

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صنعتی شاهرود

دانشکده مهندسی کامپیوتر و فناوری اطلاعات
رشته مهندسی کامپیوتر - گرایش هوش مصنوعی

رساله دکتری

بازشناسی انسان با استفاده از ویژگی‌های مدل سه‌بعدی در

سیستم‌های نظارت ویدئویی

نگارنده: علی سبطی

استاد راهنما

دکتر حمید حسن‌پور

استاد مشاور

دکتر علی سلیمانی ایوری

شهریور ۱۳۹۶

ب



فرم شماره ۱۲: صورت جلسه نهایی دفاع از رساله دکتری (Ph.D)
(ویژه دانشجویان ورودی های ۹۴ و ما قبل)

بدینوسیله گواهی می شود آقای علی سبطی دانشجوی دکتری رشته مهندسی کامپیوتر - هوش مصنوعی به شماره دانشجویی ۹۰۲۴۷۴۵ ورودی بهمن ماه سال ۱۳۹۱ در تاریخ ۱۳۹۶/۰۶/۰۷ از رساله نظری/ عملی خود با عنوان: پارشناسی انسان با استفاده از ویژگی های مدل سه بعدی در سیستم های نظارت ویدئویی دفاع و با اخذ نمره ۱۸/۱۱ به درجه کسری نائل گردید.

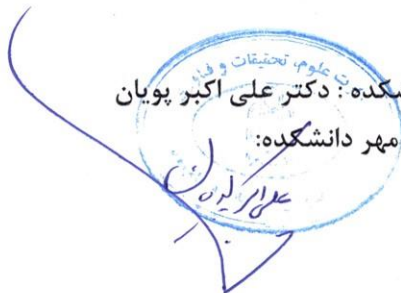
<input checked="" type="checkbox"/> الف) درجه عالی: نمره ۱۹-۲۰	<input type="checkbox"/> ب) درجه بسیار خوب: نمره ۱۸/۹۹ - ۱۷
<input type="checkbox"/> ج) درجه خوب: نمره ۱۶/۹۹ - ۱۵	<input type="checkbox"/> د) غیر قابل قبول و نیاز به دفاع مجدد دارد
<input type="checkbox"/> ه) رساله نیاز به اصلاحات دارد	

ردیف	هیئت داوران	نام و نام خانوادگی	مرتبۀ علمی	امضاء
۱	دکتر حمید حسن پور	استاد راهنما	استاد	
۲	دکتر سید علی سلیمانی ایوری	مشاورین	دانشیار	
۳	دکتر اسدالله شاه بهرامی	استاد مدعو خارجی	دانشیار	
۴	دکتر وحید ابوالقاسمی	استاد مدعو داخلی	استادیار	
۵	دکتر منصور فاتح	استاد مدعو داخلی	استادیار	
۶	دکتر هدی مشایخی	سرپرست (نماینده) تحصیلات تکمیلی دانشکده	استادیار	

مدیر محترم تحصیلات تکمیلی دانشگاه:

ضمن تأیید مراتب فوق مقرر فرمائید اقدامات لازم در خصوص انجام مراحل دانش آموختگی آقای علی سبطی بعمل آید.

نام و نام خانوادگی رئیس دانشکده: دکتر علی اکبر پویان
تاریخ و امضاء و مهر دانشکده:





پیوست شماره ۲

دانشکده: مهندسی کامپیوتر
گروه: هوش مصنوعی

رساله دکتری آقای علی سبطی

تحت عنوان: بازشناسی انسان با استفاده از ویژگی‌های مدل سه‌بعدی در سیستم‌های نظارت ویدئویی
در تاریخ ۱۳۹۶/۰۶/۰۸ توسط کمیته تخصصی زیر جهت اخذ مدرک رساله دکتری ارزیابی گردید و با درجه بسیار خوب مورد پذیرش قرار گرفت.

امضاء	اساتید مشاور	امضاء	اساتید راهنما
	نام و نام خانوادگی: دکتر سید علی سلیمانی ایوری		نام و نام خانوادگی: دکتر حمید حسن پور

امضاء	نماینده تحصیلات تکمیلی	امضاء	اساتید داور
	نام و نام خانوادگی: دکتر هدی مشایخی		نام و نام خانوادگی: دکتر اسدالله شاه بهرامی
			نام و نام خانوادگی: دکتر منصور فاتح
			نام و نام خانوادگی: دکتر وحید ابوالقاسمی

تقدیم و قدردانی

این رساله را ضمن تشکر و سپاس بیکران و در کمال افتخار و اطمینان تقدیم می‌نمایم به:

روح پاک پدرم که عالمانه به من آموخت تا چگونه در عرصه زندگی، ایستادگی را تجربه نمایم

و به مادرم، دریای بی‌کران فداکاری و عشق، که وجودم برایش همه نجب بود و وجودش برایم همه مهر

و استاد با کمالت و شایسته، جناب آقای دکتر حمید حسن پور که در کمال سعه صدر، با حسن خلق و فروتنی،

از هیچ کجی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این رساله را بر عهده گرفتند.

خدایا توفیق خدمتی سرشار از شور و نشاط و همراه و همسوا علم و دانش و پژوهش

جهت رشد و شکوفایی ایران کنسال عنایت بفرما

تعهد نامه

این جانب علی سبطی دانشجوی دوره دکتری رشته مهندسی کامپیوتر/هوش مصنوعی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی شاهرود نویسنده‌ی پایان‌نامه **بازشناسی انسان با استفاده از ویژگی‌های مدل سه‌بعدی در سیستم‌های نظارت ویدئویی**، تحت راهنمایی آقای دکتر حمید حسن‌پور متعهد می‌شوم :

- تحقیقات در این پایان‌نامه توسط این جانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورداستفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ:

امضای دانشجو:

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه ای، نرم افزارها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

چکیده

نظارت ویدئویی هوشمند یکی از کاربردهای اصلی و مهم در بینایی ماشین است. پیکربندی رایج این سیستم‌ها، مجموعه‌ای از دوربین‌های با میدان دید ناهمپوشان با هدف نهایی تشخیص وقایع غیرنرمال و تجزیه و تحلیل رفتاری افراد می‌باشد. باشناسی افراد از گام‌های اصلی این سیستم‌ها به شمار می‌رود که عبارت است از برچسب‌گذاری افراد رؤیت شده در شبکه‌ای از دوربین‌ها به طوری که برای هر فرد مشخص، یک برچسب واحد اختصاص یابد. در چنین سیستم‌هایی یک فرد می‌تواند در زمان‌ها و موقعیت‌های مختلف در مقابل مجموعه‌ای از دوربین‌ها قرار گیرد.

به دلایل مختلف مسئله باشناسی جزو مسائل پیچیده در بینایی ماشین است. دلیل این پیچیدگی را می‌توان در تغییرات ظاهری شخص از دوربینی به دوربین دیگر دانست. این تغییرات می‌تواند به دلیل محدودیت‌های سخت‌افزاری مانند کیفیت پایین دوربین‌ها، پاسخ‌های متفاوت دوربین‌ها نسبت به رنگ‌ها، تغییر شرایط نوری در موقعیت‌های مختلف نصب دوربین‌ها یا تغییرات زاویه‌ای در نحوه قرارگیری شخص در مقابل دوربین‌ها به وجود آید. یکی از نقاط ضعف اصلی در روش‌های موجود، عدم توجه به اطلاعات زاویه‌ای است. در واقع، زمانی که کیفیت تصاویر ثبت شده پایین است یا افراد از دوربین دور هستند الگوریتم‌های موجود به‌ناچار اطلاعات مهمی نظیر زاویه قرارگیری شخص در برابر دوربین را نادیده می‌گیرند.

