

فصل سوم:

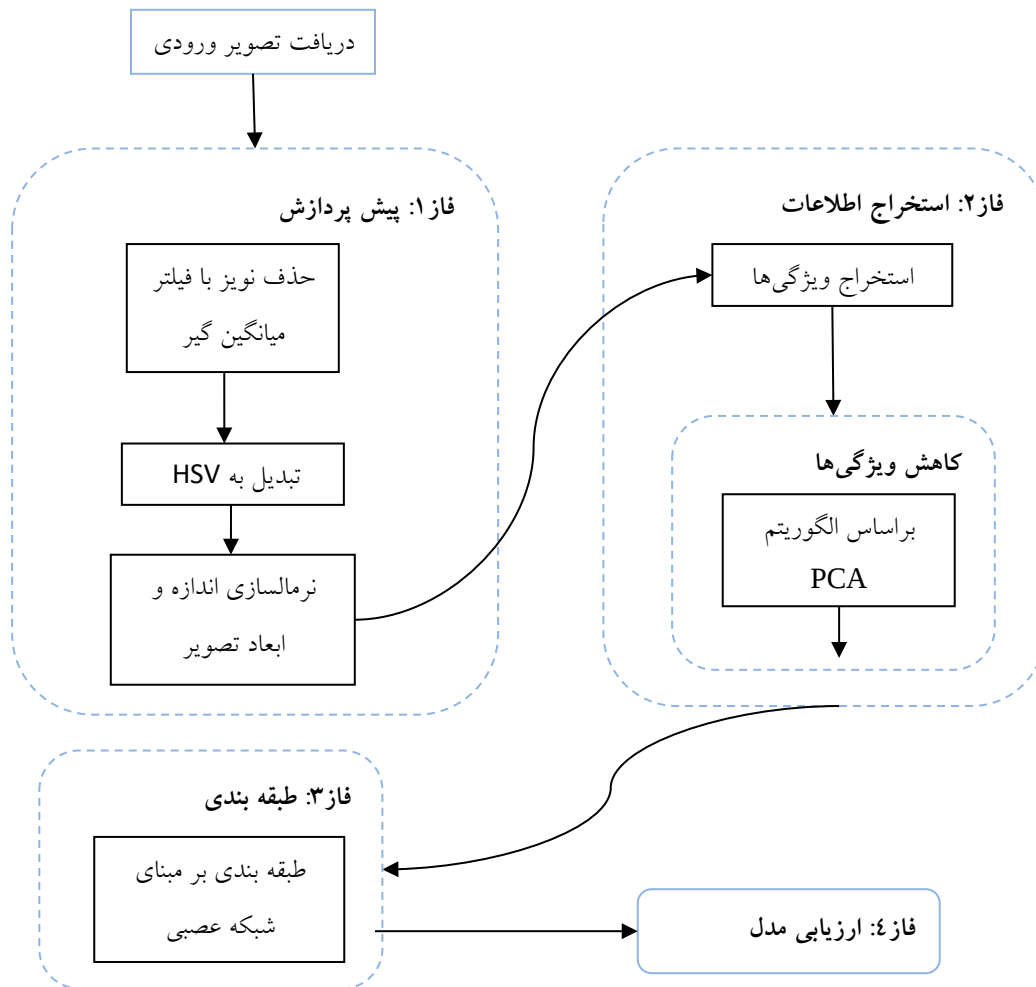
۱ روش تحقیق

۱-۱-۱- مقدمه

پردازش تصویر با پردازش سیگنال سرو کار دارد و دارای دو شاخه عمده بهبود تصویر و بینایی ماشین است. بهبود تصویر در برگیرنده روش‌هایی مثل استفاده از فیلترها و افزایش تضاد برای بهتر کردن کیفیت دیداری تصویر و اطمینان از نمایش درست آن در خروجی است و بینایی ماشین شامل روش‌هایی است که به کمک آن‌ها می‌توان معنی و محتوای تصویر را درک کرد تا از آن در کارهایی مثل رباتیک و محور تصویر استفاده شود. در این فصل ویژگی تصاویر با استفاده از بهترین تکنیک‌های استخراج ویژگی مشخص می‌شوند و سپس به شبکه عصبی مصنوعی برای یادگیری داده می‌شوند. بدین منظور در مرحله پیش‌پردازش از حذف نویزها، نرمال‌سازی و .. برای آماده‌سازی تصویر استفاده شده است. سپس به کمک شبکه عصبی مصنوعی در سه جهت افقی، عمودی و قطری تصویر را آنالیز می‌کنیم. با توجه به نمودار هیستوگرام و تعریف توسط یک سطح آستانه و اعمال آن بر روی تصویر و پس از یادگیری و تست شبکه عصبی محل و میزان گستردگی ضایعات مشخص می‌شود.

