنابراین سطرهای $R_m * W_1^{-1} * W_2^{-1} * \dots * W_m^{-1}$ حاوی ضرایب ترکیب خطی منابع مستقل آماری U می‌باشد که X_{rec} را تشکیل داده است، که X_{rec} تقریب کمینه خطای مربعی X است، مانند آنچه در PCA داشتیم. بنابراین بازنمایی مولفه‌های مستقل تصاویر چهره براساس مجموعه‌ای از m تصویر ویژگی مستقل آماری U توسط سطرهای ماتریس زیر داده شده است.

$$B = R_m \times W_I^{-1} \quad (3-5)$$

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 $B_{test} = R_{test} \times W_I^T$ محاسبه شده است.

۵-۳-۲-۲-۳-۱-استقلال در فضای چهره در مقابل فضای پیکسل

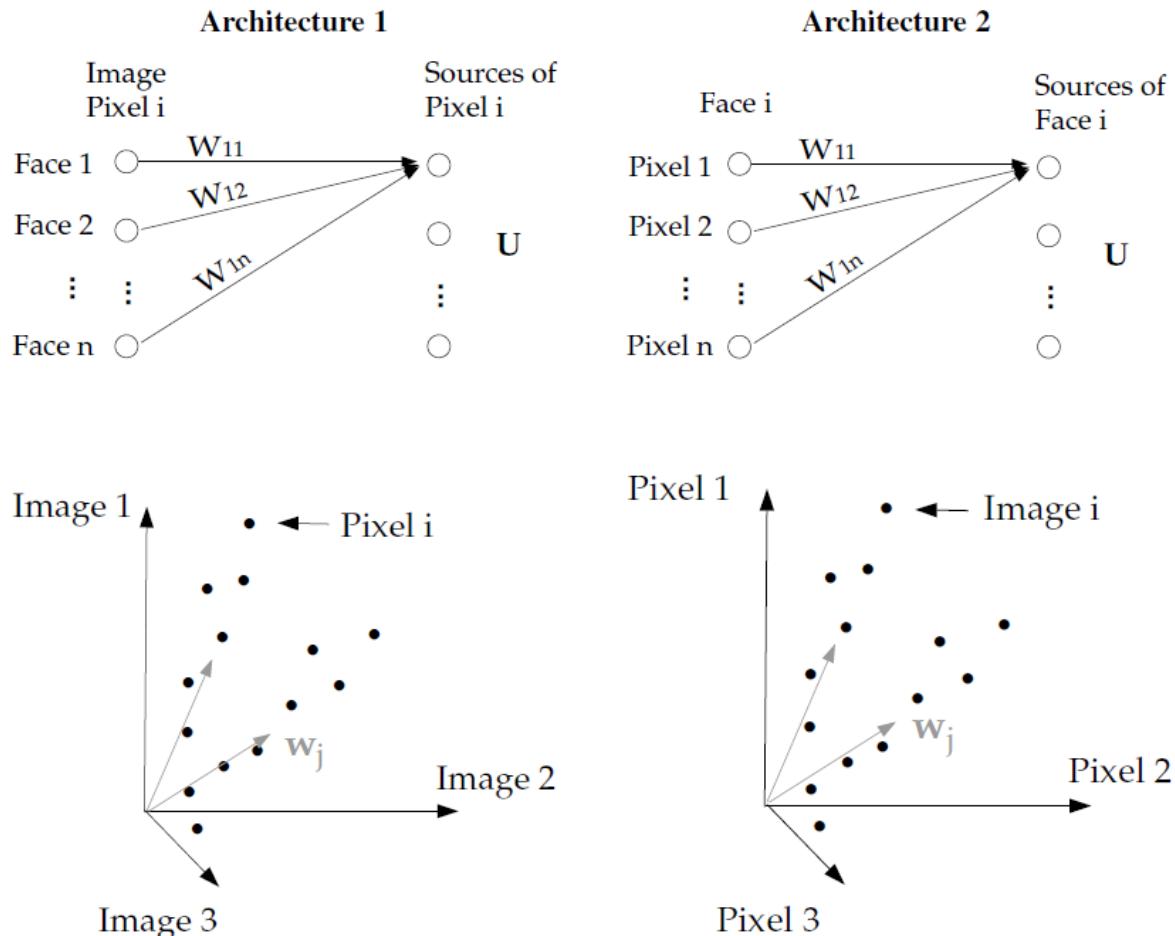
تحلیل‌ها در بخش ۱-۲-۳-۵ تصاویر پایه مستقل آماری را تولید می‌کند. الگوریتم ICA تصاویر را در میان محل پیکسل‌ها تفکیک می‌کند (شکل ۵-۵ چپ بالا). هر محل پیکسل یک مشاهده است که برای هر یک از چهره‌ها، مقادیر متفاوتی از سطوح خاکستری را اختیار می‌کند. این امر در شکل ۵-۵ نمایش داده شده است (چپ پایین) که در آن هر پیکسل برای تصویر هر چهره براساس مقدار خاکستریش 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 ناهمبسته‌اند. تحلیل مولفه‌های اصلی چهره‌ها 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 یکه‌ناهمبسته اند زیرا: ۱- آن‌ها طبق تعریف متعارف‌اند، 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 اندازه گیری 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 $\vec{X}\vec{v}_i$ 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 کدگزاری شده است، برخورد می‌شود. در معماری ۲، ICA، بردارهای وزن را در جهت استقلال آماری کدگزاری چهره‌در میان جمعیت چهره‌ها پیدا می‌کند. افکنش داده‌ها روی این وزن‌ها مجموعه‌ای از متغیرهای کدگزاری مستقل را بجای محل پیکسل‌ها تولید می‌کند، بطوری که مقدار هر متغیر کدگزاری از روی سایر متغیرهای کدگزاری قابل پیش‌بینی نمی‌باشد. هر چهره توسط مقداری که از این مجموعه جدید متغیرهای مستقل کدگزاری 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.

