
بسم الله الرحمن الرحیم

دانشگاه بیرجند

دانشکده مهندسی برق و کامپیوتر

پایان‌نامه دوره کارشناسی ارشد مهندسی برق - الکترونیک

بازشناسی بلادرنگ مستقل از اجراکننده زبان اشاره فارسی

علی اصغر زارع

استاد راهنما:

دکتر سید حمید ظهیری

تابستان ۱۳۹۳

تأییدیه هیات داوران

(برای پایان‌نامه)

یک نسخه اصل فرم مربوطه

تقدیم به روح پاک پدرم:

که عالمانه به من آموخت تا چگونه در عرصه زندگی، ایستادگی را تجربه نمایم.

تقدیم به مادرم:

دریای بی‌کران فداکاری و عشق که وجودم برایش همه رنج است و وجودش برایم همه مهر.

تقدیم به خواهران مهربانم:

که وجودشان شادی بخش و صفایشان مایه آرامش من است.

تقدیم به برادران عزیزم:

که همواره در طول تحصیل متحمل زحماتم بودند و تکیه گاه من در مواجهه با مشکلات، و وجودشان مایه دلگرمی من می‌باشد.

و تقدیم به:

همه کسانی که لحظه‌ای بعد انسانی و وجدانی خود را فراموش نمی‌کنند و بر آستان گران سنگ انسانیت سر فرود می‌آورند و انسان را با همه تفاوت‌هایش ارج می‌نهند.

تشکر و قدردانی

با عنایت خداوند متعال، پایان‌نامه فوق به پایان رسید که در بطن آن اعتقاد، هم فکری و همیاری بسیاری از عزیزان در بهبود بخش کوچکی از علم و دانش بشری و خدمت به میهن عزیزمان نهفته است. خط‌مشی این کار، بر اصول تعهد، تخصص و بازگویی نتایج بدون هیچ کم و کاستی بنا نهاده شده است. در این راستا تجارب اساتید ارجمند دانشگاه بیرجند بسیار موثر بوده‌اند. اکنون بر خود لازم میدانم از تک تک سروران و بزرگواران گرامی جناب دکتر ظهیری، جناب دکتر مهرشاد و جناب دکتر رضوی ، که در طی دو سال گذشته با راهنمایی‌ها و کمک‌هایشان راه را برای رسیدن به این نقطه هموار ساختند، با کمال سپاس قدردانی نمایم و توفیق ایزد منان را بر یکایک این عزیزان خواستارم.

علی‌اصغر زارع
دانشگاه بیرجند
تابستان ۱۳۹۳

چکیده

بینایی کامپیوتر بر پایه بازشناسی ژست، کاربردهای بالقوه‌ای در حوزه تعامل انسان-کامپیوتر مانند بازشناسی زبان اشاره دارد. بازشناسی خودکار زبان اشاره فارسی به دلیل عامل‌هایی مانند تعداد وسیع ژست‌های مشابه هم، جهت‌گیری دست‌ها، پس زمینه‌ی پیچیده و تغییرات روشنایی نور محیط، یکی از پر چالش‌ترین حوزه‌های تحقیقاتی می‌باشد. در این پایان‌نامه دو نوع سیستم متفاوت بازشناسی خودکار زبان اشاره فارسی بر پایه بینایی کامپیوتر با اهدافی چون پردازش بلادرنگ ژست‌ها، مستقل‌سازی سیستم بازشناسی در برابر افراد مختلف و همچنین عدم استفاده اجـراکننده ژست‌ها از دستکش یا مارکر پیشنهاد شده است. در سیستم پیشنهادی اول، پس از تقطیع دست‌ها از فریم‌های ویدویی، با تصحیح زاویه جهت گیری دست و برش زدن ناحیه مچ دست از مابقی ناحیه دست، مرز ناحیه دست استخراج می‌گردد و در جهت بازنمایی آن، محاسبه تابع زاویه‌ای تجمعی صورت می‌گیرد و در ادامه با محاسبه دامنه ضرایب فوریه گسسته و یک سری عملیات ریاضی در جهت مقاوم سازی آنها نسبت به انتقال، چرخش و تغییر مقیاس، ویژگی‌های مطلوب در حوزه‌ی فرکانس استخراج می‌گردد و این ویژگی‌های استخراجی جهت بازشناسی ژست‌ها به ورودی‌های یک شبکه عصبی پرسپترون چند لایه پیشرو اعمال می‌گردد اما در سیستم پیشنهادی دوم تمام عملیات پیش‌پردازش تا مرحله استخراج ناحیه کف دست و انگشتان مشابه سیستم اول انجام می‌پذیرد ولی پس از این مرحله، محاسبه ثابت‌های گشتاور در جهت بازنمایی ناحیه دست صورت می‌گیرد که این ثابت‌های گشتاور تغییرناپذیر مشابه با سیستم اول جهت بازشناسی ژست‌ها به ورودی‌های یک شبکه عصبی چند لایه اعمال می‌گردد البته طبقه‌بندی داده‌ها توسط سه طبقه‌بند بیز، K-NN و شبکه عصبی در جهت مقایسه عملکرد طبقه‌بندهای مختلف انجام شده است و همچنین این دو سیستم پیشنهادی با یکدیگر مقایسه شده‌اند. مجموعه آموزشی ژست‌ها از ۲۵۰ نمونه به ازای ۱۰ ژست در پنج موقعیت وجهت‌گیری متفاوت به وسیله پنج نفر بدست آمده است و نتایج بازشناسی این سیستم‌ها، نرخ بازشناسی ۱۰۰٪ را برای سیستم اول و ۹۵٪ را برای سیستم دوم نشان می‌دهد.