۱-۱-۲- مراحل انجام پردازش تصویر

هنگامی که تصویر از ورودی دریافت می‌شود در ابتدا باید مراحل را طی کند تا بتوانیم روش‌های موردنظر را بر روی آن اعمال کنیم. بعد از اینکه تصویر ورودی خوانده شد حال باید پیش‌پردازش روی آن اعمال شود. بعد از آن با استفاده از ویژگی‌های استخراج شده و کاهش ویژگی به منظور سرعت بخشیدن در کار ضایعات شناسایی می‌شوند سپس برای بهبود ضایعات شناسایی شده آن‌ها را به شبکه عصبی RBF می‌دهیم تا کار تشخیص ضایعات را برای ما انجام دهد. در شکل (۱-۳) چارت روش انجام تحقیق مشخص شده است.



شکل (۳-۱): چارت روش انجام تحقیق

۱-۳-۱- فاز ۱: پیش پردازش

در این مرحله الگوریتم‌های پیش‌پردازش ارائه شده است. این الگوریتم‌ها برای تبدیل تصویر رنگی به HSV شامل نرمال‌سازی اندازه، ابعاد تصویر و حذف نویز می‌باشند که برای جداسازی مولفه‌های رنگی و بهبود کیفیت تصاویر استفاده می‌شود تا اعتبار بیشتری پیدا کند. اولین گام پیش‌پردازش حذف نویز است. در پردازش تصویر نیاز به تشخیص یک محدوده رنگی مثلاً رنگ نارنجی داریم در این حالت یافتن این محدوده رنگی با ترکیب ۳ رنگ اصلی ممکن نیست یا اینکه کاری دشوار و وقت گیر است.

۱-۱-۴- حذف نویز

هدف این گام کاهش نویز و جدا کردن سیگنال از نویز است در این مرحله با حذف نویز، تصویر برای مرحله بعدی آماده می‌شود. برای حذف نویز از فیلتر میانگین گیر^۱ که از فیلترهای پر کاربرد است استفاده خواهد شد. ابتدا یک کرنر از تصویر مشخص می‌شود که اندازه آن بستگی به میزان تارشدگی تصویری که مدنظر است دارد. اگر فرض کنیم یک کرنر 3×3 انتخاب کرده‌ایم برای اینکار ابتدا کرنر را روی تصویر اولیه قرار می‌دهیم وقتی این ماسک روی تصویر اولیه قرار می‌گیرد تمام پیکسل‌هایی که در اطراف آن پیکسلی که بعنوان مرکز انتخاب شده باهم جمع می‌کند و حاصل این جمع شدن بر عدد ۹ تقسیم می‌کند (چون کرنر ما از ۹ پیکسل تشکیل شده است). نتیجه حاصل میانگین این ۹ پیکسل است که وقتی به پیکسل متناظرش در ماتریس نتیجه اعمال شود یک شدت رنگ متناسب با این پیکسل و همسایه‌های آن خواهد بود. روش آسان‌تر آن ضرب تک تک پیکسل‌ها در $1/9$ است سپس جمع آن‌ها با یکدیگر انجام می‌شود در تصویر ۲-۳ این تغییرات اعمال شده است.



تصویر (۲-۳): اعمال فیلتر میانگین گیر

۱-۱-۵- تبدیل تصویر RGB به HSV

تصاویر خروجی از دوربین‌ها معمولاً RGB هستند به همین دلیل اولین کار بعد از دریافت تصویر از دوربین این است که سیستم RGB را به HSV تبدیل کنیم این کار در نرم افزار متلب بسیار ساده است کافیت از تابع $A = \text{RGB2HSV}(K)$ استفاده کرد که K تصویر با فرمت RGB و A همان تصویر قبلی با فرمت HSV است.