هدف از پژوهش حاضر مدل‌سازی قضاوت ناظر انسانی در خصوص تغییرات بصری پوشش شخص هدف، منتج از تغییر در زاویه دید است. برای این منظور از اطلاعات زاویه قرارگیری شخص در برابر دوربین و تغییرات سه‌بعدی نشأت گرفته از آن استفاده می‌شود. در رویکرد ارائه شده، ابتدا زاویه قرارگیری شخص در تصویر با دو روش پیشنهادی، یکی بر پایه اطلاعات کانتوری شخص در تصویر و دیگری با بهره‌گیری از ویژگی هیستوگرام گرادیان‌های جهتی و کلاسه‌بند رگرسیون لجستیک استخراج می‌گردد. سپس نواحی خاصی که تحت چرخش‌های زاویه می‌توانند آشکار یا نهان شوند مشخص می‌گردد. بدین منظور از ناحیه سر در تصویر به‌عنوان یک نقطه کلیدی استفاده می‌شود که با استفاده از شبکه عصبی کانولوشن استخراج می‌گردد. در مرحله تطبیق از فرآیند باشناسی، نواحی استخراج شده وزن دهی شده یا الگوی رنگی و بافتی نواحی مجاور را به خود می‌گیرند که در رویکرد پیشنهادی از آن به‌عنوان فرآیند یکسان‌سازی یاد می‌کنیم.

در طراحی سیستم پیشنهادی سعی شده ملاحظات مرتبط با پیچیدگی محاسباتی در کلیه قسمت‌ها مورد توجه قرار گیرد. از ویژگی‌های قابل توجه این سیستم قابلیت استفاده و تلفیق در اغلب روش‌های باشناسی است. در این پژوهش، رویکرد پیشنهادی بر روی دو الگوریتم پیشگام در کاربرد باشناسی افزوده شد. نتایج حاصل از اعمال این روش بر روی مجموعه داده‌های ViPER که به‌عنوان یکی از دشوارترین مجموعه داده‌های حوزه باشناسی مطرح است، نشان‌دهنده بهبود در دقت باشناسی به میزان ۱,۳ درصد نسبت به دو الگوریتم مذکور و اثربخش بودن یکسان‌سازی پیشنهادی است.

واژگان کلیدی: باشناسی افراد، سیستم نظارت ویدئویی، مدل سه‌بعدی، تشخیص زاویه، شبکه عصبی کانولوشن

مقالات مستخرج از رساله

- علی سبطی، حمید حسن پور، "بازشناسی انسان در سیستم‌های نظارت ویدئویی"، مجله محاسبات نرم، دانشگاه کاشان، دوره سوم، شماره اول، صفحات ۶۲-۸۱، بهار ۱۳۹۳
- علی سبطی، حمید حسن پور، "بهبود الگوریتم SDALF در بازشناسی انسان با بهره‌گیری از اطلاعات زاویه شخص"، بیست و سومین کنفرانس مهندسی برق ایران، دانشگاه صنعتی شریف، تهران، ایران، ۱۳۹۴.
- علی سبطی، حمید حسن پور، "تشخیص زاویه قرارگیری شخص در تصویر با استفاده از اطلاعات کانتوری"، نشریه مهندسی برق و مهندسی کامپیوتر ایران، ب- مهندسی کامپیوتر، سال ۱۴، شماره ۴، صفحات ۳۲۲-۳۳۲، زمستان ۱۳۹۵.
- Ali Sebti, Hamid Hassanpour, "Body Orientation Estimation with the Ensemble of Logistic Regression Classifiers", *Multimedia Tools and Applications*, Springer, 2016, doi: 10.1007/s11042-016-4129-0
- علی سبطی، حمید حسن پور، "بازشناسی انسان در سیستم‌های نظارت ویدئویی با بهره‌گیری از اطلاعات زاویه‌ای"، مجله مهندسی برق دانشگاه تبریز، تابستان ۱۳۹۶. (تحت داوری)

فهرست مطالب

- ۱- مقدمه‌ای بر بازشناسی انسان در سیستم‌های نظارت ویدئویی..... ۱
- ۱-۱- تعریف مسئله..... ۳
- ۱-۲- چالش‌های موجود در بازشناسی افراد..... ۴
- ۱-۳- انواع پیکربندی‌های مسئله بازشناسی افراد..... ۶
- ۱-۳-۱- مجموعه باز در مقابل مجموعه بسته..... ۷
- ۱-۴- فازهای پیشین بازشناسی در سیستم‌های نظارت ویدئویی..... ۹
- ۱-۴-۱- آشکارساز انسان..... ۱۰
- ۱-۴-۲- حذف پس‌زمینه..... ۱۰
- ۱-۴-۳- حذف سایه..... ۱۳
- ۱-۵- هدف رساله..... ۱۴
- ۱-۶- مفروضات مورد نظر برای مسئله بازشناسی انسان..... ۱۵
- ۱-۷- مجموعه داده‌ها به منظور ارزیابی..... ۱۶
- ۱-۸- نوآوری‌ها و دستاوردهای رساله..... ۱۷
- ۱-۹- ساختار رساله..... ۱۹
- ۲- مروری بر کارهای پیشین..... ۲۱
- ۲-۱- روش‌های مبتنی بر اطلاعات تصاویر..... ۲۳
- ۲-۱-۱- روش‌های غیرفعال..... ۲۴
- ۲-۱-۱-۱- مدل ظاهری مکانی-زمانی..... ۲۵
- ۲-۱-۱-۲- بازشناسی چندتصویره انسان به کمک توصیفگر HPE..... ۲۸
- ۲-۱-۱-۲-۳- تجمیع ویژگی‌های محلی بر پایه قرینگی..... ۲۹
- ۲-۱-۱-۲-۴- روش‌های بر پایه توصیفگر کوواریانس..... ۳۲
- ۲-۱-۱-۲-۵- مدل مکانی-زمانی بر پایه بخش‌بندی بدن انسان..... ۴۰
- ۲-۱-۱-۲-۶- بازشناسی سریع بر اساس بازنمایی عدم شباهت..... ۴۱
- ۲-۱-۱-۲-۷- بازشناسی بر پایه هیستوگرام رنگی و بافت..... ۴۳

۴۸	۲-۱-۲- روش‌های فعال
۴۸	۲-۱-۲-۱- هم‌ترازی رنگی
۴۸	۲-۲-۱-۲- روش‌های یادگیری توصیفگر
۵۵	۳-۲-۱-۲- روش‌های یادگیری معیار فاصله
۶۱	۲-۲- روش‌های مبتنی بر اطلاعات پیکربندی دوربین‌ها
۶۲	۱-۲-۲- استخراج موقعیت قرارگیری نسبی دوربین‌ها
۶۴	۲-۲-۲- هم‌تراز کردن دوربین‌ها
۶۷	۳-۲- جمع‌بندی
۶۹	۳- روش پیشنهادی
۷۰	۱-۳- مقدمه
۷۷	۲-۳- تشخیص زاویه قرارگیری شخص در تصویر
۷۹	۱-۲-۳- تشخیص زاویه با استفاده از اطلاعات کانتوری
۸۳	۱-۱-۲-۳- تبدیل فاصله
۸۵	۲-۱-۲-۳- خوشه‌بندی کانتورها
۸۷	۳-۱-۲-۳- پیمایش خوشه‌بندی
۹۰	۴-۱-۲-۳- محاسبه زاویه قرارگیری شخص به کمک ترکیبی از درخت‌های کانتوری
۹۱	۵-۱-۲-۳- ترکیب هیستوگرام زاویه‌ای و HOG به‌عنوان ویژگی نهایی
۹۲	۶-۱-۲-۳- داده‌های مورد استفاده و آموزش
۹۳	۷-۱-۲-۳- آزمایش‌ها و نتایج
۹۷	۸-۱-۲-۳- تعیین حد آستانه در ساخت و پیمایش درخت خوشه‌بندی
۹۸	۲-۲-۳- تشخیص زاویه با استفاده از ترکیب کلاسه‌بندهای رگرسیون لجستیک
۹۹	۱-۲-۲-۳- توصیفگر HOG
۱۰۰	۲-۲-۲-۳- رگرسیون لجستیک چندجمله‌ای
۱۰۱	۳-۲-۲-۳- استخراج نواحی ایستا، پویا و آموزش کلاسه‌بندها
۱۰۵	۴-۲-۲-۳- ترکیب نتایج کلاسه‌بندها
۱۰۷	۵-۲-۲-۳- نتایج آزمایش‌ها

