

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه بیرجند
دانشکده مهندسی برق و کامپیوتر

پایان نامه 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 پایان نامه)

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 بی کران 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 می باشد.

و تقدیم به:

همه کسانی که لحظه ای بعد انسانی و وجدانی خود را فراموش نمی کنند و بر آستان گران سنگ انسانیت سر فرود می آورند و انسان را با همه تفاوت هایش ارج می نهند.

تشر 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.

علی اصغر 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، پس‌زمینه‌ی پیچیده و تغییرات روشنایی نور محیط، یکی از پرچالش‌ترین 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 اجراکننده ژست‌ها از دستکش یا مارکر پیشنهاد شده است. در سیستم پیشنهادی اول، پس از تقطیع دست‌ها از فریم‌های ویدئویی، با تصحیح زاویه جهت‌گیری 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 مقاوم‌سازی 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 و جهت‌گیری 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

صفحه	عنوان
ج	فهرست جدول‌ها.....
ه	فهرست شکل‌ها.....
۱	فصل ۱- مقدمه.....
۲	فصل ۲- کارهای مرتبط.....
۲	۱-۲-۱- بازشناسی ژست ایستا.....
۲	۱-۲-۱-۱- بازشناسی الفبای اشاره با استفاده از مدل مخفی مارکوف دو سطحی.....
۶	۱-۲-۱-۲- بازشناسی زبان اشاره فارسی با استفاده از تبدیل موجک و شبکه‌های عصبی.....
۱۰	۱-۲-۱-۳- رویکرد مبتنی بر بینایی کامپیوتر برای بازشناسی الفبای زبان اشاره هندی.....
۱۴	۱-۲-۱-۴- بازشناسی خودکار الفبای اشاره ترکی.....
۱۶	۱-۲-۱-۵- دیگر کارهای بازشناسی ژست ایستا.....
۱۷	۲-۲- بازشناسی ژست پیوسته.....
۱۷	۱-۲-۲- بازشناسی ژست با استفاده از شبکه‌های عصبی بازگشتی.....
۲۰	۲-۲-۲- سیستم بازشناسی ژست پیوسته‌ی بلادرنگ برای زبان اشاره.....
۲۶	فصل ۳- سیستم بازشناسی بلادرنگ و مستقل از اجرا کننده زبان اشاره فارسی.....
	۱-۳-۱- سیستم بازشناسی بلادرنگ ژست ایستای زبان اشاره فارسی مستقل از اجرا کننده بر پایه
۲۸	ضرایب فوریه با استفاده از شبکه‌های عصبی.....
۲۸	۱-۳-۱- مجموعه نمونه‌ها.....
۲۹	۱-۳-۲- انتخاب فریم کلیدی.....
۳۰	۱-۳-۳- بهبود میزان تباین.....
۳۱	۱-۳-۴- حذف نویز و فیلترگذاری.....
۳۲	۱-۳-۵- تقطیع پوست.....
۳۶	۱-۳-۶- تشخیص دست.....
۳۸	۱-۳-۷- تصحیح زاویه جهت‌گیری دست.....
۳۹	۱-۳-۸- الگوریتم برش مچ دست.....
۴۱	۱-۳-۹- استخراج مرز دست.....
۴۲	۱-۳-۱۰- استخراج ویژگی‌های دامنه ضرایب فوریه.....
۴۸	۱-۳-۱۱- طبقه‌بندی.....
	۲-۳- سیستم بازشناسی بلادرنگ ژست ایستای زبان اشاره فارسی مستقل از اجرا کننده بر پایه
۵۰	ثابت‌های گشتاور با استفاده از شبکه‌های عصبی.....
۵۰	۱-۲-۳- استخراج ویژگی‌های ثابت‌های گشتاور.....

۵۴ طبقه بندی	۲-۲-۳
۵۷ آزمایش ها و نتایج شبیه سازی	فصل ۴-۴
۵۹ سیستم مبتنی بر دامنه ضرایب فوریه	۱-۴
۶۴ سیستم مبتنی بر ثابت های گشتاور	۲-۴
 سیستم بازشناسی مبتنی بر دامنه ضرایب فوریه در مقابل سیستم بازشناسی مبتنی بر ثابت های گشتاور	۳-۴
۶۹	
۷۹ نتیجه گیری و کارهای آینده	فصل ۵-۵
۸۱	فهرست مراجع
۸۳	واژه نامه فارسی به انگلیسی
۸۵	واژه نامه انگلیسی به فارسی