¹ median

برای هر رنگ سه مشخصه اصلی وجود دارد:

- ۱- نام رنگ^۲: مشخص کننده نوع رنگ مثلاً آبی، قرمز، متمایل به نارنجی.
 - ۲- شدت رنگ^۳: هر رنگ می‌تواند پررنگ یا کم رنگ باشد اما ماهیت ذاتی آن ثابت است.
 - ۳- روشنایی یا تیرگی رنگ^۴: مولفه شدت رنگ با نور تابیده شده به آن تغییر می‌کند، بعنوان مثال اگر نور تابیده شده کم باشد رنگ متمایل به سیاه می‌شود.
- حوضه جدیدی از رنگ بنام HSI که الهام گرفته از چشم انسان و مخفف سه مشخصه اصلی رنگ است. کار کردن با HSV دقیقاً معادل سیستم HSI می‌باشد که با داشتن یک تصویر با فرمت HSV می‌توان به راحتی نقاط با رنگ‌های موردنظر را روی تصویر یافت. در این سیستم مقدار H بین ۰ تا ۳۶۰ قابل تغییر است. یعنی از قرمز تا سبز ۱۲۰ درجه و از سبز تا آبی ۱۲۰، از آبی تا قرمز ۱۲۰ درجه خلاف جهت عقربه‌های ساعت است که S بین ۰ تا ۱۰۰ تغییر می‌کند یعنی از تاریک ۰ تا روشن ۱۰۰. چند نمونه از رنگ‌ها در دو حوضه RGB و HSI در جدول ۱-۳ نشان داده شده است:

جدول (۱-۳): نشان دهنده مقایسه رنگ‌ها در دو حوضه مختلف

رنگ RGB	رنگ HIS
سیاه ۲۵۵،۲۵۵،۲۵۵	۰،۰،۰
سفید ۲۵۵،۲۵۵،۲۵۵	۱۰۰،۰،۰
قرمز ۲۵۵،۲۵۵،۰	۱۰۰،۰،۱۰۰
سبز ۰،۲۵۵،۰	۱۰۰،۱۲۰،۱۰۰
آبی ۰،۰،۲۵۵	۱۰۰،۲۴۰،۱۰۰
قهوه ای ۶۴،۱۲۸،۱۲۸	۱۸۰،۵۰،۵۰

^۲ hue

^۳ saturation

^۴ Intensity

۱-۱-۶- نرمال سازی اندازه و ابعاد تصویر

ابتدا باید تصاویر به تاثیر عواملی مثل فاصله خطوط، اندازه ها نرمال سازی شوند تا بلوک یکنواختی از تصویر حاصل شود تصاویر ورودی هر کدام ابعاد و اندازه متفاوتی برای خود دارند تعیین ابعاد امری ضروری در روش شبکه عصبی مصنوعی است چون تصویر به ماتریس های کاملاً یکسانی تبدیل می شود. دو تصویر با اندازه های متفاوت را می توانیم با دستور `imresize` در نرم افزار متلب به تصاویر با اندازه های یکسان تبدیل کنیم.

۱-۱-۷- فاز ۲: استخراج ویژگی ها

هدف از استخراج ویژگی به دست آوردن داده های خام به شکل قابل استفاده تری به منظور پردازش های آماری می باشد. در اغلب روش های موجود برای ارزیابی، شناسایی و بهبود کیفیت تصاویر، استخراج ویژگی از مراحل ویژه به حساب می آید. در حقیقت، وقتی حجم داده ورودی به یک الگوریتم زیاد است و پردازش آن مدت زیادی طول می کشد، استخراج ویژگی انجام می شود تا از داده هایی که متناسب با حجم زیادشان اطلاعات ندارند، یک مجموعه ویژگی به دست آید که نمایش مختصرتری از داده است که برای نمایش یک کلاس یا یک مجموعه نمونه باید شناخته شده باشند. انتخاب ویژگی نویسه، یا یک نمونه اصلی سعی می کند تا بهترین ویژگی ها را که بیشترین اطلاعات متمایزکننده را در خود دارند، ضمن رعایت محدود بودن تعداد ویژگی از ویژگی های استخراج شده، به دست آورد. مکان نسبی نقاط انتهایی، نقاط تقاطع و گوشه ها از ویژگی های محلی و تعداد حفره ها در یک نویسه یا تعداد تفرقه ها در مرز بیرونی از ویژگی های کلی هستند. در ادامه برخی از روش های متداول برای استخراج ویژگی مورد بررسی قرار می گیرند.