کلید واژه‌ها: الگوریتم برش مچ دست، تابع زاویه‌ای تجمعی، تبدیل فوریه گسسته، ثابت‌های گشتاور، شبکه عصبی

فهرست مطالب

عنوان **صفحه**

فهرست جدول‌ها

عنوان **صفحه**

فهرست شکل‌ها

عنوان **صفحه**

فصل ۱– مقدمه

با حضور پردازشگرهای دیجیتال در زندگی روزمره، انسان‌ها تمایل دارند تا بتوانند فرامین خود را هر چه ساده‌تر به کامپیوتر ها تفهیم کنند. در این راستا برقراری ارتباط بین انسان و کامپیوترها محدود است اکثرشیوه‌های ارتباطی رایج شامل یک وسیله واسط سخت افزاری و نرم افزاری بدون ادراک مثل صفحه کلید، موس، قلم نوری و ... می‌باشد که انتقال معنـا و مفهوم از طریق آنها در سطـوح پایینی هدایـت می‌شود. با ظهور کامپیوترهای پیشرفته‌تر و تجهیز شده به دوربین‌ها و همچنین پیشرفت الگوریتم‌های نرم‌افزاری در حوزه‌های یادگیری ماشین[۱]، بازشناسی الگو[۲]،پردازش سیگنال و تصویر، ماشین‌ها قادر به دیدن در فضای ماشینی خود شدند و با کمک الگوریتم‌های هوش مصنوعی[۳]، الهامی از قدرت ادراک خالق خود گرفته اند از این رو این حوزه دانش، تعامل انسان-کامپیوتر یا انسان- ماشین[۴] نام گرفت.

زبان اشاره یک شیوه ارتباطی اولیه برای افراد کر و لال است که یک فرم بصری از ارتباطات شامل ترکیبی از شکل، جهت‌گیری دست‌ها، حرکات دست‌ها، بازوها یا بدن و حالات چهره به جای صدا در زبان گفتاری است. با توجه به این حقیقت که بیش از ۳۶۰ میلیون نفر از مردم جهان بر طبق آمار سازمان سلامت جهانی ازنقصان شنوایی رنج می برند نیاز به طراحی سیـــــستم های مترجم اتوماتیک به عنوان واسط بین افراد کر ولال و افراد عادی به شدت احساس می‌شود این دستـه از افراد با استـفاده از این سیستم ها قادرند افکار خود را به دیگران انتقال دهنـــد و همچنین از طریق کامپیوتر ها قادر به تولید اصوات برآمده از ذهن خود، آن آرزوی رویایی‌شان هر چند مصنوعی خواهند شد. هدف سیستم‌های بازشناسی خودکار زبان اشاره، توسعه الگوریتم‌ها و روش‌های برای تشخیص صحیح یک توالی از ژست های تولید شده و فهم معنای آنها است. روش پیشنهادی ما در این پایان‌نامه بر روی بازشناسی زبان اشاره ایستا فارسی به صورت بلادرنگ بر روی مجموعه محدودی از ژست‌ها شامل ارقام و کلمات تمرکز دارد که به صورت زیر تدوین شده است:که در فصل ۲ به طور مختصر کارهای مرتبط را تشریح کرده‌ایم. در فصل۳ شیوه تقطیع دست پیشنهادی، الگوریتم‌های میانی چون تصحیح زاویه جهت‌گیری، برش مچ دست به همراه شیوه محاسبه ثابت‌های گشتاور به عنوان ویژگی های جهت بازنمایی ناحیه دست و نحوه طبقه بندی الگوها از طریق شبکه عصبی تشریح شده است در فصل ۴ نتایج آزمایش‌ها و نهایتاً نتیجه‌گیری و کارهای آینده در فصل۵، پایان‌بخش پایان‌نامه خواهد بود.

