

فصل ششم

مشخصه‌ها

5-1: مقدمه

در یادگیری ماشین و شناسایی الگو، مشخصه یک ویژگی ابداعی قابل اندازه‌گیری و منحصر به فرد در یک پدیده است. انتخاب مشخصه مستقل و گسسته، کلید موفقیت هر الگوریتم شناسایی الگو در کلاس‌بندی است. مشخصه‌ها، معمولاً عددی هستند، اما مشخصه‌های ساختاری مانند رشته‌ها و گراف‌ها نیز در شناسایی الگوهای مفهومی استفاده می‌شوند.

مجموعه مشخصه‌های یک شی اغلب در یک بردار مشخصه قرار می‌گیرند. دلیل این گروه‌بندی آن است که می‌توان بر روی بردار، عملیات ریاضی انجام داد. مجموعه بردارهای مشخصه ممکن، یک فضای مشخصه را تشکیل می‌دهند. برای مثال، بسیاری از الگوریتم‌ها با ترکیب بردار مشخصه و بردار وزن‌ها با استفاده از تابع پیش‌گویی‌کننده خطی، به کلاس‌بندی یک شی در دسته مشخص امتیازی اختصاص می‌دهند.^[1]

یک مثال از بردارهای مشخصه، زمانی است که یک نقطه از تصویر به یک کلاس مشخص تعلق داشته باشد. تصور اینکه هر نقطه از تصویر، یک بردار مشخصه بر اساس مجموعه مناسب از مشخصه‌ها دارد، به این معناست که هر کلاس متناسب با فضای مشخصه قابل جداسازی است و مثال مرتبط دیگر زمانی است که بخواهیم شبکه عصبی را به تصاویر اعمال کنیم. داده ورودی به شبکه عصبی اغلب به صورت برداری از مشخصه‌های هر نقطه از تصویر داده می‌شود.

مفهوم مشخصه ضرورتاً مانند مفهوم متغیر توضیحی به کار برده شده در تکنیک‌های آماری مانند رگرسیون خطی است.^[2]

یک مشخصه خوب باید ویژگی‌های زیر را دارا باشد:

بازنمایی¹: یک مشخصه باید به خوبی نمایانگر داده‌ای باشد که مشخصه از آن انتخاب شده است. در واقع، باید یک توصیف فشرده و مفید از داده داشته باشد.

مشخصه منحصر بفرد²: یک مشخصه باید مقادیر متفاوت برای کلاس‌های متفاوت و مقادیر مشابه برای اشیا مشابه داشته باشد تا عمل کلاس‌بندی به خوبی انجام باشد.

¹ Representation

قابل توصیف³: یک مشخصه باید به سادگی قابل تفسیر و ترجمه به ویژگی‌های شی باشد که توسط متخصصان انسانی به کار برده می‌شود.
مفید و مناسب⁴: یک مشخصه باید با معنی و مناسب باشد.

2-5: انواع مشخصه‌ها

مشخصه‌ها از سه دید مختلف به سه دسته متفاوت تقسیم می‌شوند. انواع مشخصه‌ها از یک دید عبارتند از:

1. مشخصه حوزه زمان یا فضا⁵
 2. مشخصه حوزه فرکانس یا تبدیل⁶
- انواع دیگری از مشخصه‌ها به صورت زیر هستند:
1. مشخصه‌های محلی⁷
 2. مشخصه‌های سراسری⁸
- دسته آخر مشخصه‌ها، چهار گروه زیر می‌باشند:
1. مشخصه‌های بر اساس ظاهر⁹
 2. مشخصه‌های بعد از قطعه‌بندی¹⁰
 3. مشخصه‌های هندسی¹¹
 4. مشخصه‌های گشتاوری¹²

² Characteristic

³ Interpretable

⁴ Suitable

⁵ Time-domain or space-domain

⁶ Transformation or frequency domain

⁷ Local features(local descriptors)

⁸ Global features(Holistic features)

⁹ Appearance-based features

¹⁰ After segmentation features

¹¹ Geometric features

در ادامه به توضیح هر یک از این مشخصه‌ها می‌پردازیم.

5-2-1: مشخصه حوزه زمان یا فضا¹³

این نوع مشخصه، در سیگنال‌های صوتی بسیار کاربرد دارد. مشخصه‌هایی مانند دامنه و دوره تناوب از مشخصه‌های حوزه زمان یا فضا هستند.

5-2-2: مشخصه حوزه فرکانس یا تبدیل¹⁴

مشخصه‌هایی مانند ویژگی‌های طیفی و ضرایب در یک باند فرکانسی مشخص، از جمله مشخصه‌های حوزه فرکانس هستند.^[12]

5-2-3: مشخصه‌های محلی¹⁵

این مشخصه‌ها، الگو یا تصویر را با جزئیات آن توصیف می‌کنند. در واقع بافت آن را نشان می‌دهند. بنابراین مشخصه‌های ساده‌ای مانند کنتراست، هم‌بستگی¹⁶ و یک‌جوری¹⁷ که توصیف‌کننده بافت¹⁸ هستند، از جمله توصیف‌کننده‌های محلی یا مشخصه‌های SIFT¹⁹ می‌باشند. به عنوان مثال، خطوط عمودی یک تصویر، جز مشخصه‌های محلی آن است.^[5]

5-2-4: مشخصه‌های سراسری²⁰

مشخصه‌هایی هستند که کلیت یک الگو یا تصویر را نشان می‌دهند. مثلاً در تصویر یک برگ، طول و عرض آن از مشخصه‌های سراسری می‌باشند. یا تعیین رنگ کلی یک تصویر با

¹² Moment features

¹³ Time-domain or space-domain

¹⁴ Transformation or frequency domain

¹⁵ Local features

¹⁶ Correlation

¹⁷ Homogeneity

¹⁸ Texture

¹⁹ SIFT (Scale Invariant Feature Transform)

²⁰ Global features

رسم هیستوگرام‌های آن که آیا ترکیبات رنگ تصویر موردنظر، به سمت رنگ‌های روشن است یا تیره از مشخصه‌های سراسری است. بطور کلی معمولاً هیستوگرام از جمله مشخصه‌های سراسری یک تصویر است.