فهرست جدول‌ها

عنوان	صفحه
جدول ۱-۲: نتایج دقت بازشناسی بر روی ۸ ژست الفبای اشاره هندی [۳].....	۱۴
جدول ۱-۳: نحوه جمع آوری هر ژست در فواصل و جهت گیری متفاوت از دوربین.....	۲۹
جدول ۱-۴: نتایج معیارهای بازشناسی بر روی ۴۰٪ داده های آموزشی و ۶۰٪ داده های آزمایشی.....	۵۹
جدول ۲-۴: متوسط دقت تشخیص و قابلیت اطمینان برای ۴۰٪ داده های آموزشی و ۶۰٪ آزمایشی.....	۶۰
جدول ۳-۴: نتایج معیارهای بازشناسی بر روی ۵۵٪ داده های آموزشی و ۴۵٪ داده های آزمایشی.....	۶۱
جدول ۴-۴: متوسط دقت تشخیص و قابلیت اطمینان برای ۵۵٪ داده های آموزشی و ۴۵٪ آزمایشی.....	۶۱
جدول ۵-۴: نتایج معیارهای بازشناسی بر روی ۹۰٪ داده های آموزشی و ۱۰٪ داده های آزمایشی.....	۶۲
جدول ۶-۴: متوسط دقت تشخیص و قابلیت اطمینان برای ۹۰٪ داده های آموزشی و ۱۰٪ آزمایشی.....	۶۳
جدول ۷-۴: نتایج معیارهای بازشناسی بر روی ۴۰٪ داده های آموزشی و ۶۰٪ داده های آزمایشی.....	۶۴
جدول ۸-۴: متوسط دقت تشخیص و قابلیت اطمینان برای ۴۰٪ داده های آموزشی و ۶۰٪ آزمایشی.....	۶۵
جدول ۹-۴: نتایج معیارهای بازشناسی بر روی ۵۵٪ داده های آموزشی و ۴۵٪ داده های آزمایشی.....	۶۶
جدول ۱۰-۴: متوسط دقت تشخیص و قابلیت اطمینان برای ۵۵٪ داده های آموزشی و ۴۵٪ آزمایشی.....	۶۶
جدول ۱۱-۴: نتایج معیارهای بازشناسی بر روی ۹۰٪ داده های آموزشی و ۱۰٪ داده های آزمایشی.....	۶۷
جدول ۱۲-۴: متوسط دقت تشخیص و قابلیت اطمینان برای ۹۰٪ داده های آموزشی و ۱۰٪ آزمایشی.....	۶۸
جدول ۱۳-۴: نتایج مقایسه ای SYS1 در مقابل SYS2 به ازای ۴۰٪ داده های آموزش و ۶۰٪ داده های آزمایشی.....	۶۹
جدول ۱۴-۴: ماتریس آسیمگی شبکه عصبی سیستم اول به ازای ۱۰۰ نمونه آموزشی.....	۷۰
جدول ۱۵-۴: ماتریس آسیمگی شبکه عصبی سیستم دوم به ازای ۱۰۰ نمونه آموزشی.....	۷۰
جدول ۱۶-۴: ماتریس آسیمگی طبقه بند بیز سیستم اول به ازای ۱۰۰ نمونه آموزشی.....	۷۱
جدول ۱۷-۴: ماتریس آسیمگی طبقه بند بیز سیستم دوم به ازای ۱۰۰ نمونه آموزشی.....	۷۱
جدول ۱۸-۴: نتایج مقایسه ای SYS1 در مقابل SYS2 به ازای ۵۵٪ داده های آموزش و ۴۵٪ داده های آزمایشی.....	۷۲
جدول ۱۹-۴: ماتریس آسیمگی شبکه عصبی سیستم اول به ازای ۱۳۰ نمونه آموزشی.....	۷۳
جدول ۲۰-۴: ماتریس آسیمگی شبکه عصبی سیستم دوم به ازای ۱۳۰ نمونه آموزشی.....	۷۳
جدول ۲۱-۴: ماتریس آسیمگی طبقه بند بیز سیستم اول به ازای ۱۳۰ نمونه آموزشی.....	۷۴
جدول ۲۲-۴: ماتریس آسیمگی طبقه بند بیز سیستم دوم به ازای ۱۳۰ نمونه آموزشی.....	۷۴
جدول ۲۳-۴: نتایج مقایسه ای SYS1 در مقابل SYS2 به ازای ۹۰٪ داده های آموزش و ۱۰٪ داده های آزمایشی.....	۷۵
جدول ۲۴-۴: ماتریس آسیمگی شبکه عصبی سیستم اول به ازای ۲۳۰ نمونه آموزشی.....	۷۶
جدول ۲۵-۴: ماتریس آسیمگی شبکه عصبی سیستم دوم به ازای ۲۳۰ نمونه آموزشی.....	۷۶