[1] Machine Learning
[2] Pattern Recognition
[3] Artificial Intelligence Algorithms
[4] Human-Machine Interaction

فصل ۲- کارهای مرتبط

به طور کلی بینایی کامپیوتر بر پایه بازشناسی علایم و همچنین سیستم‌های بازشناسی حرکات انسان، مجموعه‌ای متنوع از ویژگی‌های شامل توصیف‌گرهای شکل هندسی، پیکربندی انگشتان و توصیف‌گرهای مبتنی بر ظاهر را برای بازنمایی علایم به کار می‌برند که این ویژگی‌ها جهت بازشناسی علایم به کار برده می‌شود. برخی از سیستم‌های بازشناسی دست ایستا، تکنیک‌های بازشناسی الگوی، مانند تطبیق الگو و شبکه‌های عصبی را به کار می‌برند ودیگر سیستم‌ها، ژست‌های پیوسته را به وسیله شبکه‌های عصبی موقتی[5] و مدل‌های مخفی مارکوف مورد بازشناسی قرار می‌دهند.
در این فصل، مقالات به دو دسته تقسیم‌بندی شده است که اولین بخش آن شامل سیستم‌های است که ژست‌های دست ایستا را تشخیص می‌دهند در حالی که بخش دوم، سیستم‌های را پوشش می‌دهد که ژست‌های پیوسته را بازشناسی می‌کند.

۲-۱- بازشناسی ژست ایستا

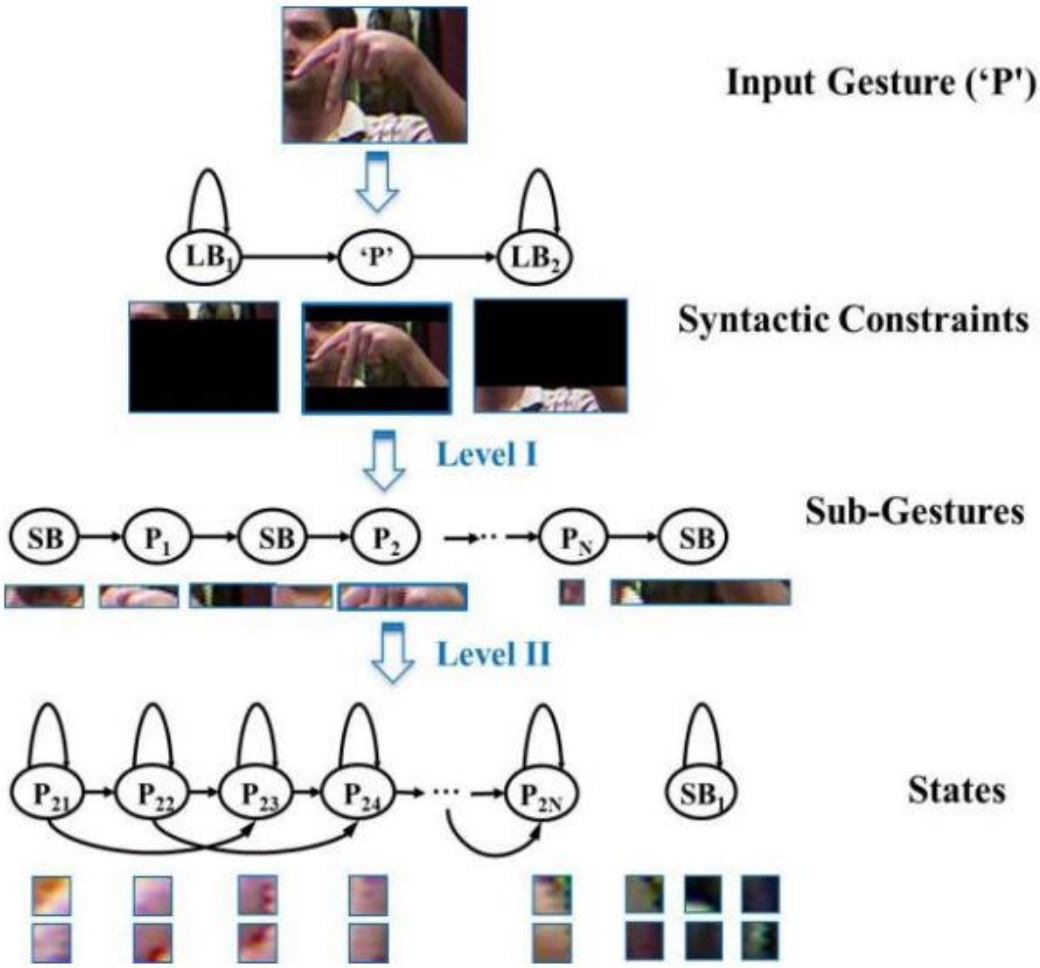
از آنجا که بیشتر تمرکز بازشناسی ژست‌های ایستا یا همان بازشناسی هجی انگشتان[6]، به پیکربندی و حالت قرارگیری انگشتان در تشکیل یک ژست معطوف است بنابراین در این راستا جهت استخراج ویژگی‌های قوی، به توصیف‌گرهایی که به خوبی بتوانند شکل هندسی یک شیء را در تصویر توصیف کنند، نیاز خواهد بود.