با وجود ویژگی‌های خوبی که دو مشخصه محلی و سراسری دارند، گاهی به تنهایی برای شناسایی یک تصویر به خوبی عمل نمی‌کنند. بنابراین می‌توان از ترکیب هر دو مشخصه استفاده کرد.^[5]

5-2-5: مشخصه‌های بر اساس ظاهر

مشخصه‌های بر اساس ظاهر، تصویر اصلی و تبدیلات آن مانند کوچکسازی تصویر²¹، اعمال فیلتر²²، تعیین آستانه²³ و ... را شامل می‌شوند که برای سیستم‌های OCR²⁴ استفاده می‌گردند. مثلاً در یک سیستم شناسایی زبان اشاره، برای شناسایی حرکات دست، به جای قطعه‌بندی و استخراج مشخصه‌های دست به تنهایی، مشخصه‌های کل تصویر که شامل تصویر انسان نیز می‌گردد به عنوان مشخصه‌های بر اساس ظاهر در نظر گرفته می‌شوند. شکل زیر را ببینید.^[14]



شکل (5-1): مشخصه‌های بر اساس ظاهر - تصویر چپ: تصویر اصلی - تصویر وسط: تصویر آستانه گذاری شده - تصویر راست: تصویر کوچکسازی شده

5-2-6: مشخصه‌های بعد از قطعه‌بندی

²¹ Down-scaling

²² Filtering

²³ Thresholding

²⁴ Optical Character Recognition (OCR)

برای شناسایی یک تصویر، ابتدا آن را قطعه‌بندی می‌کنند و سپس مشخصه‌های این قطعه‌ها را در نظر می‌گیرند. مثلاً در سیستم شناسایی زبان اشاره، برای شناسایی دست انسان، پس از قطعه‌بندی تصویر، مشخصه‌های مکانی از تصویر که دست در آنجا قرار دارد را به عنوان بردار مشخصه در نظر می‌گیرند. ایراد استفاده از این نوع مشخصه این است که اگر در مرحله اول و قطعه‌بندی، مکان صحیح شی مورد نظر (در این مثال دست) شناسایی نشود، مراحل دیگر شناسایی نیز درست انجام نخواهد گرفت.^[6]

5-2-7: مشخصه‌های هندسی

این مشخصه‌ها اطلاعات خاصی را در ارتباط با موقعیت اجسام درون یک تصویر، مانند شکل آنها یا پیکربندی‌شان نشان می‌دهند. در یک سیستم شناسایی زبان اشاره، مشخصه‌هایی مانند مکان دست، طول پیرامون دست، مرکز ثقل آن، باز یا بسته بودن دست و جهت دست به عنوان مشخصه‌های هندسی در نظر گرفته می‌شوند.^[6]

5-2-8: مشخصه‌های گشتاوری

مشخصه‌های تغییرناپذیر، مشخصه‌هایی هستند که با تبدیلات هندسی مانند تغییر اندازه، چرخش و... ثابت می‌مانند. مشخصه‌های گشتاوری این مشخصه‌ها را نشان می‌دهند.^[10] در حالت کلی گشتاورها در یک تصویر را می‌توان به دسته‌های زیر تقسیم بندی کرد:

- **گشتاورهای سطری:**

ویژگی‌های ساده‌ای که به عنوان گشتاورهای سطری شناخته می‌شوند، مساحت برای تصاویر دودویی یا جمع سطوح خاکستری و مرکز ثقل هستند.

- **گشتاورهای مرکزی:** ثابت‌های جابجا شده²⁵ از گشتاورهای مرکزی هستند.

- **گشتاورهای ثابت اندازه.**

- **گشتاورهای ثابت چرخش.**^[3]

3-5: مشخصه‌های معمول در حوزه‌های گوناگون

1-3-5: دست نوشته^[3]

در مرحله استخراج مشخصه‌ها، هر کاراکتر به صورت یک بردار مشخصه نمایش داده می‌شود. هدف اصلی استخراج مشخصه‌ها این است که مجموعه‌ای از مشخصه‌ها که نرخ شناسایی را با حداقل تعداد عناصر، ماکزیمم می‌کند، ایجاد کند.

با توجه به طبیعت دست‌نوشته‌ها با نرخ بالای تغییرپذیری مشخصه‌ها، یافتن مشخصه‌ها کار بسیار دشواری است. متدهای استخراج مشخصه در این مورد، بر اساس سه نوع مشخصه است:

- 1- مشخصه‌های آماری²⁶
- 2- مشخصه‌های ساختاری²⁷
- 3- گشتاورها و تبدیلات سراسری²⁸

1-1-3-5: مشخصه‌های آماری

برای نمایش تصویر یک کاراکتر به وسیله توزیع‌های آماری نقاط در برخی حوزه‌ها، باید به دگرگونی‌های نوع (style) توجه کرد. مشخصه‌های آماری اصلی که برای نمایش کاراکتر به کار می‌روند، عبارتند از:

- 1- تقسیم نواحی²⁹

²⁶ Statistical

²⁷ Structural

²⁸ Global transformations and moments

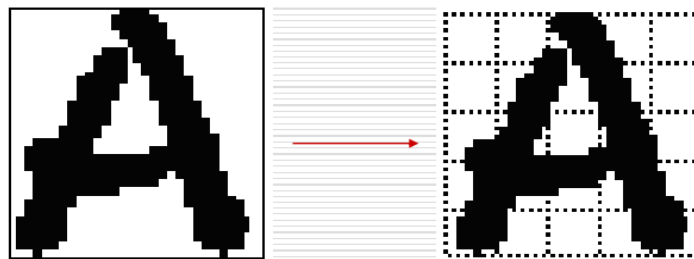
²⁹ Zoning

Projections and profiles -2

Crossings and distances -3

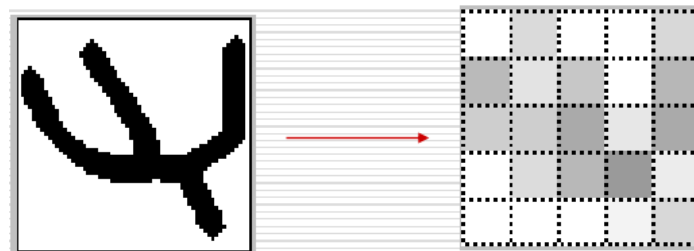
5-3-1-1: تقسیم نواحی

تصویر کاراکتر به نواحی $N \times M$ تقسیم می‌شود. از مشخصه‌های هر ناحیه برای تشکیل بردار مشخصه استفاده می‌شود. هدف تقسیم نواحی، به دست آوردن ویژگی‌های محلی به جای به دست آوردن ویژگی‌های سراسری است.