جدول ۴-۲۶: ماتریس آسیمگی طبقه بند بیز سیستم اول به ازای ۲۳۰ نمونه آموزشی ۷۷

جدول ۴-۲۷: ماتریس آسیمگی طبقه بند بیز سیستم دوم به ازای ۲۳۰ نمونه آموزشی ۷۷

فهرست شکل‌ها

صفحه	عنوان
۳	شکل ۱-۲: ساختار شبکه HMM جهت بازشناسی الفبای اشاره [۱].....
۴	شکل ۲-۲: تحلیل فریمی انجام شده برای استخراج ویژگی‌های ۹ بعدی HOG [۱].....
۵	شکل ۳-۲: ساختار دو سطحی که یک تصویر را به صورت ترکیباتی از مدل‌های LB، SB و SG مدل می‌کند [۱].....
۶	شکل ۴-۲: الفبای اشاره فارسی [۲].....
۷	شکل ۵-۲: بلوک دیاگرام سیستم مبتنی بر تبدیل موجک [۲].....
۸	شکل ۶-۲: درخت تجزیه تبدیل موجک [۲].....
۹	شکل ۷-۲: بازیابی زیر تصاویر برای علامت "ظ" [۲].....
۱۰	شکل ۸-۲: ساختار شبکه عصبی در مرحله طبقه بندی [۲].....
۱۰	شکل ۹-۲: الفبای اشاره زبان هندی [۳].....
۱۱	شکل ۱۰-۲: بلوک دیاگرام سیستم بازشناسی الفبای اشاره هندی مبتنی بر مشخصه‌های هندسی [۳].....
۱۲	شکل ۱۱-۲: تصویر اصلی [۳].....
۱۲	شکل ۱۲-۲: تقطیع ناحیه دست در فضای HSI [۳].....
۱۲	شکل ۱۳-۲: تشخیص نقاط پایه و نوک انگشتان [۳].....
۱۳	شکل ۱۴-۲: ناحیه کف دست با حذف انگشتان [۳].....
۱۳	شکل ۱۵-۲: ناحیه انگشتان با حذف کف دست [۳].....
۱۳	شکل ۱۶-۲: تقسیم بندی هر انگشت به سه بخش جهت تشخیص میزان باز و نیمه بسته بودن [۳].....
۱۴	شکل ۱۷-۲: الفبای اشاره ترکی [۴].....
۱۵	شکل ۱۸-۲: سلول‌ها و بلوک‌ها در یک تصویر نمونه [۴].....
۱۶	شکل ۱۹-۲: بلوک‌های روی هم افتاده در تصویر گرادیان به همراه هیستوگرام گرادیان جهت دار نرمالیزه [۴].....
۱۸	شکل ۲۰-۲: پیکربندی شبکه عصبی [۱۳].....
۱۹	شکل ۲۱-۲: شبکه عصبی بازگشتی جهت بازشناسی کلمات [۱۳].....
۱۹	شکل ۲۲-۲: سیستم بازشناسی ژست پیوسته [۱۳].....
۲۰	شکل ۲۳-۲: کلمات زبان اشاره ژاپنی [۱۳].....
۲۱	شکل ۲۴-۲: پنجاه و یک ژست ایستای زبان تایوانی [۱۴].....
۲۲	شکل ۲۵-۲: موقعیت‌های نسبی بین ژست‌ها و بدن در زبان اشاره تایوانی [۱۴].....
۲۲	شکل ۲۶-۲: هشت نوع حرکت موجود در ژست‌های زبان اشاره تایوانی [۱۴].....
۲۳	شکل ۲۷-۲: تغییرات تعداد TVP‌های یک رشته ورودی از ژست در طول محور زمان [۱۴].....
۲۳	شکل ۲۸-۲: بلوک دیاگرام سیستم بازشناسی ژست پیوسته بر پایه ژست ایستا [۱۴].....