۲-۱-۱- بازشناسی الفبای اشاره با استفاده از مدل مخفی مارکوف دو سطحی

در این پروژه Lu و همکارانش یک ساختار جدید که از یک مدل مخفی دو سطحی بهره برده است را پیشنهاد داده‌اند[۱]. این سیستم قادر به بازشناسی هر ژست به صورت یک توالی از زیر واحدهای ژست است. علاوه بر این فرآیند تقطیع زیر واحدها و بازشناسی به صورت یک‌جا صورت می‌پذیرد. در این پروژه از ویژگی‌هایی بر پایه هیستوگرام گرادیان جهت‌دار استفاده شده است. همانطور که در شکل ۲-۱ نشان داده شده است این سیستم برپایه مدل مخفی مارکوف پیکربندی شده است که از یک شیوه آموزش بدون نظارت، برای برچسب‌گذاری کل تصاویر استفاده می‌کند. یک سا ختار مدل مخفی[7] مارکوف مرسوم، شامل سه مرحله می‌باشد. ۱- استخراج ویژگی ۲- تخمین پارامتر با یک فرآیند یادگیری تکراری ۳- بازشناسی برپایه رمزگشایی ویتربی[8]. در معماری نشان داده شده، در هر تصویر به نواحی مستطیلی پیکسل‌های N×N تقطیع می‌شود و یک پنجره تحلیل روی هم‌افتادگی M×M پیکسل برای محاسبه

[5] Temporal Neural Networks
[6] Fingerspelling Recognition
[7] Hidden Markov Model
[8] Viterbi

ویژگی‌ها محاسبه می‌شود. برای تسهیل پردازش بلادرنگ، تصاویر از چپ به راست و از بالا به پایین اسکن می‌شوند چناچه در شکل ۲-۲ نشان داده شده است.



شکل ۲-۱: ساختار شبکه HMM جهت بازشناسی الفبای اشاره [۱].

اندازه تصاویر در الفبای اشاره امریکایی از ۹۰×۶۰ پیکسل تا ۱۳۰×۱۷۰ پیکسل با اکثریت تصاویری در اندازه ۱۰۰×۸۰ متغیر است. طبق یک سری آزمایش‌ها بر روی مجموعه‌ایی از تصاویر به منظور بهینه سازی اندازه‌های هر پنجره و فریم، بهترین نتایج با ترکیباتی از اندازه فریم ۵ پیکسل و اندازه پنجره ۳۰ پیکسل بدست آمده است. تعداد بهینه وضعیت‌ها و توپولوژی مدل به این پارامتر وابسته است. ترکیب ۵/۳۰ یک مصالحه خوب را بین پیچیدگی محاسباتی و پیچیدگی مدل و عملکرد بازشناسی ارائه می‌دهد. در این پروژه ویژگی‌های هیستوگرام گرادیان جهت‌دار[9] به دلیل شهرت آن در کاربردهای پردازش تصویر به کار گرفته شده است. برای محاسبه ویژگی‌های HOG، در ابتدا دامنه گرادیان تصویر g و زاویه θ برای هر پیکسل با به کاربردن یک فیلتر یک بعدی به صورت زیر محاسبه می‌گردد.

$$g(u,v) = \sqrt{g_x(u,v)^2 + g_y(u,v)^2} \qquad (۱-۲)$$

[9] Histogram of Oriented Gradient (HOG)