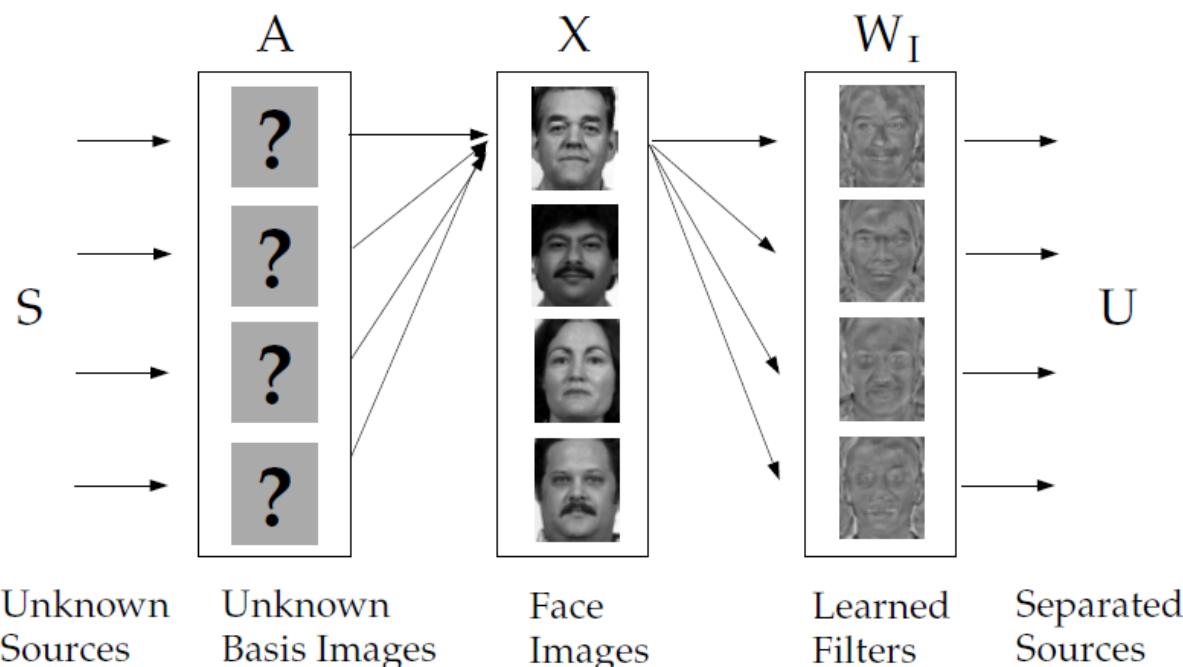
³²⁵ factorial code



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 سطرهای 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 جهتهای 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.