- شکل ۳-۱: ژست‌های انتخابی زبان اشاره فارسی ۲۷
- شکل ۳-۲: نمای دوربین با ابعاد 640×480 ۲۷
- شکل ۳-۳: فلوچارت سیستم بازشناسی پیشنهادی اول ۲۸
- شکل ۳-۴: شیوه انتخاب فریم کلیدی بر پایه تفریق فریمی ۳۰
- شکل ۳-۵: فیلتر پایین گذر گوسی متقارن با انحراف استاندارد $\sigma=6$ ۳۱
- شکل ۳-۶: تقطیع پوست در فضای رنگ RGB ۳۵
- شکل ۳-۷: تقطیع در فضای رنگ YCbCr ۳۶
- شکل ۳-۸: حاصل ترکیب دو فضای رنگ ۳۶
- شکل ۳-۹: نمایش مراکز جرم دو ناحیه اصلی تقطیع شده ۳۷
- شکل ۳-۱۰: نحوه تصحیح جهت‌گیری دست ۳۹
- شکل ۳-۱۱: نمایش ناحیه دست در ماتریس تصویر و خط برش ۴۰
- شکل ۳-۱۲: نمایش نحوه تصحیح زاویه و برش مچ دست ۴۱
- شکل ۳-۱۳: ناحیه کف دست و انگشتان ۴۲
- شکل ۳-۱۴: مرز ناحیه دست ۴۲
- شکل ۳-۱۵: نمایش رابطه تابع زاویه‌ای با تابع زاویه‌ای تجمعی ۴۳
- شکل ۳-۱۶: نمایش بازنمایی تابع زاویه‌ای مرز دست برای ژست "سه" ۴۴
- شکل ۳-۱۷: نمایش تابع زاویه‌ای تجمعی مرز دست برای ژست "سه" ۴۵
- شکل ۳-۱۸: نمایش تابع زاویه‌ای تجمعی نرمالیزه شده مرز دست برای ژست "سه" ۴۵
- شکل ۳-۱۹: مرز ناحیه دست برای ژست "سه" ۴۷
- شکل ۳-۲۰: دامنه توصیف‌گر فوریه برای ژست "سه" ۴۸
- شکل ۳-۲۱: ساختار شبکه عصبی پیشنهادی جهت بازشناسی ژست‌ها ۴۹
- شکل ۳-۲۲: ناحیه کف دست و انگشتان ۵۰
- شکل ۳-۲۳: ناحیه دست و ویژگی‌های استخراجی ۵۳
- شکل ۳-۲۴: ساختار شبکه عصبی چند لایه پیشنهادی ۵۴
- شکل ۳-۲۵: فلوچارت سیستم پیشنهادی مبتنی بر ثابت‌های گشتاور ۵۵
- شکل ۴-۱: رابط گرافیکی سیستم پیشنهادی ۷۸

فصل ۱ - مقدمه

با حضور پردازشگرهای دیجیتال در زندگی روزمره، انسان‌ها تمایل دارند تا بتوانند فرامین خود را هر چه ساده‌تر به کامپیوترها تفهیم کنند. در این راستا برقراری ارتباط بین انسان و کامپیوترها محدود است اکثر شیوه‌های ارتباطی رایج شامل یک وسیله واسط سخت افزاری و نرم افزاری بدون ادراک مثل صفحه کلید، موس، قلم نوری و ... می‌باشد که انتقال معنا و مفهوم از طریق آنها در سطوح پایینی هدایت 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 گرفته‌اند 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 از نقصان شنوایی رنج می‌برند نیاز به طراحی سیستم‌های مترجم اتوماتیک به عنوان واسط بین افراد کر و لال و افراد عادی به شدت احساس می‌شود این دسته از افراد با استفاده از این سیستم‌ها قادرند افکار خود را به دیگران انتقال دهند و همچنین از طریق کامپیوترها 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 ژست‌های 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 فصل ۳ شیوه تقطیع دست پیشنهادی، الگوریتم‌های میانی چون تصحیح زاویه جهت‌گیری، برش میچ دست به همراه شیوه محاسبه ثابت‌های گشتاور به عنوان ویژگی‌های جهت‌بازنمایی ناحیه دست و نحوه طبقه‌بندی الگوها از طریق شبکه عصبی تشریح شده است در فصل ۴ نتایج آزمایش‌ها و نهایتاً نتیجه‌گیری و کارهای آینده در فصل ۵، 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 علائم و همچنین سیستم‌های بازشناسی حرکات انسان، مجموعه‌ای متنوع از ویژگی‌های شامل توصیف‌گرهای شکل هندسی، پیکربندی انگشتان و توصیف‌گرهای مبتنی بر ظاهر را برای بازنمایی علائم به کار می‌برند که این ویژگی‌ها جهت بازشناسی علائم به کار برده می‌شود. برخی از سیستم‌های بازشناسی دست ایستا، تکنیک‌های بازشناسی الگوی، مانند تطبیق الگو و شبکه‌های عصبی را به کار می‌برند و دیگر سیستم‌ها، ژست‌های پیوسته را به وسیله شبکه‌های عصبی موقتی^۵ 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.

۲-۱-۱ - بازشناسی ژست ایستا

از آنجا که بیشتر تمرکز بازشناسی ژست‌های ایستا یا همان بازشناسی هجی انگشتان^۶، 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.