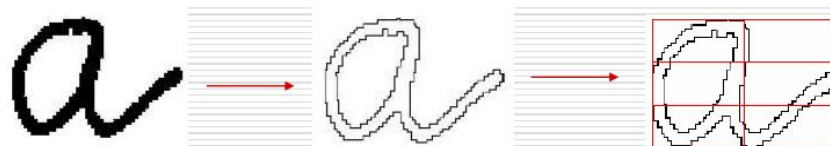
شکل (5-2): تقسیم تصویر یک کاراکتر به نواحی (zone)های مختلف

نواحی می‌توانند بر اساس تراکم پیکسل‌ها تقسیم شوند. در واقع، تعداد پیکسل‌های پشت‌زمینه یا تعداد پیکسل‌های نرمال شده در هر سلول به عنوان یک مشخصه در نظر گرفته می‌شوند. در شکل 5-3، مربع‌های تیره‌تر تراکم بیشتری از پیکسل‌های هر ناحیه (zone) را نشان می‌دهند.



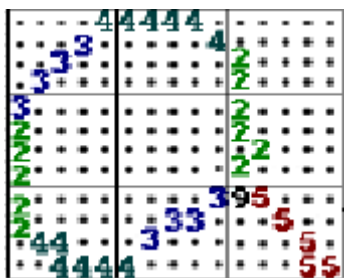
شکل (5-3): تراکم پیکسل‌ها در هر ناحیه (zone). مربع‌های روشن‌تر تراکم کمتری دارند.

تقسیم نواحی می‌تواند بر اساس خطوط پیرامونی تصویر کاراکتر باشد.



شکل 4-5: خطوط پیرامون تصویر یک کاراکتر

برای هر ناحیه‌ای که حد فاصل آن مشخص شده، یک هیستوگرام با آنالیز پیکسل‌های همسایگی 3×3 به دست خواهد آمد. سپس بر اساس ساختمان تصویر کاراکتر، بخش‌های خطی منحصر به فرد جدا می‌شوند و اطلاعات بخش خطی با یک عدد هدایت‌کننده، برچسب‌گذاری می‌گردد.



شکل (5-6): برچسب‌گذاری

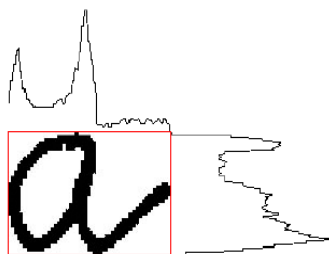


شکل (5-5): ساختمان تصویر حرف a

2-1-1-3-5: Projections and profiles

ایده اصلی برای به کار بردن projections در یک تصویر کاراکتر این است که آنها سیگنال‌های دو بعدی هستند که می‌توانند به صورت سیگنال یک‌بعدی نشان داده شوند. این مشخصه‌ها اگرچه به نویز و تغییر شکل وابسته نیستند اما نسبت به چرخش وابسته‌اند.

هیستوگرام‌های projection تعداد پیکسل‌های هر ستون و هر سطر تصویر کاراکتر را می‌شمارند. این هیستوگرام‌ها می‌توانند کاراکترهایی مانند m و n را از هم جدا کنند.



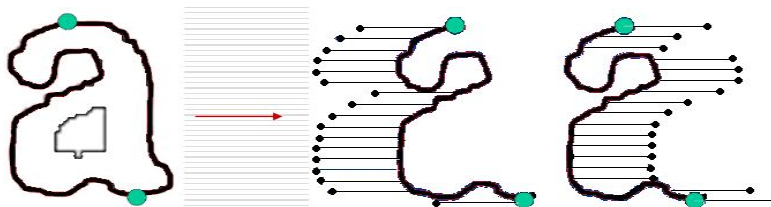
شکل (5-7): هیستوگرام‌های projection کاراکتر a به صورت سطری و ستونی.

Profile تعداد پیکسل‌های (فاصله) بین باکس اطراف تصویر کاراکتر و لبه خود کاراکتر را می‌شمارد. Profile به خوبی شکل‌های خارجی کاراکتر را توصیف می‌کند و تعداد زیادی از کاراکترها را با این مشخصه از یکدیگر تشخیص می‌دهد مانند p و q.



شکل (5-8): محاسبه فاصله بین پیکسل‌های تصویر کاراکتر و لبه آن.

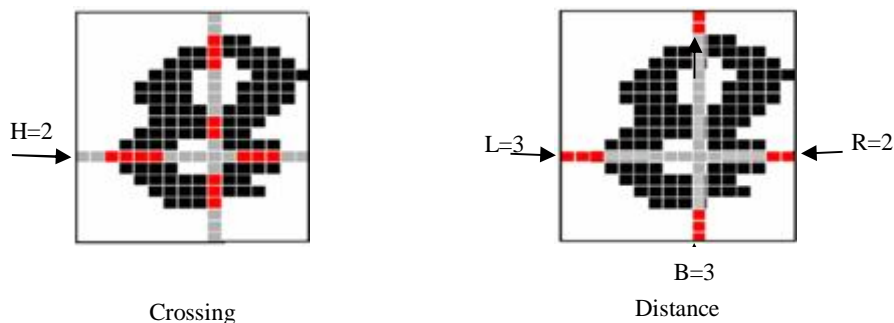
همچنین، profile می‌تواند برای استخراج خطوط حاشیه کاراکتر، قرار دادن بالاترین و پایین‌ترین نقاط خطوط حاشیه و محاسبه profile‌های بیرون و داخل حاشیه به کار برده شوند.



شکل (5-9): استفاده از profile برای یافتن بالاترین و پایین‌ترین نقاط تصویر کاراکتر.

Crossings and distances :3-1-1-3-5

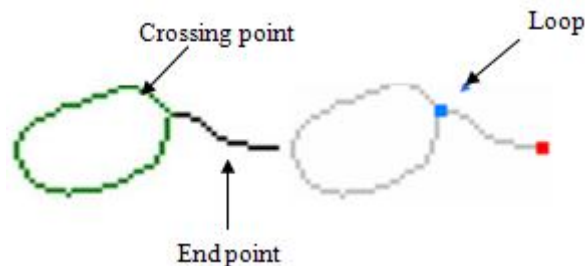
Crossings تعداد گذرها از پیکسل‌های پشت زمینه به پس زمینه را از میان خطوط افقی و عمودی تصویر کاراکتر می‌شمارد و Distance ها، فاصله‌های بین اولین پیکسل تصویر از لبه‌های بالا و پایین از میان خطوط عمودی، و لبه‌های چپ و راست از میان خطوط افقی را محاسبه می‌کند. هر دو روش، مشخصه‌های خوبی برای کاراکترها هستند.