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 \mathbf{X} است و مانند آنست که تصاویر در ستونها باشند و پیکسلها در سطرهای (شکل ۵-۵ راست). تحت این معماری، فیلترها (سطرهای \mathbf{W}_I) تصاویر هستند که بالطبع ستونهای $A = \mathbf{W}_I^{-1}$ می باشند. ستونهای A یک مجموعه‌ی جدید از تصاویر پایه برای چهره‌ها، و ضرایب آنها برای بازسازی هر چهره (که در ستونهای خروجی‌های U ICA موجود است) را تشکیل می‌دهند.



شکل ۵-۶: مدل ساخت چهره با معماری ۲ بر مبنای (Bell & Sejnowski 97) و (Olshausen & Field 96). هر تصویر در مجموعه تصاویر به عنوان یک ترکیب خطی از تصاویر پایه ماتریس به حساب می‌آید. تصاویر پایه هریک به مجموعه‌ای از عوامل مستقل منتسب می‌شود که بدرار ضرایب آن در A آمده است. این عوامل با استفاده از ماتریسی از فیلترهای یادگرفته شده بازیابی می‌شود که تلاش می‌کند که تابعهای پایه مجھول را برای تولید خروجیهای مستقل آماری U معکوس می‌شود.

معماری ۲ به تصویر مدل ترکیبی در شکل ۷-۵ وابسته است. این مدل همانند مدل موجود در شکل ۳-۵ می‌باشد، به جز این که اکنون فرض می‌شود که چهره‌ها از ضریب‌های مستقل، S ، برای یک مجموعه از تصاویر پایه A مت Shankل شده‌اند در حالی که در مدل موجود در شکل ۳-۵ به صورت دیگری بود: منابع مستقل S تصاویر پایه بودند، و ضرایب در A قرار داشتند. این مدل براساس تصویر مدل ترکیبی از اولشان و فیلد (Olshausen & Field) می‌باشد، و همچنین توسط بل و سجووسکی (Bell & Sejnowski 97) برای پیداکردن فیلترهایی که از صحنه‌های طبیعی خروجی‌های مستقل آماری تولید کند، به کار رفته است. تلاش الگوریتم ICA بر این است که منبع ضرایب را بوسیله‌ی یافتن مجموعه‌ای از صافی‌ها W_I ، که خروجی‌های مستقل آماری U را تولید می‌کنند، بازیابی کند.

$$\text{Face Image} = u_1 * \text{Filter } a_1 + u_2 * \text{Filter } a_2 + \dots + u_n * \text{Filter } a_n$$

شکل ۵-۷: بازنمایی کد فاکتوریل از ضرایب مستقل u تشکیل می‌شود تا ترکیبی خطی از تصاویر پایه در A که هر تصویر چهره x را تشکیل داده است بدست دهد. $\text{ICA Representation} = (u_1, u_2, \dots, u_n)$

ستون های ماتریس خروجی $W_I X = U$ ، ICA یک کد فاکتوریل برای تصاویر آموزشی در X تهیه می کند. هر ستون از U شامل ضرایب تصاویر پایه درون A برای بازسازی هر تصویر در X می باشد (شکل ۷-۵). کد نمایش برای تصاویر آزمایشی بوسیله ای $W_I X_{test,i} = U_{test,i}$ پیدا می شود، در حالی که X_{test} ماتریس دارای میانگین صفر تصاویر آزمایشی می باشد، و ماتریس وزن، که توسط اجرای ICA بر روی تصاویر آموزشی پیداشده می باشد.