۲-۱-۱-۱ - 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 بر پایه مدل مخفی مارکوف پیکربندی شده است که از یک شیوه آموزش بدون نظارت، برای برچسب‌گذاری کل تصاویر استفاده می‌کند. یک ساختار مدل مخفی^۷ مارکوف مرسوم، شامل سه مرحله می‌باشد. ۱- استخراج ویژگی ۲- تخمین پارامتر با یک فرآیند یادگیری تکراری ۳- بازشناسی بر پایه رمزگشایی ویتربی^۸. در معماری نشان داده شده، در هر تصویر به نواحی مستطیلی پیکسل‌های $N \times N$ تقطیع می‌شود و یک پنجره تحلیل روی هم‌افتادگی $M \times M$ پیکسل برای محاسبه

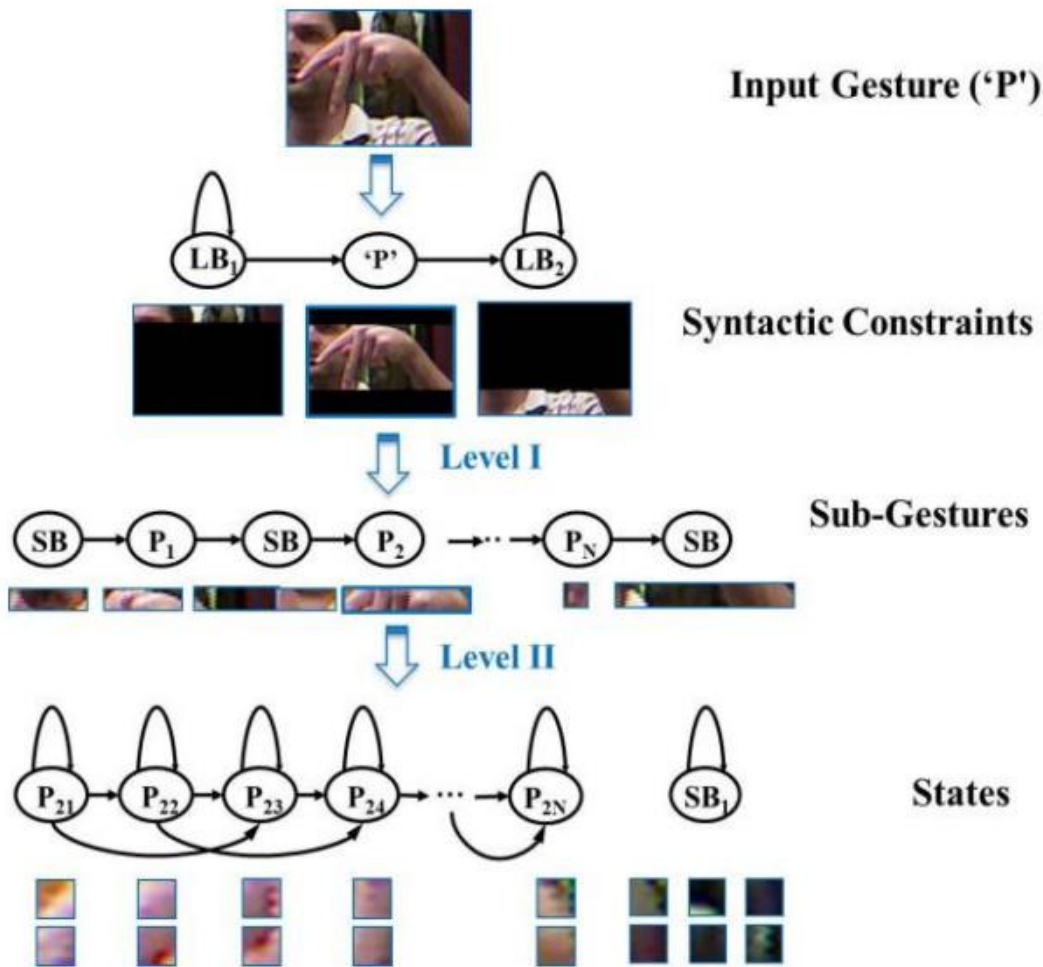
⁵ Temporal Neural Networks

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 چنانچه 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 ۶۰×۹۰ پیکسل تا ۱۷۰×۱۳۰ پیکسل با اکثریت تصاویری در اندازه ۸۰×۱۰۰ متغیر است. طبق یک سری آزمایش‌ها بر روی مجموعه‌ای 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^۹ 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 کار بردن یک فیلتر یک بعدی به صورت زیر محاسبه می‌گردد.