۱۱۲	۳-۳- استخراج ناحیه سر در تصویر
۱۱۲	۳-۳-۱- شبکه عصبی کانولوشن
۱۱۵	۳-۳-۲- آماده‌سازی داده‌ها و انتخاب معماری شبکه
۱۱۹	۳-۳-۳- استفاده از شبکه عصبی کانولوشن در تشخیص زاویه
۱۲۲	۳-۴- یکسان‌سازی تصاویر
۱۲۴	۳-۵- تلفیق مرحله یکسان‌سازی با الگوریتم‌های بازشناسی
۱۲۵	۳-۵-۱- ویژگی ماکزیمم وقوع محلی و تلفیق با روش پیشنهادی
۱۲۷	۳-۵-۲- توصیفگر گاوسی سلسله‌مراتبی و تلفیق با روش پیشنهادی
۱۲۸	۳-۶- جمع‌بندی
۱۲۹	۴- ارزیابی و نتایج آزمایش‌ها
۱۳۲	۴-۱- ارزیابی روش ناآگاهانه با دانستن اطلاعات زاویه‌ای
۱۳۴	۴-۲- ارزیابی روش آگاهانه به‌همراه تخمین خودکار زاویه
۱۳۷	۴-۳- بازشناسی چندتصویره (دنباله ویدئویی)
۱۳۹	۴-۴- جمع‌بندی
۱۴۱	۵- جمع‌بندی و کارهای آینده
۱۴۲	۵-۱- جمع‌بندی رویکرد پیشنهادی در رساله
۱۴۴	۵-۲- کارهای آینده
۱۴۷	واژه‌نامه
۱۴۹	مراجع

فهرست تصاویر

- شکل ۱-۱: انواع تغییرات ظاهری یک فرد در سیستم نظارتی ویدئویی [۱] ۶
- شکل ۲-۱: نمایی کلی از یک سیستم بازشناسی ۸
- شکل ۳-۱: تغییرات مقادیر رنگی برای یک پیکسل از سه دنباله ویدئویی در طول زمان، تصویر الف، مقدار یک پیکسل را در بازه‌های زمانی مختلف از روز نشان می‌دهد که تغییرات تدریجی موقعیت مُد اصلی توزیع را شاهدیم. تصویر ب، مقدار پیکسلی در محل آب را می‌بینیم که توزیعی دو مُده می‌باشد. تصویر پ، توزیع رنگی پیکسلی واقع بر روی ناحیه مانیتور است که به صورت تناوبی بین دو مقدار تغییر می‌کند که به صورت دو مُده قابل مشاهده است. ۱۲
- شکل ۱-۲: دسته‌بندی روش‌های بازشناسی افراد در سیستم‌های نظارتی [۱] ۲۳
- شکل ۲-۲: نتیجه قطعه‌بندی به روش حوزه آبریز در ناحیه پیش‌زمینه برای چند فریم متوالی [۶].. ۲۶
- شکل ۳-۲: استخراج لبه‌های با اهمیت در تصویر با استفاده از الگوریتم ارائه شده در [۶]. تصویر الف: تصویر ورودی، تصویر ب: عکس فرکانسی تصویر پ: نتیجه خوشه‌بندی گراف و تصویرت: مقدار رنگی میانه برای هر بخش از خروجی خوشه‌بندی. ۲۶
- شکل ۴-۲: تصویر راست مدل مثلثی، تصویر چپ: دو نمونه انطباق به کمک لبه‌های بااهمیت و قسمت ناحیه پیش‌زمینه بدست آمده از الگوریتم بخش‌بندی مبتنی بر پارتیشن‌بندی گراف [۶] ۲۷
- شکل ۵-۲: خلاصه‌ای از الگوریتم ارائه‌شده در [۱۴] ۲۹
- شکل ۶-۲: سه ویژگی از معرفی شده در الگوریتم SDALF برای شش تصویر. ستون الف: تصویر اصلی، ستون ب: استخراج نواحی تنه و پاها و محورهای قرینگی، ستون پ: تصویر back-projection هیستوگرام وزن‌دار، ستون ت: توصیفگر MSCR و ستون ث: توصیفگر RHSP [۱۵] ۳۱
- شکل ۷-۲: طرح کلی الگوریتم مقاله [۱۸] ۳۴
- شکل ۸-۲: ساختار هرمی در سه سطح به منظور استخراج توصیفگر کوواریانس در مقاله [۲۳] ۳۶
- شکل ۹-۲: خلاصه‌ای از الگوریتم محاسبه توصیفگر گاوسی سلسله مراتبی [۲۵] ۴۰
- شکل ۱۰-۲: طرح کلی از روش ارائه‌شده در [۲۶] ۴۱

- شکل ۲-۱۱: سطر الف: بخش‌بندی تصویر فرد به دو ناحیه و استخراج مؤلفه‌های ویژگی. سطر ب: ادغام کلیه مؤلفه‌های مرتبط با یک بخش. سطر پ: خوشه‌بندی هر بخش و استخراج نماینده‌ها [۲۷] ۴۲
- شکل ۲-۱۲: مرحله بازشناسی در بستر عدم شباهت [۲۷] ۴۳
- شکل ۲-۱۳: نحوه محاسبه ویژگی ماکزیمم وقوع محلی در [۲۸] ۴۴
- شکل ۲-۱۴: نمونه‌ای از بازنمایی تصویر یک فرد، بر اساس توصیفگر SCNCD برای ۱۶ رنگ برجسته [۳۱] ۴۵
- شکل ۲-۱۵: مقایسه بازسازی یک تصویر از روی هیستوگرام مکانی و هیستوگرام استاندارد [۳۲] ۴۷
- شکل ۲-۱۶: تقسیم‌بندی ناحیه تنه بر اساس روش ارائه شده در [۳۳] ۴۷
- شکل ۲-۱۷: ساختار شبکه عصبی مورد استفاده در [۳۹] (شامل ویژگی ELF16 و ویژگی‌های حاصل از شبکه عصبی کانولوشن) ۵۱
- شکل ۲-۱۸: ویژگی F بدست آمده از دو فرد برای دو دوربین A و B بهم متصل شده و بردار ساخته شده به کلاسه بند C ارسال می‌گردد [۴۰] ۵۲
- شکل ۲-۱۹: مراحل الگوریتم ارائه شده در مقاله [۴۱] ۵۴
- شکل ۲-۲۰: تناظر مکانی زیرنواحی تصاویر ثبت شده برای یک جفت دوربین، الف و ب: تناظر زیرنواحی برای تصاویر ثبت شده از دو شخص در دو دوربین، پ: ماتریس تناظر ساختاری برای یک جفت دوربین [۴۹] ۵۹
- شکل ۲-۲۱: خلاصه‌ای از مراحل آموزش و تست در روش پیشنهادی [۴۹] ۵۹
- شکل ۲-۲۲: ساختار شبکه کانولوشنی عمیق پیشنهادی در [۵۰] ۶۰
- شکل ۲-۲۳: تابع توزیع تخمین شده سمت راست ارتباط بین سرعت خروج فرد از دوربین ۲ و زمانی که طول می‌کشد که همان فرد به دوربین ۱ وارد شود (مستطیل سبز فضای سبزی را نشان می‌دهد که افراد کمتر وارد آن می‌شوند) [۵۱] ۶۳
- شکل ۲-۲۴: ارتباط بین موقعیت در راستای Y برای خروج از دوربین ۱ و ورود به دوربین ۲ (نواحی سبزرنگ فضای سبز را در محیط مذکور نشان می‌دهد) [۵۱] ۶۴
- شکل ۲-۲۵: نقشه پانورامیک استخراج شده از تصاویر بدست آمده از چند دوربین با میدان دید مشترک [۵۲] ۶۵
- شکل ۲-۲۶: نقشه پانورامیک بدست آمده برای دو شخص خاص [۵۲] ۶۵