این دو روش بازنمایی امضاهایی را برای تصویر بدست می دهند. همانگونه که گفته شد امضای یک تصویر هنگامی که بر روی بردارهای مشخص شده توسط این روشها افکنده می شود بسیار شبیه سایر چهره هاست (به طوری که حتی می توان 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 تکه های تصویر استخراج می شود و امضاهای آن با افکنش بر روی فضای مربوط بدست می آید. سپس با استفاده از روش نزدیکترین همسایه روی امضای تکه تصاویری که از چهره ها و غیر چهره های متنوعی موجود است درباره چهره بودن یا نبودن آن تصمیم گیری می شود.

۳-۳-۵ مقایسه روشها

برای مقایسه روشها چند آزمایش ترتیب داده می شود.

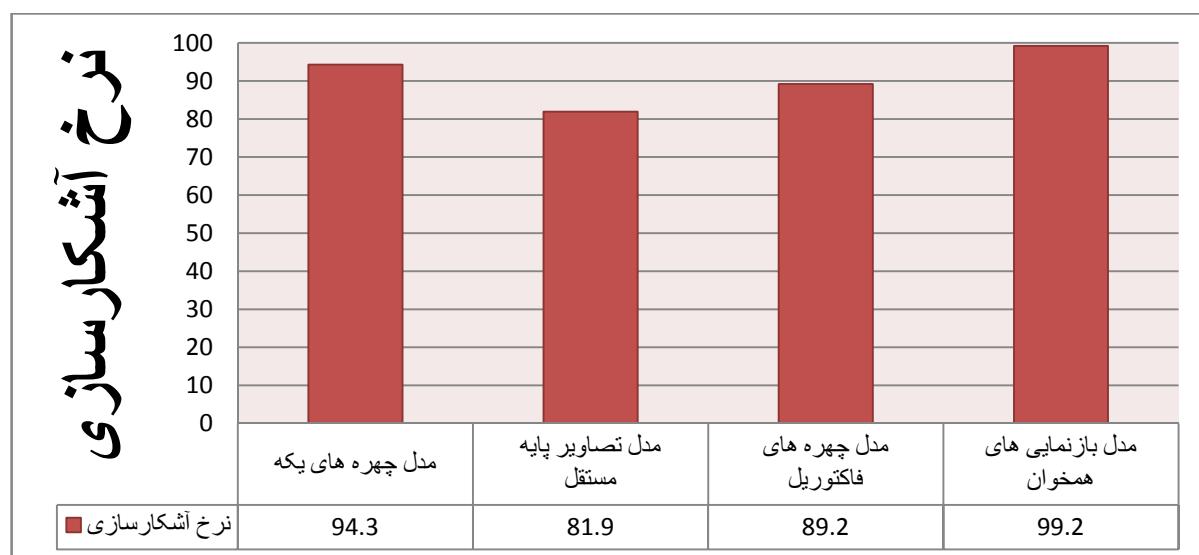
۳-۳-۱-۱ آشکارسازی چهره تمام رخ

در این آزمایش از مجموعه داده های AT&T Database (Olivetti) استفاده می شود که شامل ۴۰ فرد مختلف است که از هر کدام از آنها ۱۰ تصویر تمام رخ موجود است (شکل ۸-۵).