۱-۱-۸- استخراج ویژگی براساس هیستوگرام

هیستوگرام تصویر نموداری است که توسط آن تعداد پیکسل های هر سطح روشنایی در تصویر ورودی مشخص می شود. فرض کنید تصویر ورودی یک تصویر Grayscale با ۲۵۶ سطح روشنایی باشد، بنابراین هریک از پیکسل های تصویر مقداری در بازه [0...255] می توانند داشته باشند. بنابراین با محاسبه هیستوگرام شما برای هر تصویر یک آرایه به طول ۲۵۶ دارید که یک بردار ویژگی تصویر مورد نظر شماست. یکی از راه های موثر در استخراج ویژگی نمایش روش هیستوگرام طیفی برای عمل تشخیص و دسته بندی می باشد که دارای ویژگی تبدیل - ثابت است. این نمایش با تقسیم حوزه فرکانس به نواحی کوچک مستقل از هم به دست

می‌آید. این کار منجر به تولید یک مجموعه از فیلترها و یک نمایش که شامل توزیع‌های حاشیه‌ای^۵ و پاسخ‌های آن فیلترها می‌باشد، می‌شود. به این ترتیب حوزه فرکانس به نواحی کوچک تقسیم خواهد شد و فیلترهای مکانی مربوطه به دست می‌آید و کانولوشن تصویر داده شده با هر فیلتری محاسبه شده و توزیع حاشیه‌ای تصویر فیلتر شده حاصل را به دست خواهد آورد و آنگاه، توزیع‌های حاشیه‌ای با هم ترکیب خواهند شد. بنابراین هر تصویر به وسیله یک بردار شامل توزیع‌های حاشیه‌ای نمایش داده می‌شوند که نمایش هیستوگرام طیفی تصویر نام دارد. منظور از روش هیستوگرام برای استخراج ویژگی، استفاده از روش‌های اصطلاحاً histogram static است. در این روش برای استخراج ویژگی از تصاویر با استفاده از مشخصه‌های آماری مرتبه‌ی اول (براساس هیستوگرام شدت روشنایی پیکسل‌ها) ویژگی‌هایی از تصویر استخراج شده است و تعدادی از این ویژگی‌ها شامل میانگین، چولگی^۶، انحراف مطلق^۷، واریانس، کشیدگی^۸ و انحراف معیار^۹ می‌باشد. به این علت اسم هیستوگرام را دارد که در اون براساس روشنایی تک پیکسل (نه براساس بافت تصویر و همسایگی تعداد پیکسل) ویژگی استخراج می‌شود (ویژگی‌های استاتیک). با استفاده از روش‌های آماری مراتب بالاتر هم استخراج ویژگی‌های بافتی نتیجه می‌شود که ویژگی‌های دیگری هستن و بر اساس بافت تصویر از تصویر حاصل می‌شوند و با این ویژگی‌ها متفاوتند که مستلزم استفاده از قوانین و تکنیک‌های GLCM، GLRLM و ... است. برای طبقه‌بندی و کلاس‌بندی داده‌ها نیاز به یک سری دیتا داریم. وقتی که بر روی تصاویر کار می‌کنیم و این دیتاها در اختیارمون نیست، نیاز هست که یک سری ویژگی‌ها رو از تصویر استخراج کنیم که یکی از این روش‌ها استفاده از ویژگی‌های بافت تصویر است. همانطور که در مطالب بالاتر توضیح داده شد می‌توان از اطلاعات هر پیکسل و بر اساس مشخصه‌های آماری مرتبه‌ی اول این ویژگی‌ها را استخراج کرد که استخراج ویژگی از روش‌های استاتیک نامیده می‌شود اما می‌توان علاوه بر اون از مشخصه‌های آماری مرتبه‌ی دوم یا همان GLCM و یا مشخصه‌های آماری مراتب بالاتر یا GLRLM استفاده کرد. برخی روش هیستوگرام استاتیک/ استاتیک را هم جزو ویژگی‌های بافتی می‌دانند (چون علاوه بر اینکه از پیکسل‌های تصویر استفاده می‌کند از اطلاعات تک پیکسل هم استفاده می‌کند) اما GLCM و GLRLM براساس ویژگی‌های بافتی و بر اساس اطلاعات پیکسل‌های همسایگی هر پیکسل این ویژگی‌ها را نتیجه می‌دهند (نه یک پیکسل تک و خاص در مواردی که مثلاً دقت تصویر ما پایین هست). مثل تصاویر MRI که حاوی نویز هستند نیاز است که بر اساس اطلاعات همسایگی پیکسل‌ها این ویژگی‌ها استخراج شود. پس این روش‌ها روش‌های خوبی برای این نوع تصاویر هستند. روش