- شکل ۲-۲۷: مدل سه‌بعدی استخراجی برای انسان در مرجع [۵۳]. تصویر الف: مدل سه‌بعدی انسان به همراه جزئیات، تصویر ب: میانگین سیلوئت انسان در دو زاویه مختلف، تصویر پ: مدل سه‌بعدی ساده‌شده پیشنهادی، تصویر ت: نمونه‌برداری از رئوس مورد نیاز در مدل ساده‌سازی شده ۶۶
- شکل ۲-۲۸: فرآیند انطباق تصویر فرد بر روی مدل سه‌بعدی پیشنهادی [۵۳] ۶۶
- شکل ۲-۲۹: مدل ساخته شده برای چند نمونه انسان با استفاده از مدل سه‌بعدی پیشنهادی در [۵۳] ۶۷
- شکل ۳-۱: شش تصویر مربوط به سه شخص از مجموعه داده‌ای ViPER ۷۱
- شکل ۳-۲: نمایی کلی از سیستم پیشنهادی به همراه زیر سیستم‌ها ۷۲
- شکل ۳-۳: فلوجارت رویکرد پیشنهادی در این رساله ۷۵
- شکل ۳-۴: مرحله یکسان‌سازی و استخراج ماسک در رویکرد پیشنهادی ۷۶
- شکل ۳-۵: مدل سه‌بعدی طراحی شده از انسان در زوایا و حالت‌های مختلف راه رفتن ۸۱
- شکل ۳-۶: رنگ‌آمیزی مدل سه‌بعدی به منظور تشخیص اندام ۸۱
- شکل ۳-۷: نمونه‌ای از کانتورهای استخراج شده از مدل سه‌بعدی برای بخش‌های مختلف بدن ۸۲
- شکل ۳-۸: راست: تصویر لبه، چپ: تبدیل فاصله متناظر با تصویر سمت راست ۸۳
- شکل ۳-۹: محاسبه شباهت تصویر دو چندضلعی و شکل ۳-۷ براساس تبدیل فاصله. الف) تصاویر دو چندضلعی، ب) انطباق تصاویر متناظر با سطر الف بر روی تبدیل فاصله‌ی شکل ۳-۷ ۸۵
- شکل ۳-۱۰: خوشه‌بندی سلسله مراتبی کانتورها و تشکیل ابر کانتورها ۸۷
- شکل ۳-۱۱: هیستوگرام زاویه‌ای برای دو گره از درخت شکل ۳-۹ ۸۹
- شکل ۳-۱۲: تصاویر مجموعه داده مورد استفاده جهت آموزش (TUD) [۶۸] ۹۲
- شکل ۳-۱۳: تصاویر نمونه از مجموعه داده‌های استفاده شده در تست (ViPER) [۸] ۹۳
- شکل ۳-۱۴: نتایج به دست آمده از کلاسه‌بندی نهایی با استفاده از ویژگی پیشنهادی بر روی نمودار CMC ۹۵
- شکل ۳-۱۵: ماتریس درهم‌ریختگی برای نتایج کلاسه‌بندی نهایی ۹۶
- شکل ۳-۱۶: خلاصه گام‌های توصیفگر HOG با مقادیر پارامترهای پیش فرض ۱۰۰

- شکل ۳-۱۷: آموزش کلاسه‌بندهای مرتبط با نواحی W، UH و LH ۱۰۲
- شکل ۳-۱۸: خلاصه‌ای از عملیات انجام شده در آموزش کلاسه‌بندهای مرتبط با ناحیه HS ۱۰۴
- شکل ۳-۱۹: مثالی از مکانیزم وزن‌دهی روش پیشنهادی برای خروجی یکی از کلاسه‌بندها ۱۰۶
- شکل ۳-۲۰: خلاصه‌ای از کلیه بخش‌های سیستم پیشنهادی برای تشخیص زاویه ۱۰۷
- شکل ۳-۲۱: ساختار کلی یک شبکه عصبی کانولوشن برای مسئله تشخیص ارقام دست‌نویس لاتین ۱۱۴
- شکل ۳-۲۲: پنج فیلتر آموزش یافته در لایه دوم شبکه عصبی ۱۱۶
- شکل ۳-۲۳: استخراج موقعیت ناحیه سر با استفاده از ترکیب خروجی دو شبکه عصبی کانولوشن ۱۱۸
- شکل ۳-۲۴: مقایسه هیستوگرام فراوانی درصد همپوشانی با ناحیه مطلوب، برای سه روش موقیت‌یابی ناحیه سر، بر روی داده‌های ViPER ۱۱۹
- شکل ۳-۲۵: نرم‌افزار نوشته‌شده برای برچسب‌گذاری زاویه برای مجموعه داده HollywoodHeads ۱۲۱
- شکل ۳-۲۶: مثالی از اجرای الگوریتم پیشنهادی در استخراج نواحی خاص به‌منظور یکسان‌سازی ۱۲۴
- شکل ۳-۲۷: خلاصه الگوریتم نمونه‌افزایی برای یک تصویر و ماسک استخراج شده متناظر آن ۱۲۷
- شکل ۴-۱: یک نمونه از نمودار CMC ۱۳۱
- شکل ۴-۲: نمونه‌ای از مرحله یکسان‌سازی غیرآگاهانه برای تصویری از روبرو، بر اساس وزن‌دهی پیشنهادی ۱۳۳
- شکل ۴-۳: مقایسه نسخه استاندارد SDALF و نسخه اصلاح شده به روش پیشنهادی ۱۳۳
- شکل ۴-۴: مقایسه الگوریتم LOMO+XQDA در دو حالت (تلفیق رویکرد پیشنهادی- ماسک‌گذاری و بدون تلفیق) در قالب نمودار CMC ۱۳۵
- شکل ۴-۵: مقایسه الگوریتم GoG+XQDA در سه حالت (تلفیق با رویکرد پیشنهادی- ماسک‌گذاری و نمونه‌افزایی و بدون تلفیق) در قالب نمودار CMC ۱۳۶

فهرست جداول

- جدول ۱-۳: مقایسه نتایج کلاسه‌بندی کلی روش پیشنهادی و روش‌های موجود ۱۰۸
- جدول ۲-۳: ماتریس درهم‌ریختگی برای الگوریتم ارائه شده در [۶۳] ۱۰۸
- جدول ۳-۳: ماتریس درهم‌ریختگی برای روش پیشنهادی ۱۰۹
- جدول ۴-۳: مقایسه پیچیدگی محاسباتی روش پیشنهادی و روش‌های دیگر موجود ۱۱۱
- جدول ۵-۳: معماری شبکه عصبی کانولوشن برای تشخیص ناحیه سر از ناحیه غیرسر ۱۱۶
- جدول ۶-۳: معماری شبکه عصبی کانولوشن برای تشخیص ناحیه سر و شانه ۱۱۷
- جدول ۱-۴: نتایج بازشناسی برای تلفیق رویکرد پیشنهادی با الگوریتم‌های XQDA+LOMO و GoG+XQDA و مقایسه با روش‌های دیگر برای سه رتبه ۱، ۱۰ و ۲۰ ۱۳۷
- جدول ۲-۴: مقایسه دقت تخمین زاویه برای حالت تک تصویر و دو تصویر ۱۳۸

فهرست علائم و نشانه‌ها

نماد مجموعه	\mathbb{A}
مجموعه‌ای شامل دو عضو ۰ و ۱	$\{0,1\}$
برداری با n مؤلفه	(x_1, x_2, \dots, x_n)
عنصر a از بردار a	a_i
شناسه تخصیص یافته به عضو x	$id(x)$
توزیع احتمال روی متغیر تصادفی a	$p(a)$
اندیس‌دهی در دو سطح، سطح بالاتر (j)، سطح پایین‌تر (i)	a_i^j
نماد زمان در لحظه t	t
مقدار رنگ پیکسل در مختصات (x, y)	$I(x, y)$
نماد وزن	w
تابع فاصله بین دو متغیر	$d(x, y)$
نماد میانگین	μ
ترانپوز ماتریس A	A^T
همبستگی دو متغیر x و y	$\rho(x, y)$
لگاریتم طبیعی x	$\ln(x)$
نماد گرادیان	∇
تابع نمایی در پایه ثابت e	$exp(x)$
نماد انحراف معیار	σ
توزیع نرمال روی متغیر x با میانگین μ و کوواریانس Σ	$\mathcal{N}(x; \mu, \Sigma)$
تابع f با دامنه مجموعه \mathbb{X} و برد مجموعه \mathbb{Y}	$f: \mathbb{X} \rightarrow \mathbb{Y}$
L^p نرم x	$\ x\ ^p$
تبدیل فاصله تصویر باینری I	$Dt(I)$
مجموع ضرب درایه‌های متناظر دو ماتریس A و B	$A \circ B$
مجموع درایه‌های ماتریس A	$ A _{\oplus}$
امید ریاضی متغیر تصادفی X	$E[X]$