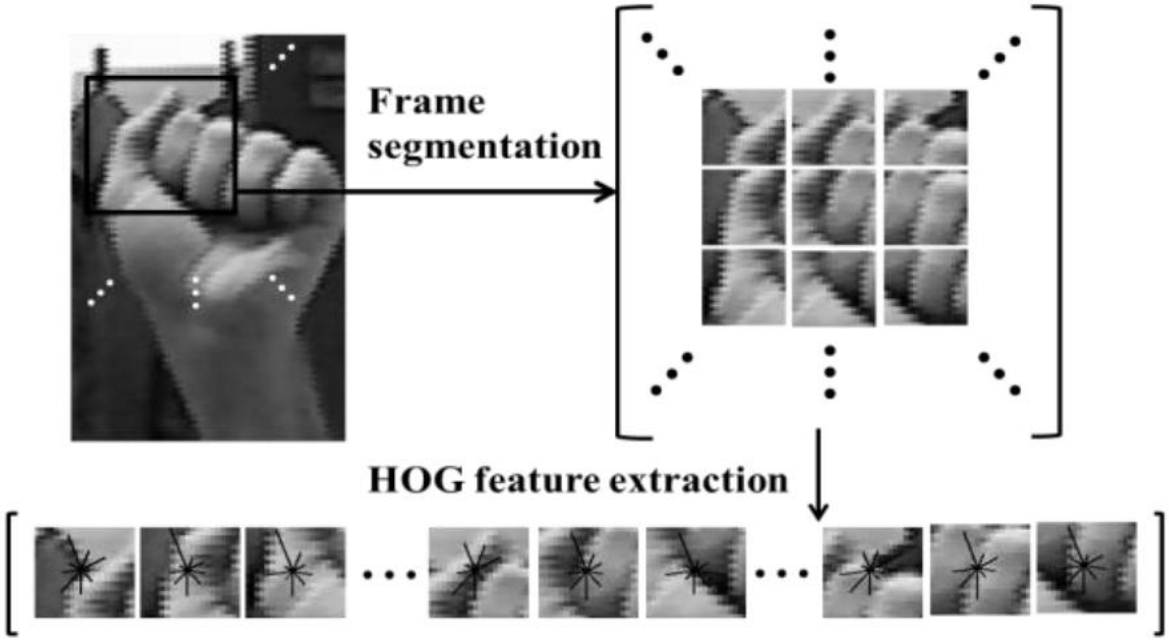
$$\theta(u,v) = arctg\,\frac{g_{y(u,v)}}{g_{x(u,v)}} \qquad (۲-۲)$$

برای هر فریم، یک بردار ویژگی استخراج شده است که f_i به وسیله روقومی کردن جهت علامت‌گذاری شده درون N مجموعه جهت‌دار وزن‌دهی شده به وسیله دامنه گرادیان به صورت زیر تعریف می‌شوند.

$$f_i = [f_i(n)]^T_{n\in[1,2,...N]} \qquad (۳-۲)$$

$$f_i(n) = \sum_{(u,v)\in f_i} g(u,v)\,\delta\left[bin(u,v) - n\right] \qquad (۴-۲)$$

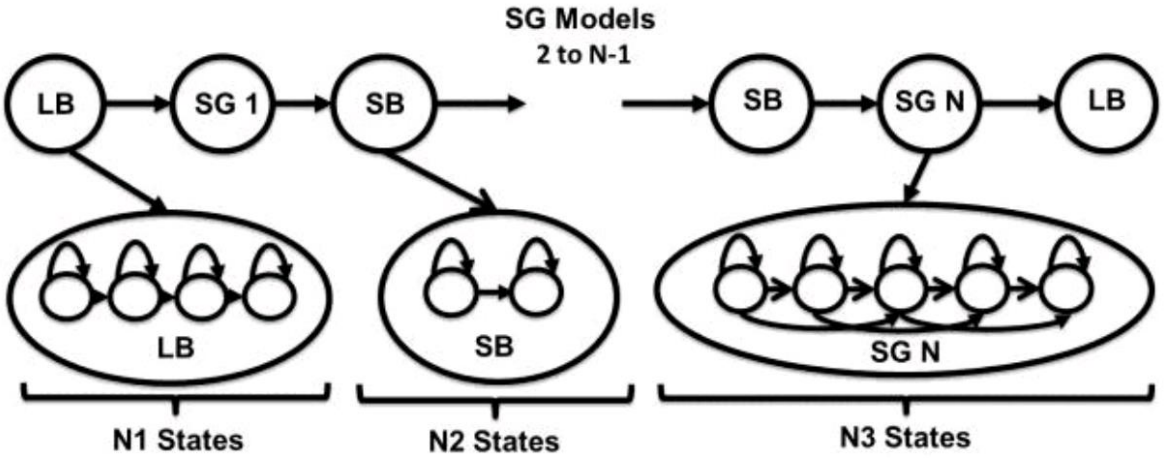
تابع bin(u,v) اندیس مجموعه مرتبط با پیکسل (u,v) را بر می‌گرداند. سپس این ویژگی‌ها به عنوان ورودی به ساختار HMM دو سطحی اعمال می‌شود سطح اول این مدل مخفی مارکوف، قیود نحوی را در این سیستم اعمال می‌کند. این سطح به دو منظور استفاده شده است. اول اینکه، این سطح سیستم را به خروجی یک فرض در هر تصویر محدود می‌کند یعنی یک علامت در هر تصویر شامل خواهد شد. تمام ۲۴ علامت به وسیله گره مرکزی در سطح بالای HMM می‌توانند خروجی باشند که در شکل ۲-۱ به قیود نحوی اشاره دارد. این قیود از خطاهای الحاق[۱۰]، به وسیله ممانعت از فرضیات چندگانه در هر تصویر جلوگیری می‌کند. دوم اینکه هر فریم تصویر را مجبور به طبقه بندی شدن به دو بدنه پس‌زمینه و علامت می‌کند که ترجیح داده شده است که به دو صورت پس‌زمینه بزرگ و ژست، این طبقه‌بندی صورت گیرد.