⁵ Marginal distributions

⁶ Skewness

⁷ Absolute deviation

⁸ Kurtosis

⁹ Standard deviation

^{۱۰} GLCM از ماتریس‌های هم‌رخداد که اولین بار توسط هارالیک معرفی شده مشخصه‌ها و ویژگی‌های هارالیک را استخراج کرده است. این روش بر اساس ماتریس‌هایی کار می‌کند که در جهت و زوایای مختلف می‌توانند ویژگی‌هایی از بافت رو نتیجه دهند و در واقع بر اساس احتمال اتصال هر جفت پیکسل در کنار هم عمل می‌کنند.

در ^{۱۱} GLRLM از ماتریس run length برای استخراج ویژگی می‌توان استفاده کرد که میزان ساییدگی یک بافت را در هریک از جهت‌های 0° , 45° , 90° , 135° تعیین می‌کند. این ماتریس هم مثل ماتریس‌های هم‌رخداد در GLCM، اگر تصویرمون n تا سطح خاکستری داشته باشه، $n*n$ تا عنصر می‌تواند داشته باشد. بعنوان مثال در تصویر ۳-۷ اگر فاصله‌ها رو یک در نظر بگیریم. سطر i در ماتریس یک عدد خاکستری رو از ماتریس سمت چپ (تصویرمون) نشان می‌دهد و ستون‌ها (j) بیانگر تعداد مرتبه‌ای است که این‌ها کنار هم قرار گرفته‌اند. مثلاً در ماتریس سمت راست اگر در خونه ی ($i=2, j=3$) عدد یک نوشته شده، به این معنی است که در ماتریس تصویرمون (سمت چپی) عدد ۲ تنها یک بار به صورت سه تایی در تصویر قرار گرفته است. به عنوان مثال وقتی به ازای مقدار $i=4, j=2$ در ماتریس مقدار ۱ مشاهده می‌شود یعنی این که پیکسل دارای سطح خاکستری ۴ تنها یک بار در ماتریس تصویر اصلی به صورت دوتایی در کنار هم قرار گرفته اند. نمونه‌ای از ویژگی‌های استخراج شده با این روش: عدم انحراف سطح خاکستری^{۱۲}، تاکید بلند مدت^{۱۳}، تاکید کوتاه مدت^{۱۴} و