فهرست اختصارات

A

ACM Active Color Model

B

BTF Brightness Transfer Function

C

CFS Correlation based Feature Selection

CMC..... Cumulative Match Curve

CNN..... Convolutional Neural Network

D

DCD..... Dominant Color Descriptor

DT Distance Transform

E

ELF Ensemble of Localized Features

F

FAR..... False Acceptance Rate

G

GMM Gaussian Mixture Model

H

HOG... Histogram of Oriented Gradients

HPE..... Histogram Plus Epitome

HSV Hue-Saturation-Value

L

Lab Lightness-a-b

LBP Local Binary Pattern

LNCC..... Local Normalized Cross-Correlation

LOMO..... Local Maximal Occurance

M

MAP..... Maximum a Posteriori

MLE Maximum Likelihood Estimation

MRCG.... Mean Reimannian Covariance Grid

MSCR Maximally Stable Color Regions

P

PCA..... Principal Component Analysis

R

ReLU Rectified Linear Units

RGB Red-Green-Blue

RHSP Recurrent High-Structured Patches

RMC..... Representative Meta Colors

S

SCNCD Salient Color Names based Color Descriptor

SDALF Symmetry-Driven Accumulation of Local Feature

SIFT Scale Invariant Feature Transform

SILTP Scale Invariant Local Terrary Pattern

Stel Structure element

SVM..... Support Vector Machine

T

TP True Positive

X

XQDA Crossview Quadratic Discriminant Analysis

فصل ۱. مقدمه‌ای بر بازشناسی انسان در
سیستم‌های نظارت ویدئویی

مقدمه

سیستم‌های نظارت ویدئویی امروزه در موقعیت‌های مختلفی از جمله اماکن عمومی به‌طور گسترده‌ای مورد استفاده قرار می‌گیرند. یکی از اهداف استقرار این سیستم‌ها نظارت بر رفتار افراد در محیط می‌باشد. متأسفانه به دلیل افزایش حملات تروریستی و رفتارهای غیرنرمال در سال‌های اخیر، تقاضا برای سیستم‌های نظارتی تحت شبکه، بیشتر گردیده تا بدین طریق امنیت بیشتری فراهم گردد. از جمله اماکنی که در آن، چنین سیستم‌هایی مورد استفاده قرار می‌گیرد می‌توان به فرودگاه‌ها، ایستگاه‌های قطار و حتی محیط‌های داخل هواپیما و قطار، مراکز خرید، مجتمع‌های تجاری و محیط‌های درون سازمان‌ها اشاره نمود [۱].

بازشناسی افراد یکی از گام‌های اولیه و پیش‌نیاز در پردازش‌های هوشمندی نظیر شناسایی و آنالیز فعالیت رفتاری افراد، بازشناسی وقایع و رویدادها و تجزیه و تحلیل صحنه می‌باشد. پردازش‌های فوق زمانی بهترین نرخ تشخیص را خواهند داشت که بازشناسی افراد با دقت خوبی صورت گرفته باشد. سخت‌افزارهای مورد استفاده، دوربین‌های مدار بسته‌ای می‌باشند که داده‌های ویدئویی را فراهم می‌آورند. در چنین سیستم‌هایی هدف، طراحی سیستم‌های خودکار و مستقل از کاربر است که بتواند موارد ذکرشده در بالا را با دقت خوبی استخراج نمایند. فرض کنید یک فرد در برابر یک دوربین در حال حرکت باشد. تا زمانی که فرد در میدان دید این دوربین قرار دارد، سیستم هوشمند، با استفاده از اطلاعات مکانی قادر به ردیابی فرد مذکور است. مشکل زمانی رخ می‌دهد که این فرد از میدان دید دوربین مذکور خارج شده و وارد میدان دید دوربین دیگری در شبکه دوربین‌ها می‌گردد. از آنجایی که موقعیت دوربین‌ها نسبت به یکدیگر مشخص نیست، بنابراین اطلاعات همسایگی مکانی و زمانی برای افراد حاضر در سیستم، در دسترس نیست. به عبارتی، زمانی که یک فرد از میدان دید دوربینی خارج می‌گردد مشخص نیست که چه زمانی و در میدان دید کدام دوربین دیگر ظاهر خواهد گردید. در نتیجه

تشخیص اینکه فرد جدید رؤیت شده، کدام فرد دیده شده در دوربین‌های دیگر است، تحت عنوان مسئله بازشناسی افراد تعریف می‌گردد.

از مفروضاتی که در اکثر روش‌های ارائه‌شده مورد استفاده قرار می‌گیرد این است که ظاهر افراد از یک دوربین به دوربین دیگر بدون تغییر است. در نتیجه ارائه روش و سیستم‌های هوشمند و خودکاری که بتواند بر مشکلات اشاره شده در بالا غلبه نماید از اهمیت ویژه‌ای برخوردار است. این سیستم‌ها علاوه بر اینکه باید پیچیدگی محاسباتی قابل قبولی داشته باشند همچنین باید نسبت به تغییرات سه‌بعدی حالت افراد در دوربین‌های مختلف، شرایط نوری گوناگون و وقوع انواع همپوشانی مقاوم باشند.

۱-۱- تعریف مسئله

طراحی سیستم‌های نظارت ویدئویی هوشمند از زمینه فعالیت‌های تحقیقاتی مورد توجه و جدید در علوم کامپیوتر می‌باشد. این کاربرد به دلیل پیچیدگی فازهای مختلف آن، شامل انواع گوناگونی از روش‌های بینایی ماشین و یادگیری ماشین است. یکی از قسمت‌های پایه‌ای و اصلی در هر سیستم نظارت ویدئویی بازشناسی افراد می‌باشد. این بازشناسی در واقع با در نظر گرفتن تعدادی از دوربین‌ها در محیط معنی می‌یابد. به دلیل هزینه‌بر بودن دوربین‌ها و نصب و نگهداری آن‌ها، معمولاً سعی می‌شود که با کمترین تعداد دوربین، کل محیط مورد نظر پوشش داده شود. بدین ترتیب در اکثر موارد این دوربین‌ها دارای میدان دید ناهمپوشان هستند.

با این پیش‌زمینه می‌توان بازشناسی افراد را چنین تعریف نمود: بازشناسی افراد عبارت است از فرآیند شناسایی افراد از تصاویر و یا ویدئوهای بدست آمده از دوربین‌های مختلف. به عبارت دیگر، هدف این فرآیند، برچسب‌زنی افراد در دوربین‌های مختلف می‌باشد. این برچسب می‌تواند مشابه برچسب افرادی باشد که تاکنون در سیستم به صورت یک تصویر و یا دنباله‌ای از فریم‌های یک ویدئو دیده شده‌اند. تخصیص برچسب جدید به افراد زمانی انجام می‌شود که برای مثال فردی در یکی از دوربین‌ها دیده

شده و می‌خواهیم بدانیم این فرد احتمالاً در گذشته در کدام یک از دوربین‌های دیگر رؤیت گردیده است. حتی زمانی که در یک دوربین، فرآیند ردیابی یک شخص بنا بر علتی مختل و گم می‌گردد، فرآیند بازشناسی به منظور شناسایی مجدد فرد و بازبازی مجدد ردیابی مورد استفاده قرار می‌گیرد. تعداد زیاد افراد که منجر به بزرگ شدن فضای جستجو می‌گردد، کیفیت پایین دوربین‌ها، شرایط نوری متفاوت دوربین‌ها و زاویه‌های مختلف افراد در دوربین‌های مختلف، شرایط بازشناسی را مشکل‌تر می‌سازد.