شکل ۸-۵: نمونه ای از مجموعه داده های AT&T (Olivetti) Database

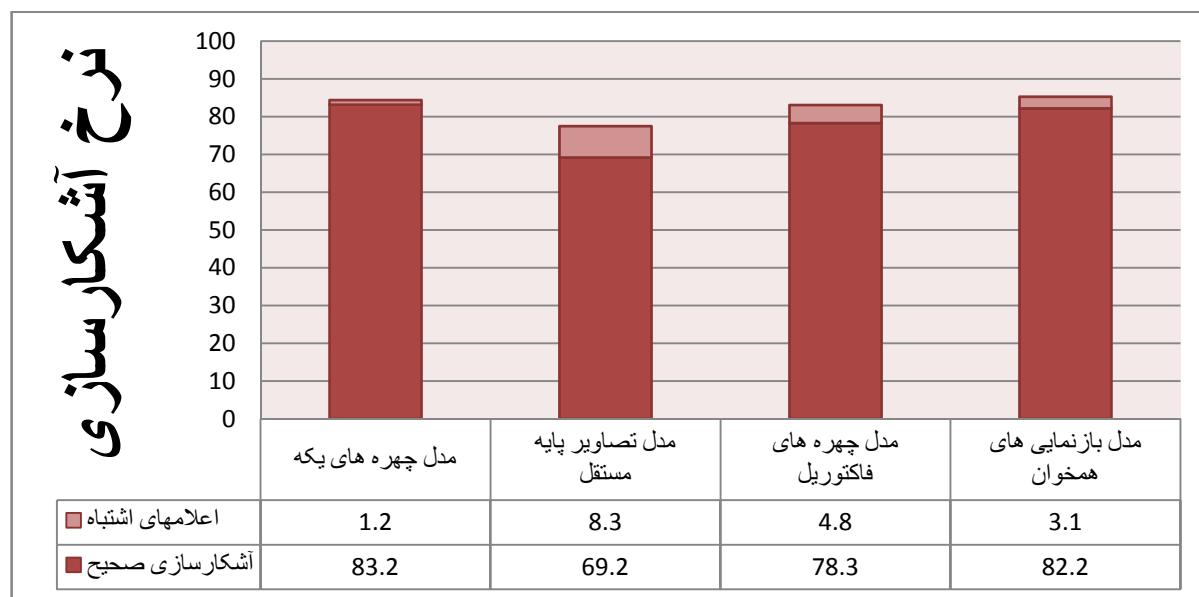
در این آزمایش از ۱۵۰ تصویر نیم رخی که وجود دارد استفاده نمی شود. هر چهار مدل در آموزش خود تنها به یک تصویر از ۱۰ تصویر هر فرد (علاوه بر مجموعه آموزشی FERET که برای هر چهار استفاده شده است) دسترسی داشته اند. هدف از این آزمایش مقایسه نرخ آشکارسازی چهره در تصویری است که تنها یک چهره تمام رخ تمام فضای آنرا پوشانده است. نمودار شکل ۹-۵ نتیجه هر چهار مدل را مقایسه می کند.



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.

۲-۳-۳-۵- 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 (جز مجموعه آموزشی FERET که برای هر چهار استفاده شده است) دسترسی نداشته اند.



شکل ۱۰-۵: 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 روشهای است. روش پیشنهادی نرخ آشکارسازی نسبتاً مشابهی دارد ولی به دلیل اعلامهای اشتباه (غلب اشتباه 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 زمینه‌ای غیر ساده قرار دارد. نمودار شکل ۵-۱۰ نتیجه هر چهار مدل را مقایسه می‌کند.

۴-۵- کارهای آتی

چنانکه تمامی تحقیقات علمی بر می‌تابد، پژوهش حاضر نیز کامل نیست و نویسنده 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 می‌شوند که در ادامه آورده شده‌اند:

- میانگین توزیع درستنمایی صفر فرض شده است. این ساده سازی 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 هردوی این مجموعه ضریب‌های مثبت و منفی یک مجموعه بهینه از مؤلفه‌های کواریانس را می‌سازد. چنین فرضی درباره متقارن 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 پایان نامه، 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 شبیب نورده صعودی تکراری با گامهای پیشخور تقریب زد. این کار بخصوص وقتی مفید است که بخواهیم استنتاج بسیار سریع اطلاعات بینایی 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 نمونه گیری 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 شبیب نورده صعودی استفاده شده است. همانند تخمین متغیرهای نهفته، می‌توان از تعریف رسمی گرادیان استفاده کرد که به تجزیه طیفی A در هر تکرار نیازمند است، و یا می‌توان از تقریب سریعتری که روی بسط سریپها 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 کوارینس لگاریتمی تک نمونه‌ای تقریب می‌زنند 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 واپتینینگ 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 تعامل یک پایه فراکامل را بیشینه می کنند نیز به مؤلفه های راستدار و محلی مشابهی منجر می شوند. همچنین معیار پراکندگی که تعییه 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 وزنهای لایه V1 را شبیه سازی کرد و از معیارهایی مانند کورتوسیس، نگانتروپی، اطلاعات متقابل، پراکندگی، درستنمایی استفاده شده در این پژوهش و ... به عنوان تابع برآش 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. Hegdé (2004). "Visual cortex: The continuing puzzle of area V2." Current Biology **14**(13): R523-R524.
- Brehm, H. and W. Stammller (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*. Chamroisse, 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.
- Derugin, 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.
- Hegdé, J. and D. Van Essen (2000). "Selectivity for complex shapes in primate visual area V2." *Journal of Neuroscience* **20**(5): 61.
- Hegdé, J. and D. Van Essen (2003). "Strategies of shape representation in macaque visual area V2." *Visual Neuroscience* **20**(03): 313-328.
- Hegdé, 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. Worshop 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 (CIMSVF). Nashville, TN, IEEE.