شکل ۲-۲: تحلیل فریمی انجام شده برای استخراج ویژگی‌های ۹ بعدی HOG [۱].

[10] Insertion

پس زمینه بزرگ(LB) می‌تواند در ابتدا و یا به دنبال یک فرض از یک علامت الفبا بیاید. تصاویر به صورت یک تعداد اختیاری از فریم‌های طبقه‌بندی شده به عنوان پس‌زمینه که به دنبال آن علامت الفبا آمده باشد و همچنین پس از آن مجددا به صورت یک تعداد اختیاری از فریم‌های پس‌زمینه مدل می-شوند.از این رو، سطح اول مدل مخفی مارکوف تقطیع ضمختی از تصویر را به صورت بخشی از فرآیند بازشناسی را پیاده‌سازی می‌کند. سطح دوم، زیر ژست‌ها را برچسب‌گذاری می‌کند. بر طبق شکل ۲-۱، هر علامت الفبا به صورت یک مدل HMM چپ به راست که مابین یک مدل پس‌زمینه کوتاه(SB) دنبال شده با یک مدل متناظر با یک زیر ژست که جایگزین می‌کند مدل می‌شود. این سطح هر علامت را به صورت یک توالی از زیر ژست‌ها(SG) مدل می‌کند. تعداد بهینه مدل‌های SG برای هر علامت در ادامه آمده است. مدل SB به بخش‌های از تصویر که مابین ژست‌ها قرار گرفته‌اند اجازه می‌دهد تا به صورت پس زمینه مدل شوند. تابع این سطح، هر علامت را به توالی از واحدی از زیر ژست‌ها تجزیه می‌کند. آموزش این مدل‌ها به یک روش بدون نظارت انجام گرفته شده است. هر ژست به صورت یک HMM چپ به راست معمولی با وضعیت‌های پرش پیاده‌سازی شده است. این مدل با عنوان مدل Bakies شناخته شده می‌باشد. بر خلاف بسیاری از سیستم‌های که ژست‌ها را به صورت کلی بازشناسی می‌کنند، در اینجا هر علامت الفبا را با یک توالی که به طور نوعی شامل LB و SB و یک توالی از مدل‌های SG است به صورت شکل ۲-۳ مدل شده‌اند. دلیل اینکه دو مدل پس‌زمینه مورد استفاده واقع شده است این



شکل ۲-۳: ساختار دو سطحی که یک تصویر را به صورت ترکیباتی از مدل‌های LB، SB و SG مدل می‌کند[۱] .

است که SB بیشتر روی نواحی پس‌زمینه کوچک همراه با انگشتان و مرزهای تصویر متمرکز شده است در حالی که مدل LB برای مدل‌سازی نواحی پس‌زمینه پیشین و پسین یک علامت در تصویر به کار برده شده است. مدل SB یک مدل HMM تک وضعیتی با انتقال به خود می‌باشد. مدل LB یک مدل ۱۱ وضعیتی است و همچنین به هر وضعیت اجازه انتقال به خود را می‌دهد. از این رو، یک فرآیند آموزش بدون نظارت پیاده‌سازی شده است که مدل‌های LB، SB و SG بدون نیاز به هر نوع رونوشت دستی[۱۱]، علامت ارائه شده به وسیله تصویر را شناسایی می‌کند. هر نمونه یک رونوشت کلی را به صورت شروع با