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

^۹ Histogram of Oriented Gradient (HOG)

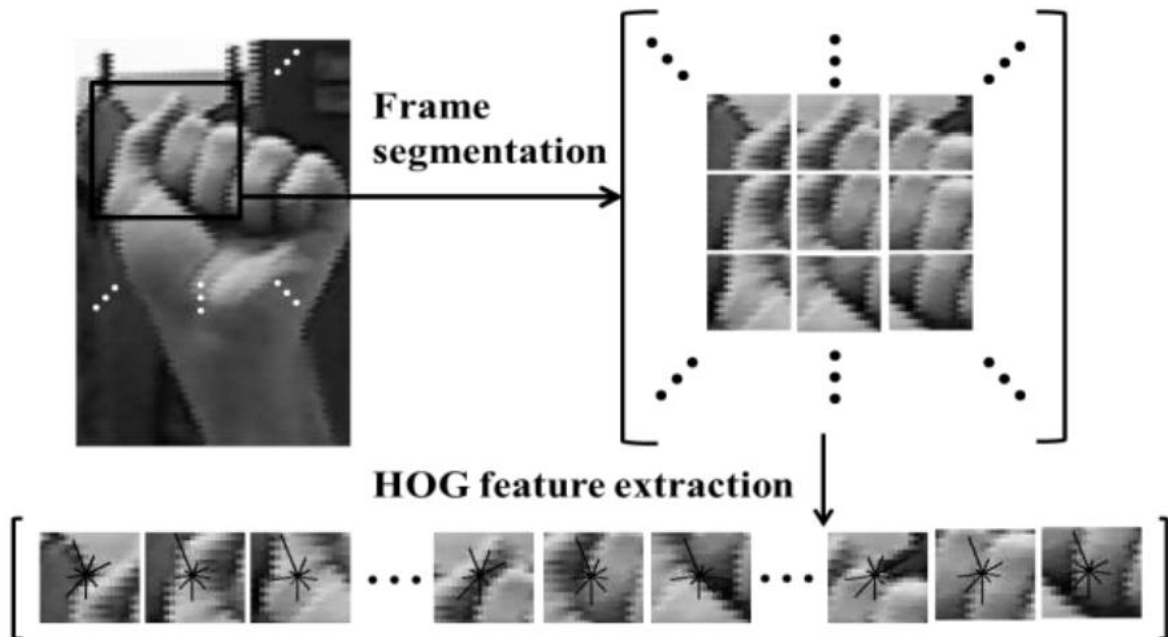
$$\theta(u,v) = \arctg \frac{g_y(u,v)}{g_x(u,v)} \quad (2-2)$$

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 دامنه‌گرادیان به صورت زیر تعریف می‌شوند.

$$f_i = [f_i(n)]_{n \in \{1,2,\dots,N\}}^T \quad (3-2)$$

$$f_i(n) = \sum_{(u,v) \in f_i} g(u,v) \delta[\text{bin}(u,v) - n] \quad (4-2)$$

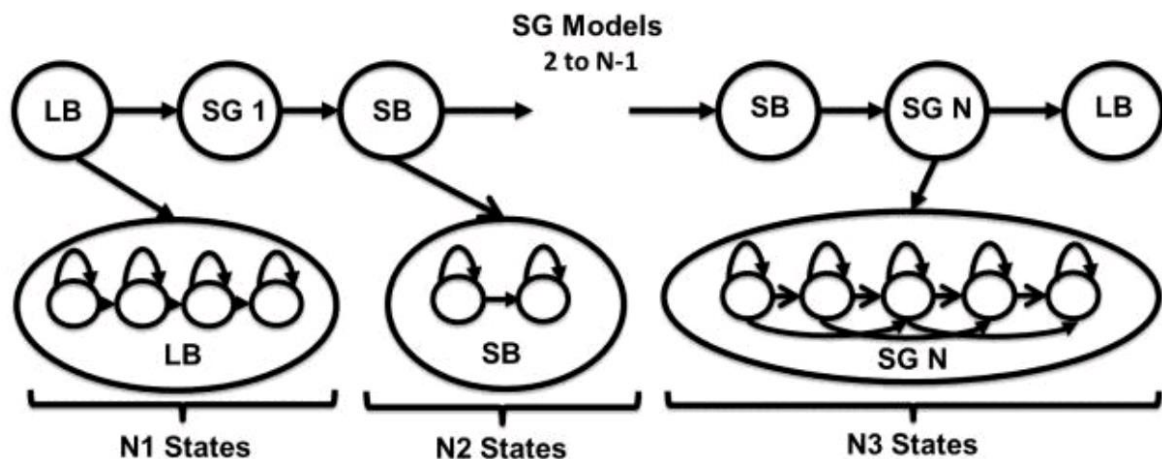
تابع $\text{bin}(u,v)$ اندیس مجموعه مرتبط با پیکسل (u,v) را بر می‌گرداند. سپس این ویژگی‌ها به عنوان ورودی به ساختار HMM دو سطحی اعمال می‌شود سطح اول این مدل مخفی مارکوف، قيود نحوی را در این سیستم اعمال می‌کند. این سطح به دو منظور استفاده شده است. اول اینکه، این سطح سیستم را به خروجی یک فرض در هر تصویر 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 قيود نحوی اشاره دارد. این قيود از خطاهای الحاق^{۱۰}، به وسیله ممانعت از فرضیات چندگانه در هر تصویر جلوگیری می‌کند. دوم اینکه هر فریم تصویر را مجبور به طبقه‌بندی 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 مجدداً به صورت یک تعداد اختیاری از فریم‌های پس‌زمینه مدل می‌شوند. 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 ۱-۲، 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 پس‌زمینه 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. برخلاف 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