- Martó-Âñez, 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. Touretzky. 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 Francisco, 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.
- Tailor, 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.
- Valpoli, 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. Zetzsche (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.
- Zetzsche, 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.
- Zetzsche, C. and E. Barth (1990). "Fundamental limits of linear filters in the visual processing of two dimensional signals." *Vision Research* **30**(7): 1111-1117.
- Zetzsche, 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.
- Zetzsche, 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
محرك	stimuli
میدان گیرنده‌گی	receptive field
تداعی موقتی	temporal association
برانگیخته کننده	excitatory
فرونشاننده	suppressve
پنجره های سینوسی	sinusoidal gratings
حول مرکز	center surround
موقتی مکانی	spatio-temporal
مسیر پشتی (خلفی)	dorsal pathway
مسیر شکمی (بطنی)	ventral pathway
حافظه انجمنی	associative memory
راستایی 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

رابطه	conjunctive
شبکیه	retina
قرنیه	cornea
مایع زلاییه	aqueous humor
مردمک	pupil
عنبیه	iris
مایع زجاجیه	vitreous humor
دورنمای بهینه سازی	optimization landscape
تفریق شناختی	cognitive subtraction
عقده 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 مژگانی	ciliary muscle
(تطابق (چشم)	accommodation
صفحه بینایی	optic disk
نقطه کور	blind spot
ماکولا (لکه)	macula
تیزبینی	visual acuity
سلولهای عقده 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
ساختر مشبك	reticular formation
کولیکولوس تحتانی	inferior colliculus
هسته بالشتک	pulvinar
پرتفکتم	prectectum
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	infrotemporal cortex

نواحی فوق گیجگاهی چند سنسوری	superior temporal polysensory
گزینندگی	selectivity
تکنیک ضبط تک نورون	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
قانون توجه	focus-of-attention
تحریک	stimulate
فرونشانی اطرافیان	surround suppression
بازداری فعالیت به ازای راستهای متقاطع	cross-orientation inhibition
اشباع پاسخ کنترast	contrast response saturation
فرونشانی ماسک	mask suppression
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
محرك	stimuli
میدان گیرندگی	receptive field
تداعی موقتی	temporal association
برانگیخته کننده	excitatory
فرونشاننده	suppressive
پنجره های سینوسی	sinusoidal gratings
حول مرکز	center surround
موقتی مکانی	spatio-temporal
مسیر پشتی (خلفی)	dorsal pathway
مسیر شکمی (بطنی)	ventral pathway

حافظه انجمنی	associative memory
راستای مرچح	preferred orientation
تصویربرداری تشید مغناطیسی	magnetic resonance imaging
ناحیه دوک مانند چهره	fusiform face area
مدلهای تبدیل شونده سه بعدی	3D morphable model
روشهای مبتنی بر ظاهر	appearanc-based methods
کلیشه اولیه ارجاعی	apriori elastic model
مدل فعال شکل	active shape model
مسئله انتساب	assignment problem
جبابها و رگه ها	blobs and streaks
خلوص رنگی	chrominance
بستان (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 تطبیق	disparity
نقشه عدم توافق	disparity map
متمايز	distinct
تطابق گراف دسته انعطاف پذیر	elastic buch graph matching
لبه	edge
چهره یکه	eigenface
تعقیب تکاملی	evolutionary persuit
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
اشتباهه منفی	false negative
اشتباهه مثبت	false positive
روشهای نامتغیر با ویژگی	feature invariant approaches
تمام رخ	frontal face
خاکستری فام	grayscale