شکل (5-10): Crossings and distances

2-3-5: مشخصه‌های ساختاری

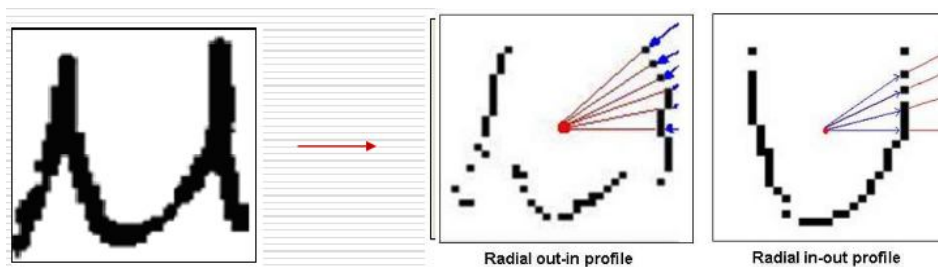
کاراکترها می‌توانند به وسیله مشخصه‌های ساختاری با مقاومت بالا در مقابل اعوجاجات و تغییرات نوع نمایش داده شوند. این نوع از نمایش ممکن است همچنین دانشی درباره ساختار یک شی را کدگذاری کند یا دانشی را فراهم کند که چه ترتیبی از مولفه‌های شی را بازآرایی می‌کنند. مشخصه‌های ساختاری بر اساس ویژگی‌های توپولوژیکی و هندسی کاراکتر مانند **cross points**، **aspect ratio**، حلقه‌ها، نقاط **branch**، **strokes** و جهات آنها، خمیدگی بین دو نقطه، منحنی‌های افقی در بالا یا پایین، هستند.



شکل (11-5): Loop و نقطه‌ی انتها و crossing point کاراکتر a.

به عنوان مثال، یک روش استخراج مشخصه ساختاری برای شناسایی کاراکترهای دست‌نویس یونانی بر اساس سه نوع از مشخصه‌ها وجود دارد:

- 1- هیستوگرام‌های projection عمودی و افقی.
- 2- هیستوگرام‌های شعاعی.
- 3- Profile های شعاعی in-out و out-in. (شکل 12-5 این Profile را نشان می‌دهد).



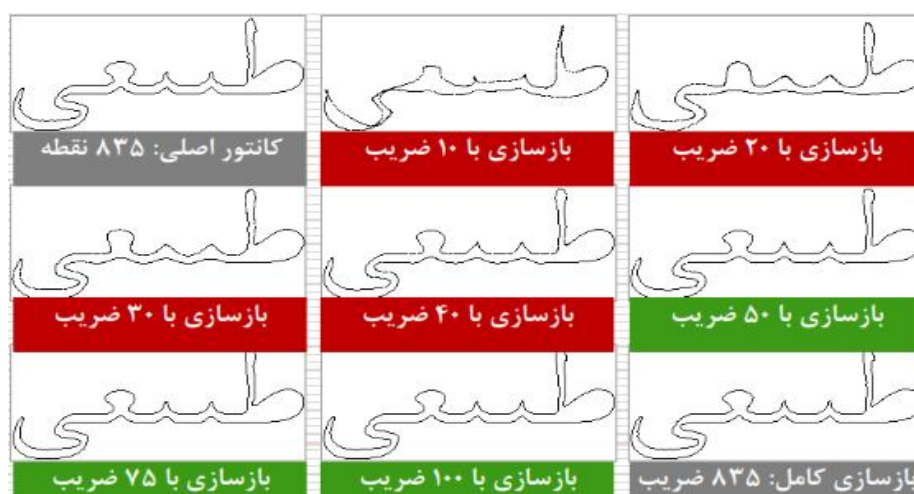
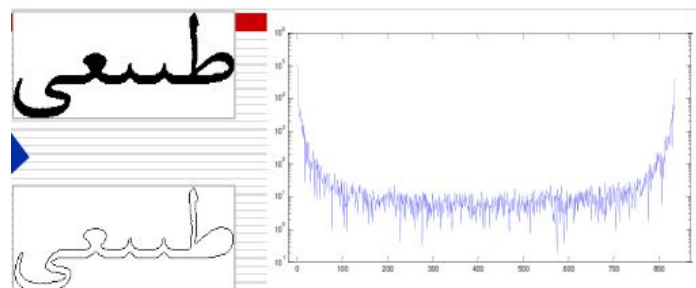
شکل (12-5): profile های شعاعی in-out و out-in.

Global Transformations and Moments: 3-3-5

تبدیل فوریه³⁰ (FT) کانتور تصویر محاسبه می‌شود. از آنجایی که اولین n ضریب FT، می‌توانند برای دوباره ساختن کانتور³¹ استفاده گردند، بنابراین این n ضریب به عنوان بردار مشخصه‌های n بعدی که کاراکتر را نشان می‌دهند، در نظر گرفته می‌شوند.

³⁰ Fourier Transform

³¹ contour



شکل (5-13): بازسازی تصویر با ضرایب تبدیل فوریه

گشتاورهای زرنیک و مرکزی، فرآیند شناسایی ثابت‌های اندازه‌گیری، ترجمه و جهت یک شی را انجام می‌دهند. تصویر اصلی می‌تواند به طور کامل از ضرایب گشتاوری دوباره ساخته شود.

گشتاورهای زرنیک در سال 1943 توسط آقای زرنیک (فیزیکدان) مطرح شد. آقای ختن‌زاد در سال 1990 از اندازه آنها برای بازشناسی تصاویر دوبعدی استفاده کرد. این گشتاور دارای تبدیل معکوس است و امروزه رواج زیادی یافته است. اما متأسفانه سرعت استخراج ویژگی در آن بسیار کند است و کارایی چندانی در مورد ارقام فارسی ندارد. با استفاده از روابط زیر، گشتاور زرنیک محاسبه می‌شود:

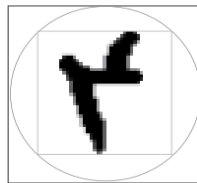
$$- \quad = (+ 1) / \sum_{-} \sum_{-} \mathbb{I} (,) [- (,)] * , \quad ^2 + ^2 \leq 1 \mathbb{I}$$

$$- \quad (,) = - \quad (,) ^ (\mathbb{I} \mathbb{I} ^ (-1) (/ -)) \quad h$$

$$= \sqrt{-1}, \quad |x| \leq 1, \quad -1 \leq x \leq 1$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n (x^2 + 1)^{n/2} (x - 1)^{n/2}}{n! (x + 1)^{n/2} (x - 1)^{n/2}}$$