[11] Transcription

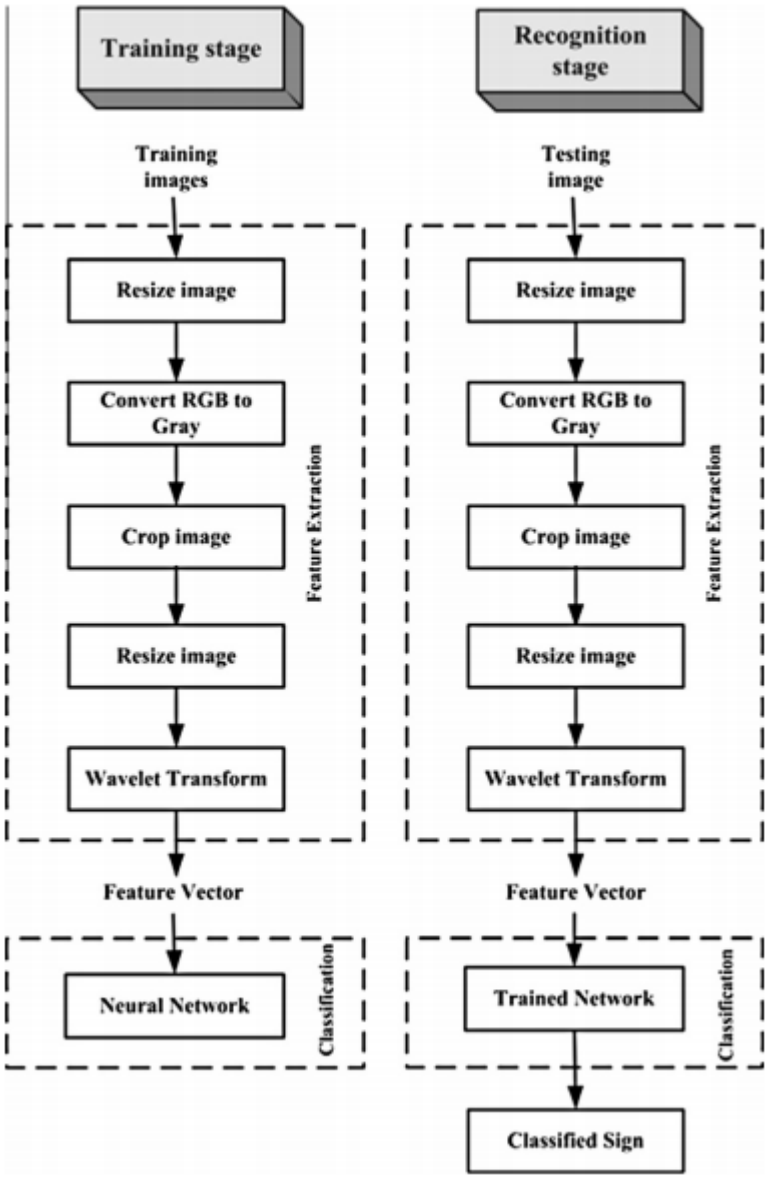
LB,SG1,SB,SG2,…SGN و پایان با یک LB به کار می‌برد. در اینجا Start LB به یک مدل LB اشاره دارد که بایستی در ابتدای هر فایل تصویر روی دهد. اتصال Start LB و End LB به یکدیگر بدین معناست که از یک مدل HMM مشابه بهره گرفته شده است. در این پروژه از الگوریتم بیشینه سازی امید ریاضی EM و به طور خاص‌تر، الگوریتم BW برای آموزش مدل مخفی مارکوف استفاده شده است لازم به ذکر است که تمام مدل‌های HMM به طور همزمان در هر تکرار به روز می‌شوند.

۲-۱-۲- بازشناسی زبان اشاره فارسی با استفاده از تبدیل موجک و شبکه‌های عصبی

زبان اشاره فارسی، زبان کاملی است که علامت‌های ساخته شده به وسیله دست‌ها و دیگر ژست‌ها شامل حالات چهره و حالات قرارگیری بدن را به کار می‌گیرد. PSL شامل ۱۰۷۵ ژست الفبای معمولی و کلماتی است که در بردارنده ژست‌های پویا و ایستا است. PSL دارای چهار المان کلی است. ۱- وضعیت دست‌ها ۲- حرکت دست‌ها ۳- مکانی که دست‌ها در آنجا می‌ایستند ۴- جهت کف دست. زبان اشاره فارسی برای ارائه ژست‌ها تنها از یک دست استفاده می‌کند که از ۳۵ ژست ایستا و ۲ ژست پویا برای ارائه ۳۷ حرف الفبای فارسی به صورت شکل ۲-۴ استفاده می‌کند.

شکل ۲-۴: الفبای اشاره فارسی [۲].