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 پیشینه سازی امید ریاضی EM و به طور خاص تر، الگوریتم BW برای آموزش مدل مخفی مارکوف استفاده شده است لازم به ذکر است که تمام مدل های HMM به طور همزمان در هر تکرار به روز می شوند.

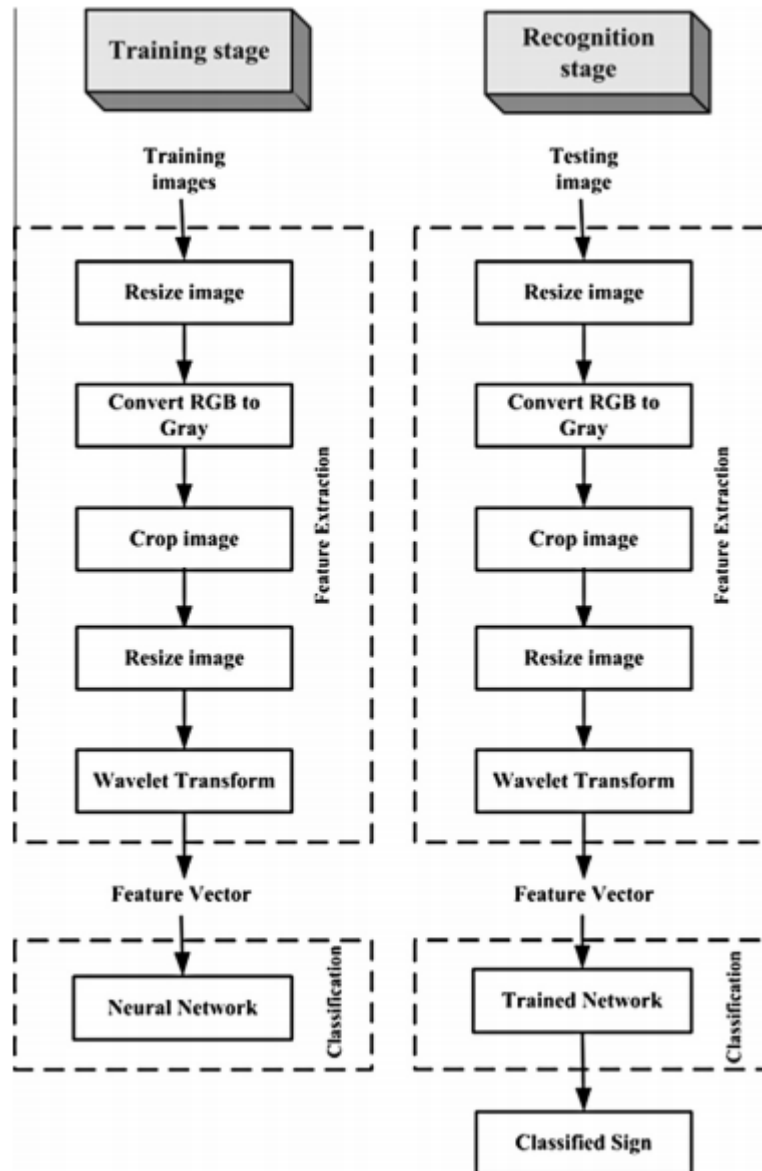
۲-۱-۲ - بازشناسی زبان اشاره فارسی با استفاده از تبدیل موجک و شبکه های عصبی

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



شکل ۲-۴: الفبای اشاره فارسی [۲].