در محاسبه گشتاور زرنیک، x و y باید داخل دایره واحد باشند و نیز تنها قسمتی از تصویر را می‌توان بازسازی کرد که داخل دایره مذکور قرار دارد. تصویر اصلی را باید داخل یک دایره محاطی قرار داده و اطراف آن را با زمینه پر کنیم:



شکل (5-14): تصویر قرار گرفته در دایره واحد برای محاسبه گشتاور زرنیک

از اندازه گشتاور زرنیک برای استخراج مشخصه از تصویر استفاده می‌کنند که مستقل از چرخش است. طول بردار مشخصه 135 است که معمولاً گشتاورهای مرتبه 1 تا 15 را به کار می‌برند. در این صورت نرخ بازشناسی ارقام 94,86% می‌باشد. همچنین می‌توان تنها از مولفه حقیقی گشتاور زرنیک استفاده کرد که طول بردار مشخصه برابر 135 و نرخ بازشناسی 96,14% خواهد بود. در صورت استفاده از هر دو مولفه حقیقی و موهومی گشتاور زرنیک، طول بردار مشخصه برابر 270 و نرخ بازشناسی 99,13% خواهد بود. مزیت این روش، یادگیری خوب داده‌های آموزش در صورت استفاده از هر دو مولفه حقیقی و موهومی است. قدرت تعمیم گشتاور زرنیک خوب نیست و روی شناسایی ارقام 3 و 4 خطای زیادی دارد. (94,3% و 94,4%) و سرعت استخراج مشخصه آن 66 بار کندتر از مکان مشخصه است.

5-3-2: سیگنال صوتی

دو نوع از اطلاعات در صوت وجود دارند. بخش مفهومی که اطلاعات زبان‌شناسی را در خود دارد که مطابق با قوانین تلفظ یک زبان هستند. اطلاعات **paralinguistic** به حالت حس

سخنگو یا صحبت‌کننده اشاره می‌کنند. در شناسایی حالت حسی یک صوت، اولین مرحله شناسایی مشخصه‌های paralinguistic است. مشخصه‌هایی که می‌توان برای شناسایی یک سیگنال صوتی و دو بخش اطلاعاتی آن استفاده کرد، عبارتند از:

1- ویژگی‌های طیفی³²

2- (MFCC)³³

به توضیح هریک از این مشخصه‌ها می‌پردازیم:

5-3-2-1: ویژگی‌های طیفی

از این مشخصه‌ها، برای شناسایی حالات حسی صوت استفاده می‌گردد. Williams و Stevens (1981)، از لحاظ فیزیولوژی تولید صوت، تغییرات سیستم عصبی را در هنگام ترس، خوشحالی و عصبانیت در نظر گرفتند. در نتیجه، فشار خون و ضربان قلب افزایش یافته و دهان خشک می‌شود. صدا نیز بلند، سریع و دارای انرژی فرکانسی زیادی است. در مقابل در هنگام ناراحتی، ضربان قلب و فشار خون کاهش یافته و صدا با انرژی فرکانسی کم و آرام تولید می‌شود. بنابراین تغییرات فیزیولوژیکی، تغییرات انرژی، توزیع انرژی در طیف فرکانسی و فرکانس و طول پازهای سیگنال صوتی را ایجاد می‌کنند. با توجه به مشخصه‌های به دست آمده از حالات حسی، سه نوع مختلف از مشخصه‌ها هستند که حالات حسی را شرح می‌دهند:

1- فرکانس اصلی (F0).

2- متغیرهای شنوایی پیوسته.

3- کیفیت صدا.

کانتور فرکانس اصلی، برای توصیف تغییرات فرکانس اصلی در الگوهای هندسی به کار می‌رود. متغیرهای شنوایی پیوسته شامل اندازه فرکانس اصلی مانند شدت، نرخ صحبت³⁴ و توزیع انرژی

³² Spectral characteristics

³³ Mel-frequency cepstrum coefficient

³⁴ Speaking rate

از طیف است. مواردی که برای توصیف کیفیت صدا استفاده می‌شوند عبارتند از: **tense, harsh** و **breathy**. این سه نوع از متغیرهای صوت با یکدیگر در ارتباطند. برای مثال، اطلاعات فرکانس اصلی و کیفیت صدا توسط متغیرهای شنوایی پیوسته گرفته می‌شوند.

متاسفانه، بعضی از حس‌های مشخص، ویژگی‌های خیلی شبیه به هم در مجموعه مشخصه‌ها دارند. از این رو، سیستم‌های کلاس‌بندی بر اساس این مشخصه‌ها قادر نیستند بیش از دو سیگنال را به درستی دسته‌بندی کنند.

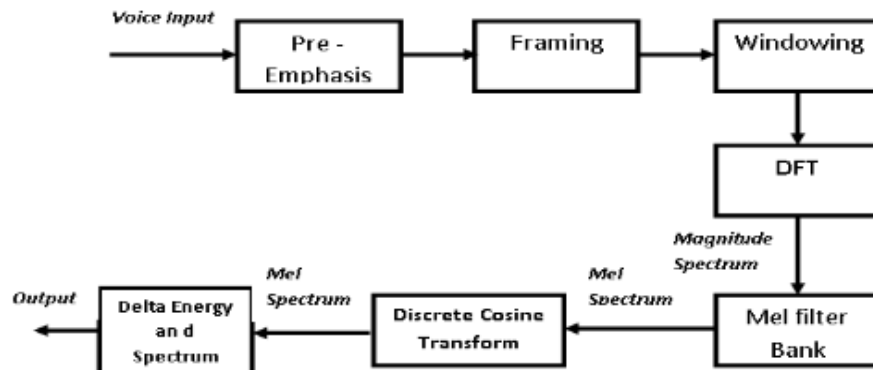
2-2-3-5: (MFCC)

برای شناسایی صوت، LPCC و MFCC انتخاب‌های مناسبی برای مشخصه‌هایی هستند که می‌توان توسط آنها بخش تلفظ یا زبان‌شناسی یک صوت را شناخت.^[15]

در پردازش صوت، **MFC** یک نمایش از طیف صوتی با توان کم بر اساس یک تبدیل خطی کسینوسی از یک طیف با توان بزرگ روی یک **scalemel** غیرخطی از فرکانس است.