۱-۲- چالش‌های موجود در بازشناسی افراد

مسائل و مشکلاتی که فرآیند بازشناسی افراد را با دشواری همراه می‌سازد وابسته به ویژگی‌های محیطی است که سیستم نظارتی در آن نصب می‌شود. برای نمونه در یک فروشگاه شلوغ یا در یک ایستگاه قطار دشواری‌های متفاوتی وجود دارد. با این حال می‌توان مشکلات مشترکی در بین محیط‌های مختلف با شرایط گوناگون را در نظر گرفت که در ادامه به بررسی آن‌ها می‌پردازیم. اغلب دشواری‌های فرآیند بازشناسی به دلیل تغییرات ظاهری فرد از یک دوربین به دوربین دیگر رخ می‌دهد. این تغییرات ظاهری به دلایل زیر بروز می‌یابد [۱]:

- **زاویه دید و وضعیت‌های گوناگون نسبت به دوربین:** مسیر حرکت فرد نسبت به دوربین و فاصله فرد تا دوربین، باعث ثبت تصاویری با زاویه دیدهای مختلف و ظاهر گوناگون می‌شود.
- **همپوشانی:** تمام یا بخش‌هایی از تصویر یک فرد، ممکن است به دلیل انسداد با اشیاء دیگر در محیط، اشخاص دیگر، وسایل جانبی فرد (به عنوان مثال کیف و کوله‌پشتی) مخفی گردد. در چنین شرایطی الگوریتم‌های بخش‌بندی که قصد جداسازی ناحیه تصویر فرد از پس‌زمینه را دارند با مشکل مواجه می‌شوند و در نتیجه توصیف‌گرهایی که در مراحل بعد بر روی تصویر فرد اعمال می‌گردند اطلاعات نادرستی را بازنمایی خواهند نمود.

- **تغییرات نوری:** شرایط نوری می‌تواند به دلیل موقعیت قرارگیری دوربین یا شرایط محیطی متفاوت باشد. حتی در یک دوربین نیز ممکن است شرایط نوری در ساعات مختلف روز متفاوت باشد.

- **پاسخ‌های رنگی متفاوت:** دوربین‌های مختلف به دلیل تفاوت در ساخت، می‌توانند پاسخ‌های رنگی متفاوتی داشته باشند که در نهایت منجر به تفاوت در ظاهر افراد گردد.

- **تعداد افراد زیاد:** زمانی که تعداد افراد دیده شده در سیستم زیاد باشد، احتمال بروز شباهت بین پوشش و ظاهر افراد مختلف نیز بالا می‌رود. بنابراین توصیف‌گرهایی که سعی در بازنمایی افراد دارند ممکن است تفاوت‌های اندک در ظاهر افراد را در بازنمایی خود منعکس نکرده و در نهایت منجر به بازشناسی‌های اشتباهی گردند.

- **تفکیک‌پذیری پایین:** تصاویر ثبت‌شده از دوربین‌های نظارتی به دلیل ملاحظات در حجم ذخیره‌سازی داده‌ها، اغلب دارای تفکیک‌پذیری پایینی بوده یا خروجی ویدئویی آن‌ها به دلیل فشرده‌سازی حاوی نویز می‌باشد. همچنین فاصله زیاد از دوربین نیز منجر به ثبت تصاویری با تفکیک‌پذیری پایین می‌شود.

شکل ۱-۱ تغییرات در زاویه دید، شرایط نوری و حتی تغییر در پوشش فرد را در زمان‌های مختلف نشان می‌دهد. علی‌رغم اینکه پوشش فرد در طول زمان می‌تواند تغییر کند، عدم تغییر پوشش در بازه‌های زمانی کوتاه، فرضی منطقی در سیستم‌های نظارتی و بازشناسی است.



شکل ۱-۱: انواع تغییرات ظاهری یک فرد در سیستم نظارتی ویدئویی [۱]

۱-۳- انواع پیکربندی‌های مسئله بازشناسی افراد

از آنجایی که کاربرد ردیابی و بازشناسی افراد در سیستم‌های نظارتی، متشکل از فازهای متعددی است، ملاحظات و پیکربندی‌های گوناگونی را می‌توان در نظر گرفت که در زیر به چند دسته‌بندی اشاره خواهیم نمود.

سیستم تک‌دوربین در مقابل چنددوربین: در سیستم‌های نظارتی به لحاظ تعداد دوربین، دو دسته‌بندی وجود دارد. از آنجایی که هدف در سیستم‌های نظارت ویدئویی، پوشش کل محیط است، استفاده از چندین دوربین در موقعیت‌های مختلف مرسوم می‌باشد.

دوربین‌های با میدان دید مشترک در مقابل دوربین‌های غیرهمپوشان: در حالتی که سیستم نظارتی شامل چندین دوربین در محیط است، میدان دید این دوربین‌ها می‌تواند دارای اشتراک باشد. در چنین حالتی مسئله بازشناسی می‌تواند راحت‌تر حل گردد. به عبارتی در این حالت شخص هیچ‌گاه از میدان دید دوربین‌ها خارج نمی‌شود و به نوعی با مسئله ردیابی روبرو خواهیم بود. در مقابل دوربین‌های با میدان دید ناهمپوشان مطرح می‌گردد که این حالت، پیکربندی مرسوم‌تری است. درواقع

به دلیل کاهش هزینه‌ها سعی می‌شود بیشترین استفاده از دوربین‌ها به منظور پوشش محیط به عمل آید.

بانظارت و تعامل کاربر در مقابل کاملاً خودکار: استفاده از یک سیستم نظارتی هوشمند می‌تواند به صورت تعامل با کاربر یا به صورت کاملاً خودکار باشد. منظور از تعامل با کاربر این است که سیستم نظارتی اطلاعاتی به عنوان ورودی از کاربر دریافت کند. برای مثال کاربر در صورت نیاز، یک فرد خاص در تصویر را با ابزار نشانگر ماوس انتخاب کرده و سیستم فرد مذکور را ردیابی می‌کند. به عنوان مثالی دیگر می‌توان به سیستم‌های نظارتی اشاره نمود که در آن کاربر امکان ورود تصویر یک شخص به سیستم جهت بازشناسی را دارد. خروجی چنین سیستمی می‌تواند چندین پیشنهاد از شبیه‌ترین افراد رؤیت شده در سیستم باشد. در سیستم‌های کاملاً خودکار کلیه افراد در سیستم، ردیابی می‌شوند و خروجی سیستم به صورت سندی از موقعیت‌های افراد و زمان حضور آن‌ها در دوربین‌های مختلف، ذخیره می‌گردد.

دوربین‌های ثابت در مقابل دوربین‌های متحرک: مثالی از دوربین‌های متحرک زمانی است که، دوربین‌ها قابلیت حرکت دارند و توسط دستورات کاربر در صورت لزوم حرکت داده می‌شوند. در حالت دیگر دوربین‌ها می‌توانند به صورت گردشی حرکت کرده و منطقه‌ای را تحت پوشش خود قرار دهند. پیکربندی مرسوم در سیستم‌های نظارتی دوربین‌های ثابت می‌باشد.