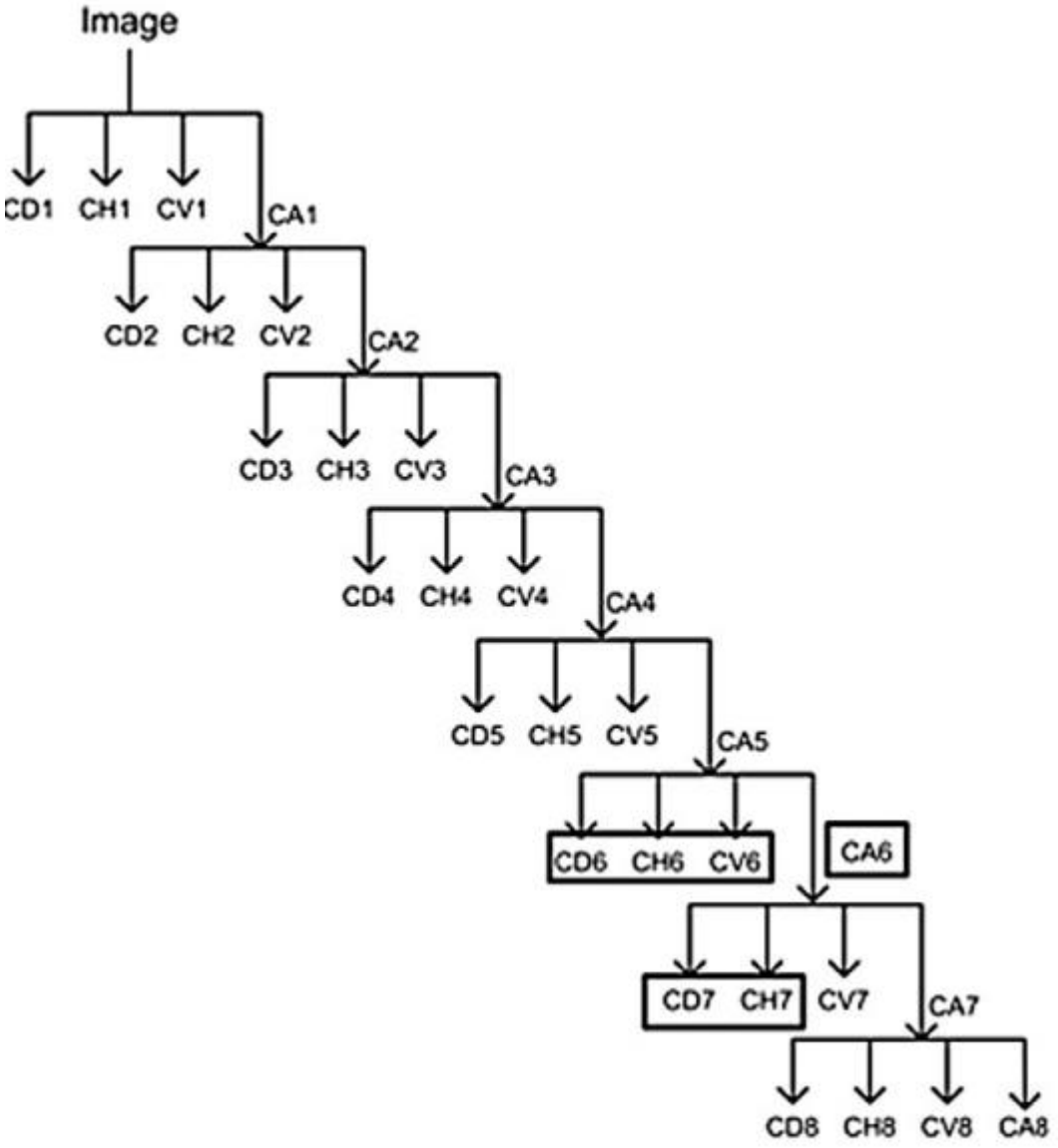
کریمی و همکارانش یک سیستم بازشناسی الفبای اشاره فارسی مبتنی بر تبدیل موجک و شبکه عصبی ارائه کرده‌اند[۲]. این سیستم شامل دو فاز استخراج ویژگی و طبقه‌بندی می‌باشد. بلوک دیاگرام این سیستم در شکل ۲-۵ نشان داده شده است.



شکل ۲-۵: بلوک دیاگرام سیستم مبتنی بر تبدیل موجک [۲].

در مرحله استخراج ویژگی، تصاویر رنگی به ۲۸۰×۳۵۰ پیکسل تغییر اندازه داده می‌شوند و همچنین این تصاویر RGB به سطح خاکستری تبدیل می‌شوند سپس در این تصاویر که فقط ناحیه دست را در خود دارند با یک عمل میانگین‌گیری، تنها ناحیه بالقوه دست استخراج می‌شود. پس از این مرحله تصاویر دومرتبه به تصاویر ۲۰۰×۳۰۰ پیکسل تغییر اندازه داده می‌شوند. در این سیستم جهت استخراج ویژگی از هر تصویر به دلیل سادگی و موثر بودن تبدیل موجک هار در توصیف بخش‌های نرم بدن انسان از این

نوع تبدیل استفاده شده است. به منظور استخراج تقریبات[۱۲] و جزئیات[۱۳] برای هر تصویر، ۹ سطح تجزیه[۱۴] به کار رفته است. با این حال ممکن است هر المان از تصویر اصلی و یا زیر تصاویر به عنوان ورودی‌های شبکه عصبی انتخاب گردند اما تقریبا غیرممکن است که تمام ۳۰۰×۲۰۰ المان یک تصویر (۶۰۰۰) به عنوان ورودی‌های شبکه عصبی انتخاب گردند. برای حل این مشکل، آنها ماتریس ضرایب بدست آمده از سطح ۶ ام و ۷ ام و سطوح بالاتر زیر تصویر را برای استخراج ورودی‌های شبکه عصبی به کار برده‌اند. با این عمل تعداد ورودی‌های شبکه عصبی کوچک‌تـر می‌شود. به عبـــارت دیگر برای انتخاب



شکل ۲-۶: درخت تجزیه تبدیل موجک [۲].

[12] Approximation
[13] Details
[14] Decomposition