1	2	3	4
1	3	4	4
3	2	2	2
4	1	4	1

Gray level	Run length(j)			
i	1	2	3	4
1	4	0	0	0
2	1	0	1	0
3	3	0	0	0
4	3	1	0	0

¹⁰ gray level co-occurrence matrix

¹¹ gray level run length matrix

¹² Gray level nonuniformity

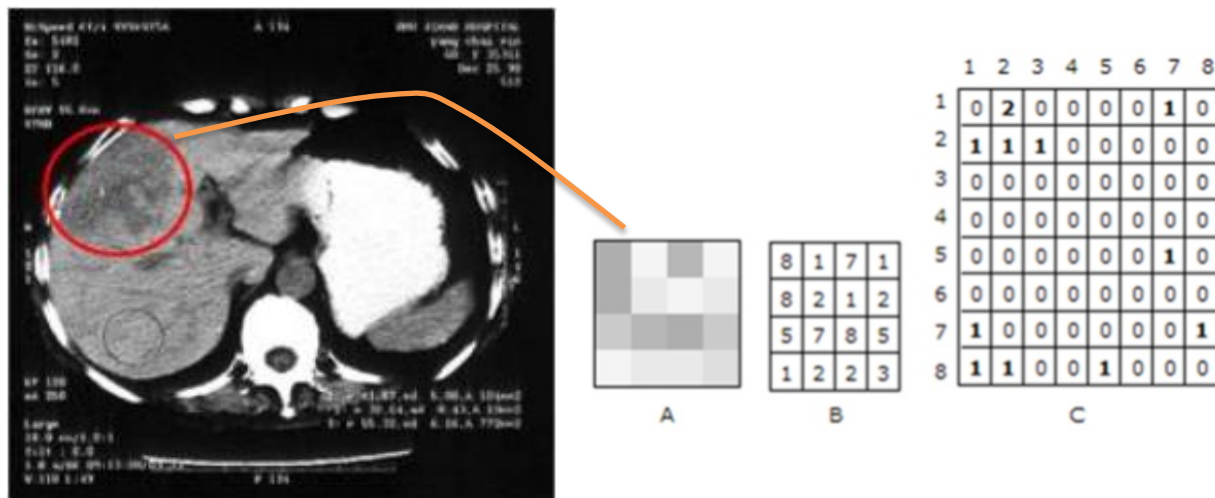
¹³ Long run emphasis

¹⁴ Short run emphasis

شکل (۷-۳): نمایش ماتریس Run length برای استخراج ویژگی [38]

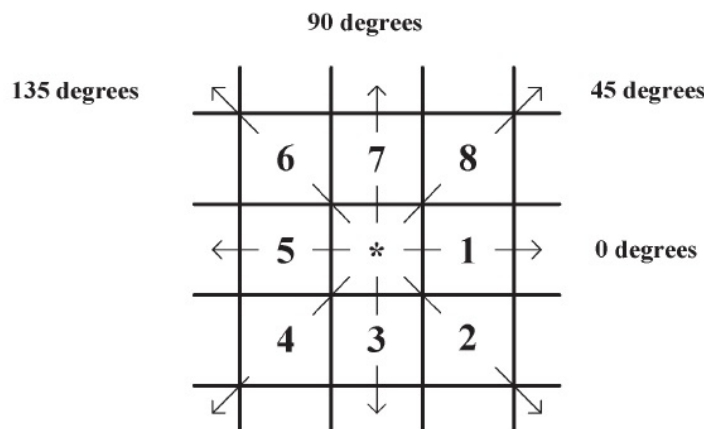
۹-۱-۱- ویژگی های نشان داده شده توسط ماتریس هم رخداد با استفاده از ویژگی مبتنی بر سطح خاکستری

همانطور که گفته شد در GLCM که مشخصه آماری مرتبه دو است، ویژگی هایی را از ماتریس هم رخداد استخراج کرده است. ویژگی مبتنی بر سطح خاکستری براساس تفاوت بین سطح خاکستری در پیکسل موردنظر و ارزش آماری یک نماینده از محیط اطراف استخراج می شود. بعنوان مثال غده های سرطانی دارای رنگ تضاد نسبت به اطراف خود هستند این ویژگی در توصیف تنوع سطح خاکستری در اطراف پیکسل مورد نظر انتخاب خوبی است. بعنوان مثال تصویر ۳-۸ یک ماتریس ۸*۸ را نشان می دهد. اگر ماتریس سمت چپ (B) را ماتریس تصویر اصلی فرض کنیم مشاهده می کنیم که حاوی عددهایی است، که این اعداد مربوط به سطح خاکستری هست که هر پیکسل از تصویر می تواند داشته باشد. اما در ماتریس سمت راست (C) سطر و ستون آن شامل بزرگترین عدد سطح خاکستری تصویر اصلی است. در ماتریس B احتمال وقوع اینکه عدد ۱ در کنار عدد ۱ قرار گرفته باشد چقدر هست: می بینیم که در ماتریس B، ۱ در هیچ حالتی کنار ۱ قرار نگرفته، پس در ماتریس نتیجه یعنی سمت راستی هم عدد صفر در خانه ی ۱ او ۱ قرار می گیرد. و یا مثلاً ۸ و ۱ در ماتریس B یکبار کنار هم قرار گرفتن پس در ماتریس سمت راست در خانه ی ۱ او ۸ عدد ۱ قرار می گیرد.