MFCCها، ضرایبی هستند که یک MFC را می‌سازند. MFCCها به عنوان مشخصه‌هایی در یک سیستم شناسایی صوت مانند سیستم‌هایی که به طور خودکار ارقام گفته‌شده به یک تلفن را شناسایی می‌کنند، هم‌چنین در تشخیص صدا برای شناسایی افراد از روی صداهایشان کاربرد دارد.^[4]

MFCC بر اساس فرکانس‌های دریافتی انسان است که فرکانس‌های بیش از 1 کیلوهرتز را شامل نمی‌شود. به عبارت دیگر، MFCC در بازه فرکانس شنوایی انسان عمل می‌کند. MFCC دو نوع فیلتر دارد که به صورت خطی در فرکانس کم زیر 1000 هرتز قرار گرفته و فضای لگاریتمی آن بالای 1000 هرتز است. روند کلی MFCC به صورت زیر است:



شکل (5-15): بلوک دیاگرام MFCC.

بنابراین، MFCC شامل 7 مرحله است که هر مرحله توابع و نگرش‌های ریاضی خود را دارد:

1- Pre-emphasis:

این مرحله، عبور سیگنال از یک فیلتر فرکانس بالا را انجام می‌دهد که باعث می‌شود انرژی سیگنال در فرکانس بالاتری قرار گیرد.

2- Framing:

عمل قطعه بندی سیگنال برای تبدیل آنالوگ به دیجیتال (ADC) به فریم کوچک با اندازه‌ای در بازه 20 تا 40 میلی ثانیه را گویند. به عبارت دیگر سیگنال صوتی به تعداد نمونه‌ها فریم‌بندی می‌شود.

3- Hamming windowing:

پنجره Hamming تمام نزدیک‌ترین خطوط فرکانسی را مشخص می‌کند. معادله پنجره hamming به صورت زیر است:

$$\begin{aligned}
 () &= () \times () \\
 () &= 0.54 - 0.46 \left(2^{(n/2 - 1)} \right) () \\
 &= 0.54 - 0.46 \left(2^{(n/2 - 1)} \right) \quad 0 \leq n \leq N-1
 \end{aligned}$$

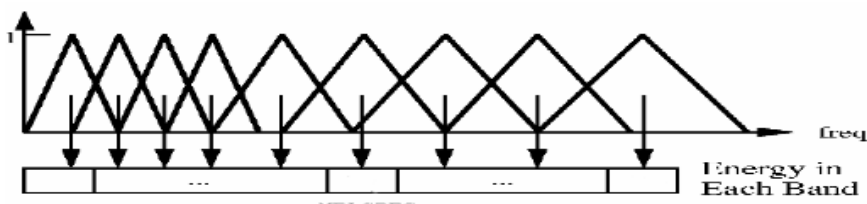
که () سیگنال خروجی، () سیگنال ورودی، () پنجره همینگ و تعداد نمونه‌ها در هر فریم است.

4- Fast Fourier Transform:

برای تبدیل N نمونه از فضای زمان به فضای فرکانسی استفاده می‌شود.

5- Mel Filter Bank Processing:

بازه فرکانسی سیگنال بسیار گسترده است و سیگنال صوتی از مقیاس خطی پیروی نمی‌کند. بنابراین بانکی از فیلترها بر اساس **mel scale** مطابق شکل اعمال می‌شوند.



شکل (5-16): بانک فیلترها برای تبدیل به مقیاس mel.

این شکل نشان می‌دهد که یک مجموعه از فیلترهای مثلی برای مجموع وزن‌دار مولفه‌های طیفی فیلتر استفاده می‌شوند که خروجی به مقیاس mel تقریب زده شود.

6- Discrete Cosine Transform:

فرآیندی برای تبدیل طیف mel به حوزه زمان با استفاده از تبدیل گسسته کسینوسی است. نتیجه این تبدیل Mel Frequency Cepstrum Coefficient (مجموعه‌ای از ضرایب که بردارهای شنوایی نامیده می‌شوند) است. بنابراین هر سیگنال صوتی ورودی به یک بردار شنوایی تبدیل می‌شود.

7- Delta Energy and Delta Spectrum:

شیب و مشتق سیگنال صوتی و فریم‌ها در تبدیل‌ها تغییر می‌کنند. بنابراین نیاز است که مشخصه‌هایی برای تغییر مشخصه‌های **cesptal** اضافه شوند. 13 مشخصه دلتا و سرعت (12 مشخصه ی **cesptal** و یک انرژی) و 39 مشخصه شتاب به بردار مشخصه اضافه می‌گردند.

13 مشخصه دلتا تغییر بین فریم ها و spectralهای متناظر و هر یک از 39 مشخصه شتاب تغییر بین فریم‌ها با مشخصه دلتا متناظرشان را نشان می‌دهند.^[16]

5-3-3: شناسایی چهره

تکنولوژی شناسایی چهره (FRT) مطالعه فرآیند شناسایی یک چهره در الگوی تصویر داده شده و جدا کردن آن از بقیه تصویر (پشت زمینه) است. در حقیقت چهره‌ها در موقعیت، اندازه، مقادیر رنگ پوست، تاثیرات روشنایی، حالات چهره و وجود یا عدم وجود مو یا عینک با یکدیگر متفاوتند. همه این موارد سبب می‌شوند که قادر به تشخیص درست چهره‌ها نباشیم. فرآیند شناسایی چهره معمولاً سه مرحله دارد: در اولین مرحله محل چهره پیدا می‌شود که به آن face localization می‌گویند. در مرحله دوم مشخصه‌های چهره استخراج می‌شوند که این مرحله خیلی بحرانی است چرا که خروجی نهایی و نتیجه به این مرحله وابسته است. آخرین مرحله کلاس‌بندی تصویر صورت بر اساس مشخصه‌هاست.

تعداد زیادی از الگوریتم‌ها برای استخراج مشخصه‌های چهره وجود دارند که هر کدام مشخصه‌های خاصی را انتخاب می‌کنند. دو نوع از مشخصه‌های معمول مشخصه‌های internal و external هستند. منظور از مشخصه‌های internal، چشم، دهان و بینی و مشخصه‌های external، مو و پیرامون صورت است.^[13]

مشخصه‌های دیگری که برای شناسایی چهره به کار برده می‌شوند، می‌توانند مشخصه‌های بعد از قطعه‌بندی باشند. می‌دانیم که تکنیک قطعه‌بندی تصویر به دو دسته تقسیم می‌شود:

1- تکنیک‌های بر اساس پیرامون³⁵: عمومی‌ترین متد قطعه‌بندی بر اساس مرز³⁶ پویای هیستوگرام تصویر است.