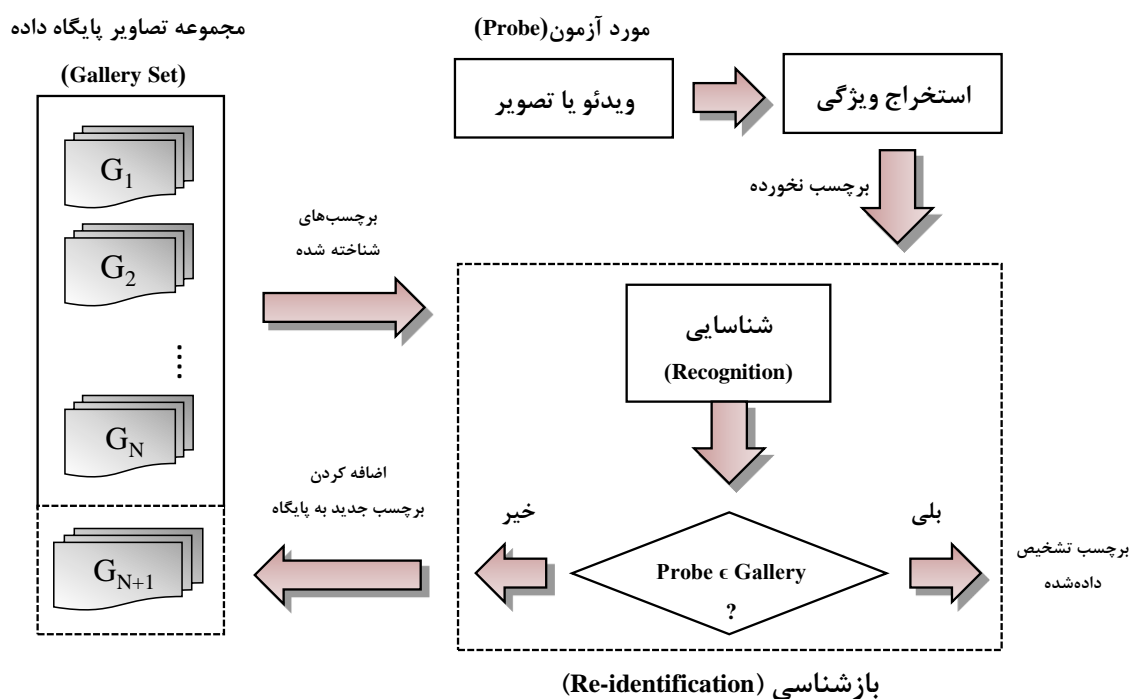
۱-۳-۱- مجموعه باز در مقابل مجموعه بسته

سیستم بازشناسی افراد مشابه یک سیستم شناسایی، شامل مجموعه تصاویری از افراد شناخته شده و دیده شده^۱ و تصویر فرد شناخته نشده^۲ می‌باشد. فرض کنید مجموعه تصاویر افراد شناخته شده به صورت $\mathbb{G} = \{g_1, g_2, \dots, g_N\}$ باشد و همچنین برچسب‌های تخصیص داده شده به این افراد به صورت

^۱ Gallery set

^۲ Probe

$id(\mathbb{G}) = \{id(g_1), id(g_2), \dots, id(g_N)\}$ تعریف گردد. در این رابطه تابع id شناسه آرگومان خود را تعیین می‌کند. فرض کنید مجموعه مورد آزمون $\mathbb{P} = \{p_1, p_2, \dots, p_M\}$ را داریم. می‌خواهیم شناسه این عناصر را مشخص کنیم. بدین منظور هر یک از عناصر مجموعه \mathbb{P} با هر یک از عناصر مجموعه \mathbb{G} مقایسه و شباهت آن‌ها محاسبه می‌شود. سپس میزان شباهت تعیین شده در مجموعه \mathbb{G} به ترتیب نزولی مرتب می‌گردد. بنابراین شناسه عنصر ناشناخته برابر با شناسه شبیه‌ترین عنصر در مجموعه \mathbb{G} خواهد بود. تصویری کلی از یک سیستم بازشناسی در شکل ۲-۱ نمایش داده شده است.



شکل ۲-۱: نمایی کلی از یک سیستم بازشناسی

بر این اساس دو رویکرد در سیستم بازشناسی افراد خواهیم داشت. در رویکرد اول که تحت عنوان مجموعه بسته^۱ می‌باشد، اعضای مجموعه \mathbb{P} زیرمجموعه‌ای از مجموعه گالری می‌باشند. در این سناریو هدف تعیین شناسه درست از مجموعه گالری است. در این حالت $id(\mathbb{P}) \subseteq id(\mathbb{G})$ بوده و برچسب

^۱ Close set

عضوی مانند p_j بصورت $id(p_j) = id(g_{i^*})$ محاسبه می‌گردد که در آن مقدار i^* از رابطه زیر تعیین می‌شود.

$$i^* = \arg \max_{i \in \{1, \dots, N\}} p(g_i | p_j) \quad (1-1)$$

که در آن $p(g_i | p_j)$ احتمال درستی $id(p_j) = id(g_i)$ می‌باشد که اغلب برای محاسبه این احتمال از معیار شباهت استفاده می‌گردد. در رویکرد دوم که مجموعه باز^۱ نامیده می‌شود، اعضای مجموعه P ممکن است در مجموعه گالری وجود نداشته باشد. در این سناریو ابتدا باید مشخص شود که تصویر آزمون عضوی از مجموعه گالری هست یا نه، تا در صورت وجود، شناسه آن به درستی انتخاب گردد. بر اساس فرمول‌بندی یادشده شرط دیگری که باید ارضا شود تا شناسه تعلق گیرد، به صورت زیر می‌باشد.

$$p(g_{i^*} | p_j) > \tau \quad (2-1)$$

در این رابطه مقدار آستانه τ تضمین می‌کند که p_j زیرمجموعه \mathbb{G} باشد. اگر این شرط برقرار نباشد تصویر جدید به مجموعه گالری اضافه شده و شناسه جدیدی به خود می‌گیرد. بطور معمول سیستم‌های نظارتی از سناریو مجموعه باز پیروی می‌کنند. در رویکرد مجموعه باز مجموعه گالری در طول زمان تغییر می‌یابد. در واقع در شروع کار سیستم در حالت مجموعه باز، مجموعه گالری تهی بوده و به مرور زمان بزرگ می‌گردد.

۴-۱- فازهای پیشین بازشناسی در سیستم‌های نظارت ویدئویی

به طور کلی قبل از انجام فرآیند بازشناسی، پیش‌پردازش‌هایی انجام می‌شود که در ادامه اشاره‌ای به آن‌ها خواهیم داشت. یکی از پردازش‌های اولیه در سیستم‌های نظارتی، ردیابی افراد است. منظور از ردیابی، دنبال کردن تصویر فرد در یک دوربین می‌باشد. بدیهی است که کلیه مشکلات و نکات مهم در

^۱ Open set

بحث ردیابی در این بخش از سیستم لحاظ می‌گردد. از آنجایی که در سیستم‌های نظارتی هدف مورد نظارت، انسان است می‌توان از این دانش پیشین چه در مرحله ردیابی و چه در مرحله بازشناسی استفاده نمود و کمک بسزایی به ساده‌سازی پردازش‌ها و بالا بردن دقت سیستم نمود. در این راستا یکی از روش‌های مرسوم و مورد استفاده، یافتن محدوده تصویر انسان در هر فریم از ویدئو خواهد بود. بدین ترتیب در صورتی که اشیاء دیگری در محیط نظارتی حرکت داشته باشند، در نظر گرفته نمی‌شوند.

۱-۴-۱- آشکارساز انسان

موقعیت‌یابی ناحیه انسان در تصویر از پردازش‌های کلیدی در بینایی ماشین است، و روش‌های بسیار متنوعی تاکنون ارائه شده‌اند و همگی سعی در ارائه روشی با دقت و سرعت بالاتر دارند. بسیاری از این روش‌ها به صورت بانظارت بر روی مجموعه عظیمی از تصاویر انسان، آموزش یافته و سپس در مرحله تست، درجه تعلق به کلاس انسان برای نواحی مختلف مشخص می‌شود. در یکی از مرسوم‌ترین این روش‌ها، با استفاده از اطلاعات توصیفگر^۱ HOG و کلاسه‌بند ماشین بردار پشتیبان (SVM)^۲ عمل موقعیت‌یابی انسان در تصویر انجام می‌شود [۲، ۳]. پیاده‌سازی‌های قابل دسترسی از این الگوریتم در ابزارهایی از قبیل OpenCV قابل استفاده می‌باشد.