تصویر (۸-۳): نمایش ماتریس هم رخداد با استفاده از ویژگی مبتنی بر سطح خاکستری [38]

این مثال براساس جهت صفر درجه این ماتریس رو بررسی کرده است. همانطور که در تصویر ۳-۹ نشان داده شده است می‌توانیم از زوایای ۴۵، ۹۰ و ۱۳۵ درجه هم برای استخراج این ویژگی‌ها استفاده کنیم.



تصویر (۳-۹): نمایش استخراج ویژگی از زوایای مختلف [38]

از ماتریس‌های هم‌رخداد ویژگی‌های (مشخصه‌های هارالیک) مختلفی نظیر: انرژی، آنتروپی، کنتراست، همگنی و همبستگی استخراج می‌شود.

- ۱- انرژی^{۱۵}: مجموع توان دو عنصر موجود در ماتریس است.
- ۲- آنتروپی^{۱۶}: توصیف ویژگی‌های بافتی در یک همسایگی محلی.
- ۳- کنتراست^{۱۷}: مقایسه شدت نور بین پیکسل و همسایه‌ی آن.
- ۴- همگنی^{۱۸}: اندازه‌گیری میزان صافی تصویر.
- ۵- همبستگی^{۱۹}: ارتباط تراکم توزیع عناصر موجود در ماتریس هم‌رخداد نسبت به قطر ماتریس.

۱-۱۰- کاهش ویژگی‌ها

پیشرفت‌های به‌وجود آمده در جمع‌آوری ویژگی و قابلیت‌های ذخیره‌سازی در طی دهه‌های اخیر باعث شده در بسیاری از علوم با حجم بزرگی از ویژگی‌ها و اطلاعات روبرو شویم. در مقایسه با بسترهای داده‌ای قدیمی و کوچک‌تر، بسترهای داده‌ای امروزی چالش‌های جدیدی در تحلیل داده‌ها بوجود آورده‌اند. روش‌های

¹⁵ Energy

¹⁶ Entropy

¹⁷ Contrast

¹⁸ Homogeneity

¹⁹ Correlation

آماري سستی به دو دليل امروزه کارائی خود را از دست داده‌اند: علت اول افزایش تعداد مشاهدات است، و علت دوم که از اهمیت بالاتری برخوردار، افزایش تعداد متغیرهای مربوط به یک مشاهده می‌باشد.

تعداد متغیرهایی که برای هر مشاهده باید اندازه‌گیری شود ابعاد داده نامیده می‌شود. عبارت "متغیر" بیشتر در آمار استفاده می‌شود در حالی که در علوم کامپیوتر و یادگیری ماشین بیشتر از ویژگی استفاده می‌گردد. بسترهای داده‌ای که دارای ابعاد زیادی هستند علیرغم فرصت‌هایی که به وجود می‌آورند، چالش‌های محاسباتی زیادی را ایجاد می‌کنند. یکی از مشکلات داده‌ها با ویژگی زیاد این است که در بیشتر مواقع تمام ویژگی‌های داده‌ها برای یافتن دانشی که در داده‌ها نهفته است مهم و حیاتی نیستند. به همین دلیل در بسیاری از زمینه‌ها کاهش ویژگی

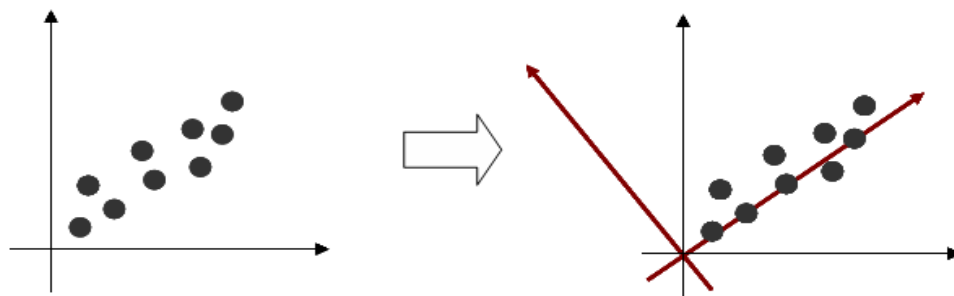
یکی از مباحث قابل توجه باقی مانده است.