یادگیری هیب	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
شكل	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 ویترbi	Viterbi segmentation
محلي بودن	locality
مدلهای تبدیل شونده سه بعدی	3D morphable model
روشهای مبتنی بر ظاهر	appearanc-based methods
کلیشه اولیه ارجاعی	apriori elastic model

مدل فعال شکل	active shape model
مسئله انتساب	assignment problem
حبابها و رگه ها	blobs and streaks
خلوص رنگی	chrominance
بستن (عملکر 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
متمايز	distinct
تطابق گراف دسته انعطاف پذیر	elastic buch graph matching
لبه	edge
چهره یکه	eigenface
تعقیب تکاملی	evolutionary persuit
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 ویژگی های چهره	facial feature detection
اشتباهه منفی	false negative
اشتباهه مثبت	false positive
روش های نامتفیر با ویژگی	feature invariant approaches
تمام رخ	frontal face
خاکستری فام	grayscale
یادگیری هیبی	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
شكل	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 ویتری	Viterbi segmentation
محلي بودن	locality
مدلهای تبدیل شونده سه بعدی	3D morphable model
روشهای مبتنی بر ظاهر	appearanc-based methods
کلیشه اولیه ارجاعی	apriori elastic model

علم آمار

افزونگی	redundancy
هم وقوعی های مشکوک	suspicious coincidences
قاعده مندیهای 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 موآد	generative model
حفظ اطلاعات	information preservation
درستنمایی	likelihood
توزیع احتمال	probability distribution
احتمال اولیه	prior probability
احتمال ثانویه	posterior probability
شیب نورده 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
ناریب	unbiased
کد پراکنده	sparse code
امید ریاضی، مقدار مورد انتظار	expected value
کدگزاری 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 یگه	orthonormal

فراکامل	overcomplete
تجزیه موجک متعامد یکه	orthonormal wavelet decomposition
همگنی	homogeneity
نوسان	fluctuation
آشوب	turbulence
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 لاپلاسین	Laplacian distribution
توزیع گوسی	Gaussian distribution
توزیع نمایی	exponential distribution
توزیع رایلیخ	Rayleigh distribution
طیف بزرگی	amplitude spectrum
فرکانس مکانی	spatial frequency
آمارگان	statistics
تعامل	interaction
تحلیل جدای پذیری خطی	linear discriminant analysis
افتراء	divergence
واریانس-متغیر	heteroscedastic
کمترین طول توصیف	minimum description length
تناوب کاری	duty cycle
دور بر درجه	cycle per degree
تفکیک کورکوانه منابع	blind source separation
متعامد کردن	orthogonalization
تبدیل وايتینگ	whitening transform
زیرگوسی	subgaussian, platikurtotic
زیرگوسی	supergaussian, leptokurtotic
تابع احتمال (نوك تیز)	spiky, peaky
ذنباله های سنگین	heavy tails
اندازه گیریهای دور از مرکز	outliers
نگانتروپی	negentropy
تبدیل قابل بازگشت	invertible transform

اصل اینفوگریفیکس infomax principle

اصل بیشینه سازی آنتروپی شبکه network entropy maximization principle

لغات عمومی

پایین به بالا bottom up

روش قضاآت arbitrate scheme

بالا به پایین top-down

ابرایانه supercomputer

مخف ها:

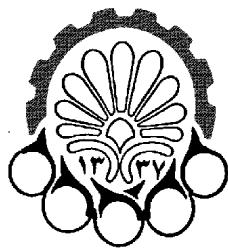
<i>Abbreviation</i>	<i>Full Text</i>
PCA	principle component analysis
LFA	local feature analysis
ICA	independent component analysis
GSM	Gaussian scale mixture
SVD	singular value decomposition
LDA	linear discriminant analysis
cpd	cycle per degree
LGN	lateral geniculate nucleus
IID	independent and identically distributed
EM	expectation maximization
MC	Markov chain Monte Carlo
LMC	large monopolar cells
MTF	modulation transfer function
PDM	point distribution model
ASM	active shape model
FA	factor analysis
SNoW	sparse network of winnows
HMM	hidden Markov model
EP	evolutionary pursuit
EBGM	elastic buck 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 consistant 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 localizion of faces appeared in a still image. The robustness of the algorithm in cluettered 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