2- تکنیک‌های بر اساس ناحیه³⁷: این الگوریتم‌ها شامل رشد ناحیه، جداسازی و ادغام نواحی می‌شوند.

³⁵ boundary-based

³⁶ threshold

³⁷ region-based

برای جداسازی تصویر چهره از پشت زمینه، کانتورهای فعال یا منحنی شکل را به کار می برند. بدین منظور از یک تابع اسپلاین پیوسته برای جابجایی تصویر منحنی استفاده می کنند تا مقدار تابع انرژی مرتبط با حالت اولیه منحنی مینیمم شود³⁸. منحنی نهایی، پیرامون خارجی شی را دنبال خواهد کرد. اما این روش حساس به نویز است و یافتن یک نقطه اولیه خوب برای تخمین بسیار سخت است.

برای قطعه بندی تصویر چهره، از جابجایی فضای رنگ ها به نام HSV³⁹ و اطلاعات شکل استفاده می کنند. ابتدا، نواحی شبیه پوست بدن بر اساس اطلاعات مولفه های hue و saturation قطعه بندی می شوند و سپس یک جستجوی خوب در هر یک از نواحی برای کشف اشکال بیضی شکل استفاده می شود. این روش خطاهای زیادی ایجاد می کند. به علاوه ایجاد یک مدل رنگ پوست یک کار آسان نیست. برای مثال، Ref، از 43 میلیون پیکسل پوست از 900 تصویر برای آموزش مدل رنگ پوست استفاده کرد. او مجموعه ای شامل نواحی پوست را برای تولید مدل پوست قطعه بندی کرد.

استفاده از template روش دیگری برای شناسایی چهره و قطعه بندی آن است. Tan و Wang از تمپلیت استفاده می کردند: دو **تمپلیت** چشم و یک **تمپلیت** دهان برای شناسایی چهره و مکان یابی مشخصه های اصلی آن استفاده شدند. سپس دو تمپلیت گونه و یک تمپلیت چانه برای شناسایی کانتور چهره به کار گرفته شدند. در این روش چهره هایی با سایه یا شرایط بد روشنایی قابل شناسایی نیستند.^[6]

Zhang از دو نوع مشخصه برای شناسایی حالات چهره از تصاویر ایستا استفاده کرد: موقعیت هندسی 34 نقطه وابسته و یک مجموعه از ضرایب wavelet در این نقاط. این دو نوع مشخصه هر دو به صورت مستقل ولی با هم در یک شبکه پرسپترون چندلایه به کار برده شدند. Guo و Dyer یک نمایش چهره مشابه ترکیب شده با برنامه خطی (LP) را برای کلاس بندی و انتخاب مشخصه های همزمان استفاده کردند.^[9]

³⁸ Energy minimization

³⁹ (Hue, saturation, value)

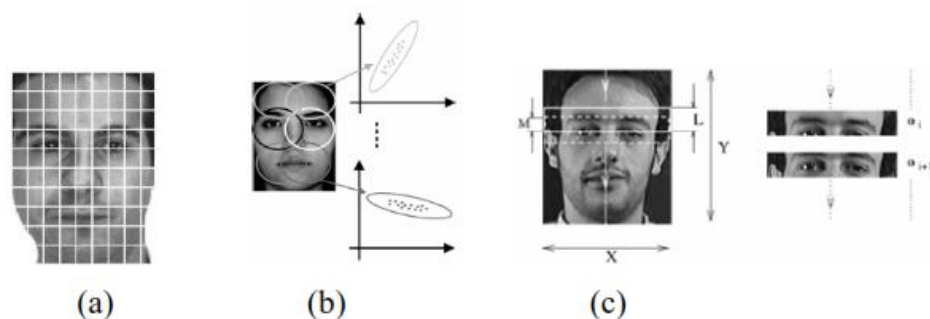
گاهی برای شناسایی چهره از مشخصه‌های **holistic**، گاهی مشخصه‌های محلی و گاهی ترکیبی از هر دو استفاده می‌کنند. مشخصه‌های holistic مانند مشخصه‌های عمومی هستند و تمام تصویر را برای شناسایی انتخاب می‌کنند؛ به همین سبب بردار مشخصه ابعاد زیادی دارد. این روش دو مزیت دارد: اول، اطلاعات تصویر و بافت آن را با جزئیات کامل شامل می‌شود و دوم اینکه نسبت به توصیفگرهای محلی جنبه‌های عمومی بیشتری را در نظر می‌گیرد اما از آنجایی که تنها یک بردار برای هر کلاس وجود دارد، اختلافات بین کلاسها به تعداد زیادی تکنیک‌های PR نیاز دارد که به طور مستقیم قابل تخمین و شناسایی نیستند.

استفاده از مشخصه‌های محلی برای شناسایی چهره تاریخچه طولانی دارد. در مقایسه با مشخصه‌های holistic، مشخصه‌های محلی به دلایل زیر برای مساله‌ای با وجود تنها یک نمونه برای هر نفر مناسب‌تر هستند: اولاً با مشخصه‌های محلی، چهره واقعی با مجموعه‌ای از بردارهای مشخصه محلی بعد کم نمایش داده می‌شود. بنابراین مشکل بعد می‌تواند از همان ابتدا نادیده گرفته شود. ثانیاً، متدهای محلی انعطاف‌پذیری بیشتری برای شناسایی یک چهره بر اساس بخش‌هایی از آن را دارند. بنابراین مشخصه‌های مشترک داخل کلاسی به آسانی قابل شناسایی اند. ثالثاً، مشخصه‌های چهره متفاوت می‌توانند تنوع طبقه‌بندی را افزایش دهند که برای شناسایی چهره مفید است. متدهای محلی به دو دسته تقسیم می‌شوند: متد براساس مشخصه محلی⁴⁰ و متد بر اساس ظاهر محلی⁴¹. در اولین روش، ابتدا مشخصه‌های محلی به دست آمده و سپس مشخصه‌ها روی نقاط مشخصه‌های محلی استخراج می‌شوند. در روش دوم، تصویر چهره را به sub-region‌هایی بر اساس مشخصه‌های محلی که مستقیماً استخراج می‌شوند تقسیم می‌کنند. در ابتدا نواحی محلی باید تعریف شوند که برای این منظور دو فاکتور را در نظر می‌گیرند: شکل و اندازه نواحی محلی. ساده‌ترین و عمومی‌ترین شکل ناحیه، پنجره مستطیلی است. پنجره‌ها می‌توانند با یکدیگر هم‌پوشانی داشته باشند یا خیر. در کنار مستطیل‌ها شکل‌های دیگر مانند بیضی و نوار⁴² هم استفاده می‌شوند. تعداد نواحی محلی هم تحت تاثیر مستقیم تعداد مشخصه‌های محلی است.