۱-۱-۱- براساس PCA^{۲۰}

PCA، یک الگوریتم یا مدل خطی است که برای استخراج چند مولفه‌ی مهم و کاهش بعد فضای ویژگی استفاده می‌شود تا مشخصه‌های مهم تصویر تحت تشخیص را به‌دست آورد. ایده‌ی اصلی این روش در یافتن مولفه‌هایی متعامد است که در حد امکان بیشترین مقدار واریانس را بیان کند. در عمل، نوعاً می‌خواهیم که داده را با استخراج چند مولفه مهم در ابعاد کوچک‌تر نشان دهیم به طوری که بیشتر ساختارهای موجود در داده حفظ گردد. برای مثال اگر یک تصویر ۵۰ در ۵۰ داشته باشیم، این تصویر ۲۵۰۰ پیکسل دارد، در واقع هر پیکسل یک ویژگی محسوب می‌شود. در صورتی که بخواهیم بدون کاهش ویژگی، کلاس‌بندی را انجام دهیم لازم است تا تمام ۲۵۰۰ پیکسل هر تصویر با تصویر مرجع مقایسه شود، که کار بسیار پرمحاسبه و زمان‌گیری است و معمولاً عملی نیست. بنابراین لازم است تا با تکنیک‌هایی این تعداد ویژگی را کاهش دهیم، در واقع تنها ویژگی‌هایی را نگه داریم که تصاویر مربوط به دو کلاس در آن ویژگی تفاوت زیادی دارند. می‌توان یک تصویر را به صورت یک سیگنال نوشت و از PCA برای کاهش ویژگی استفاده کرد. برای این کار تمام پیکسل‌های تصویر را به ترتیب پشت سر هم قرار می‌دهیم تا یک بردار به‌دست آید. بنابراین برای هر تصویر یک بردار ویژگی به اندازه ۲۵۰۰ داریم. سپس با اعمال PCA و مشخص کردن تعداد دلخواه بعد نهایی، ویژگی‌های جدید را بدست می‌آوریم. اینکه تعداد ویژگی‌های نهایی چه مقدار باشد دست خودمان هست و باید با آزمون و خطا

²⁰ Principal Component Analysis

به بهترین عدد رسید. بعنوان مثال اگر تعداد ۱۰ را انتخاب کنیم، یعنی ۲۵۰۰ ویژگی را به ۱۰ ویژگی کاهش داده‌ایم. بهترین روش برای کاهش ویژگی به صورت خطی می‌باشد. یعنی با حذف ضرایب PCA تکنیک کم اهمیت بدست آمده از این تبدیل، اطلاعات از دست رفته نسبت به روش‌های دیگر کمتر است. البته کاربرد PCA محدود به کاهش ویژگی نمی‌شود و در زمینه‌های دیگری مانند شناسایی الگو و تشخیص چهره نیز مورد استفاده قرار می‌گیرد. در این روش محورهای مختصات جدیدی برای داده‌ها تعریف شده و داده‌ها براساس این محورهای مختصات جدید بیان می‌شوند. اولین محور باید در جهتی قرار گیرد که واریانس داده‌ها ماکسیمم شود (یعنی در جهتی که پراکندگی داده‌ها بیشتر است) دومین محور باید عمود بر محور اول به گونه‌ای قرار گیرد که واریانس داده‌ها ماکسیمم شود. به همین ترتیب محورهای بعدی عمود بر تمامی محورهای قبلی به گونه‌ای قرار می‌گیرند که داده‌ها در آن جهت دارای بیشترین پراکندگی باشند. در شکل ۳-۹ این مطلب برای یکسری از داده‌ها نشان داده شده است.



شکل (۳-۹): کاهش ویژگی با کمک پراکندگی داده‌ها بصورت خطی [37]