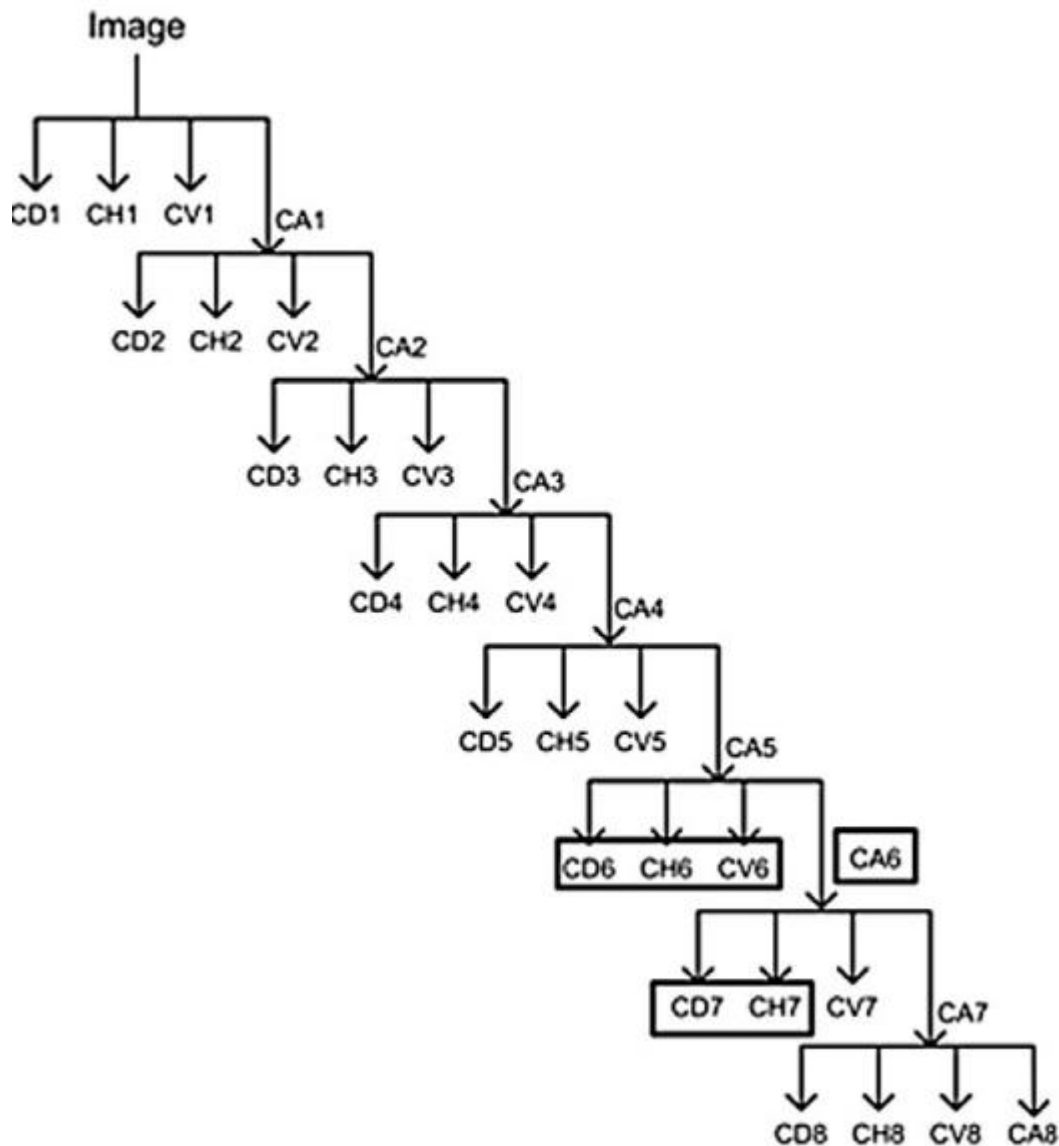
کریمی و همکارانش یک سیستم بازشناسی الفبای اشاره فارسی مبتنی بر تبدیل موجک و شبکه عصبی ارائه کرده‌اند [۲]. این سیستم شامل دو فاز استخراج ویژگی و طبقه‌بندی می‌باشد. بلوک دیاگرام این سیستم در شکل ۲-۵ نشان داده شده است.



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

در مرحله استخراج ویژگی، تصاویر رنگی به 280×350 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 200×300 پیکسل تغییر اندازه داده می‌شوند. در این سیستم جهت استخراج ویژگی از هر تصویر به دلیل سادگی و موثر بودن تبدیل موجک هر 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 تقریباً غیرممکن است که تمام 200×300 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 کوچک‌تر می‌شود. به عبارت دیگر برای انتخاب



شکل ۲-۶: درخت تجزیه تبدیل موجک [۲].

¹² Approximation

¹³ Details

¹⁴ Decomposition