⁴⁰ local feature-based

⁴¹ local appearance-based

⁴² strip



شکل (5-17): شکل‌های محلی معمول برای شناسایی چهره: a)rectangular shape, b)ellipse shape, c)stripshape.

بعد از تعریف نواحی محلی، باید تصمیم بگیریم که چطور اطلاعات آنها را نمایش دهیم. عمومی‌ترین مشخصه‌های استفاده شده، مشخصه‌های gray-level و تنوعی از مشخصه‌های مشتقی مانند Gaborwavelet، Harrwavelet و مشخصه‌های **fractal** هستند که متناسب با کاربردهای متفاوت یکی از مشخصه‌ها موفق عمل می‌کند. مشخصه gray-level ساده‌ترین مشخصه بدون فقدان اطلاعات بافت است درحالی‌که مشخصه‌های Gabor و سایر مشخصه‌های مشتقی در مقابل تغییرات روشنایی و بعضی جابجایی‌های هندسی مقاوم‌ترند.

جدول زیر، مقایسه بین حساسیت‌های مشخصه‌های محلی و عمومی را نسبت به تغییرات نشان می‌دهد:

جدول (5-1): مقایسه بین حساسیت‌های مشخصه‌های محلی و عمومی نسبت به تغییرات.

Variation factors	Local features	Holistic features
Small variations	not sensitive	Sensitive
Large variations	sensitive	very sensitive
Illumination	very sensitive	Sensitive
Expressions	not sensitive	Sensitive
Pose	sensitive	very sensitive
Noise	Very sensitive	Sensitive
Occlusion	not sensitive	very sensitive

در صورت استفاده از هر دو مشخصه با هم، مسائلی مطرح می‌شود که روی عملکرد تاثیر می‌گذارد: چگونه مشخص کنیم که کدام یک از این مشخصه‌ها باید با یکدیگر ترکیب شوند و اینکه چگونه آنها را با هم ترکیب کنیم. همانطور که در جدول بالا حساسیت‌های این مشخصه‌ها را نسبت به تغییرات دیدیم می‌توان با توجه به کاربرد و تغییرات در آن کاربرد، مشخصه‌هایی که کمتر تحت تاثیر تغییر قرار می‌گیرند برگزید و با یکدیگر ترکیب کرد. این عمل می‌تواند یک راه موثر برای کاهش پیچیدگی طبقه‌بندی‌کننده‌ها و بهبود توانایی عمومی آنها باشد.^[7]

همانطور که پیش‌تر توضیح دادیم، مشخصه‌های SIFT⁴³ مشخصه‌هایی هستند که نسبت به اندازه و جهت ثابتند و در یک تصویر بسیار متمایزند. این مشخصه‌ها برای شناسایی تصاویر چهره نیز کاربرد فراوانی دارند. در 4 مرحله از یک تصویر استخراج می‌شوند: اولین مرحله مکان نقاط **interest** را با کشف مینیمم و ماکزیمم مجموعه‌ای از اختلافات فیلترهای گوسین (DoG) اعمال شده به اندازه‌های مختلف سرتاسر تصویر، محاسبه می‌کند. سپس این مکانها با نادیده گرفتن نقاط با کنتراست کم به روز رسانی می‌شوند. یک جهت برای هر نقطه کلیدی بر اساس

⁴³ ScaleInvariant Feature Transform

مشخصه‌های محلی تصویر تعیین می‌شود. سرانجام یک توصیفگر مشخصه محلی در هر نقطه کلیدی محاسبه می‌شود. این توصیفگر، بر اساس گرادیان محلی تصویر است که مطابق با جهت نقطه کلیدی برای فراهم کردن ثبوت جهتی⁴⁴ جابجا می‌شود. هر مشخصه یک بردار با طول 128 است که همسایگی‌های نقطه کلیدی را نشان می‌دهد.^[8] شکل 19، تصاویری از چهره‌ها و مشخصه‌های SIFT آنها را نشان می‌دهد:



الف) پنج تصویر نمونه



ب) تصاویر نمونه با مشخصه‌های SIFT آنها.
 شکل (5-18): چهره‌ها و مشخصه‌های SIFT آنها.

⁴⁴ orientationin variance

مراجع:

1. [online], http://en.wikipedia.org/wiki/Features_%28pattern_recognition%29.
2. [online], http://en.wikipedia.org/wiki/Feature_detection_%28computer_vision%29.
3. [online], http://en.wikipedia.org/wiki/Image_moment
4. [online], http://en.wikipedia.org/wiki/Mel-frequency_cepstrum.
5. Combination Of Local Descriptors and Global Features For Leaf Recognition, MaliheShabanzade, MortezaZahedi and Seyyed Amin Aghvam, Shahrood university of Technology, Shahrood, Iran.
6. Exploiting Voronoi diagram properties in face segmentation and feature extraction, Abbas Cheddad, DzulkifliMohamad, AzizahAbdManaf, Northern Ireland- Johar, Kuala lumpur,Malaysia.
7. Face Recognition from a single Image per Person: A Survey, Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, Fuyan Zhang, China.
8. Face Recognition using SIFT features, Mohamed Aly.
9. Facial Expression Recognition with Local Binary Patterns and Linear Programming, X.Feng, M.Pietikainen, A.Hadid, Finland, China.
10. Fingerprint Verification Based on Invariant Moment Features and Nonlinear BPNN, Ju Cheng Yang and Dong Sun Park.
11. Optical Character Recognition for Hand-written Characters, GiorgosVamvakas, Computational Intelligence laboratory(CIL).
12. Pattern Recognition, Sharif University Lecture Notes.
13. Recognizing degraded faces: Contribution of internal and external features, IzzatN.Jarudi and PawanSinha, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology Cambridge, MA 02139.
14. Robust Appearance-based Sign Language Recognition, MortezaZahedi, PHD thesis.
15. Speech emotion recognition using hidden Markov models, Tin Lay Nwe, Say Wei Foo, LiyanageC.De Silva, Department of Electrical and Computer Engineering, National university of Singapore.

16. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques, LindasalwaMuda, MumtajBegam and I. Elamvazuthi.