۱-۴-۲- حذف پس‌زمینه

یکی از عملیات اصلی در پردازش ویدئو حذف پس‌زمینه از هر فریم ویدئویی است. زمانی که شخص در تصویر حرکت می‌کند، ناحیه مستطیلی استخراج شده شامل بخش‌هایی از پس‌زمینه است که با حرکت شخص تغییر می‌یابد. این تغییر در پس‌زمینه، منجر به کاهش کیفیت استخراج ویژگی از شخص و در نتیجه کاهش کیفیت بازشناسی خواهد شد. یکی از روش‌های مرسوم در حذف پس‌زمینه استفاده از مدل مخلوط گاوسی^۳ برای مدل‌سازی پس‌زمینه است [۴]. از محدودیت‌های استفاده از مدل مخلوط

^۱ Histogram of Oriented Gradients

^۲ Gaussian mixture model

^۳ Support Vector Machine

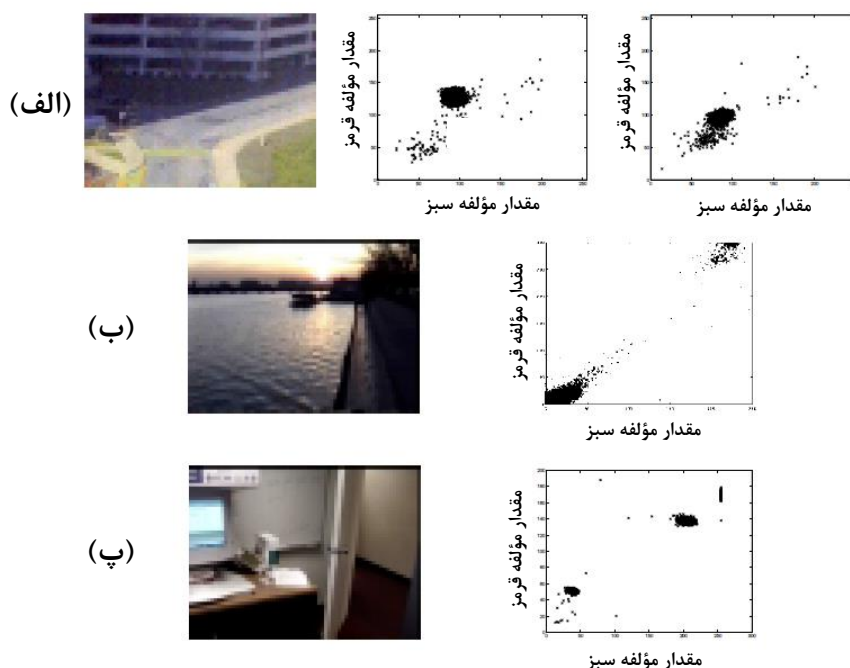
گاوسی، قید ثابت بودن دوربین می‌باشد. در کاربرد سیستم‌های نظارتی معمولاً این محدودیت وجود دارد و دوربین‌ها حرکتی ندارند. لذا می‌توان از این روش به‌منظور مدل‌سازی پس‌زمینه استفاده نمود. در دنباله‌ای از تصاویر ویدئویی برای یک پیکسل خاص (با سه مؤلفه رنگی)، مجموعه‌ای از مقادیر را خواهیم داشت. توزیع آماری این مقادیر برای یک پیکسل در کلی‌ترین حالت از یک مدل مخلوط گاوسی پیروی می‌کند. بنابراین در این روش به دنبال تخمین پارامترهای این مدل آماری در طول اجرای سیستم هستیم.

یکی از فرض‌های مورد استفاده در این روش، تعلق رنگی یک پیکسل در اغلب زمان‌ها به تصویر پس‌زمینه است. به عبارتی، پیکسل‌های یک ویدئو در اغلب اوقات، تصویر پس‌زمینه را نشان می‌دهند و در برخی موارد به‌صورت گذرا رنگ‌های شیء متحرک را به خود می‌گیرند. در نسخه‌های ساده‌تری از این الگوریتم از مدل‌های گاوسی تک‌مده^۱ استفاده می‌شود. از مشکلات روش تک‌مده، عدم توانایی مدل‌سازی پس‌زمینه‌هایی است که شامل چندین رنگ بوده که به‌صورت تناوبی تغییر می‌کنند. برای مثال برای یک پیکسل از فریم‌های پی‌درپی یک ویدئو که بر روی تصویر یک برگ درخت در ویدئو قرار دارد، ما دو رنگ خواهیم داشت. یک رنگ برای آسمان و یک رنگ برای برگ درخت که به دلیل حرکت برگ به تناوب بین این دو رنگ تغییر می‌کند. توزیع آماری رنگی این چنین پیکسلی دو مده خواهد بود که مدل مخلوط گاوسی می‌تواند آن را مدل نماید. از ویژگی‌های دیگر این روش تطبیق الگوریتم نسبت به تغییرات تدریجی پس‌زمینه است. شکل ۱-۳ حالات مذکور را با مثال نشان می‌دهد. از مشکلات دیگر این روش مدل‌سازی، می‌توان به عدم تطبیق با تغییرات ناگهانی نور محیط اشاره نمود. در چنین مواردی پیش‌زمینه‌های غیر معتبری نیز در تصویر تشخیص داده خواهد شد.

بدین ترتیب از ناحیه‌ای که به‌عنوان انسان تشخیص داده شده است پیکسل‌هایی که متعلق به مدل پس‌زمینه هستند را می‌توان حذف نمود. بنابراین ناحیه محصور در اطراف انسان خواهیم داشت. غیر

^۱ Unimodal distribution

از مدل‌سازی پس‌زمینه، روش‌های متعدد دیگری به‌منظور حذف پس‌زمینه معرفی شده است [۵، ۶]. گروهی از این روش‌ها با استفاده از بخش‌بندی تصویر عمل جداسازی ناحیه شخص از پس‌زمینه را انجام می‌دهند. برخی دیگر از روش‌ها با تمرکز بر روی ناحیه مرکزی تصویر که احتمال حضور پیکسل‌های متعلق به شخص بیشتر است عمل جداسازی را انجام می‌دهند. در [۵] با استفاده از مفهوم جدیدی با نام $Stel^1$ سعی در جداسازی پیش‌زمینه از پس‌زمینه دارد. در واقع یک $Stel$ مؤلفه‌ای ساختاری برای یک کلاس از اشیاء می‌باشد که علاوه بر تکرار در تصاویر کلاس مذکور، درون یک تصویر هم رؤیت می‌گردد. در واقع در این روش اجزایی از تصویر که خاصیت خودمتمشابهی دارند به‌عنوان بخش‌های بامعنی و مهم تصویر استخراج می‌گردد. در کاربرد کنونی دو مؤلفه یکی برای پیش‌زمینه و دیگری برای پس‌زمینه در نظر گرفته شده است.



شکل ۱-۳: تغییرات مقادیر رنگی برای یک پیکسل از سه دنباله ویدئویی در طول زمان، تصویر الف، مقدار یک پیکسل را در بازه‌های زمانی مختلف از روز نشان می‌دهد که تغییرات تدریجی موقعیت مَد اصلی توزیع را شاهدیم. تصویر ب، مقدار پیکسلی در محل آب را می‌بینیم که توزیعی دو مده می‌باشد. تصویر پ، توزیع رنگی پیکسلی واقع بر روی ناحیه مانیتور است که به‌صورت تناوبی بین دو مقدار تغییر می‌کند که به‌صورت دو مده قابل‌مشاهده است.

¹ Structure element (Stel)

در [۶] با استفاده از بخش‌بندی حوزه آبخیز^۱، تصویر ناحیه یک فرد بخش‌بندی می‌گردد. سپس لبه‌هایی که در مرز بین این نواحی قرار می‌گیرند را به‌عنوان لبه‌های بااهمیت در نظر گرفته به‌طوری که ادعا می‌شود این لبه‌ها مابین نواحی مختلف رنگی لباس فرد واقع شده و در نتیجه چین‌وچروک‌های درون یک ناحیه رنگی منظور نمی‌گردد. پس از استخراج نواحی رنگی، گرافی بر اساس همسایگی مکانی و زمانی این نواحی، ساخته می‌شود. منظور از همسایگی زمانی، همسایگی در فریم‌های متوالی است. وزن لبه دو ناحیه متوالی نیز بر اساس میزان اشتراک پیکسلی در مرز این دو ناحیه تعیین می‌گردد. در مرحله بعد به دو منظور گراف بدست آمده خوشه‌بندی می‌گردد. اول، بخش‌بندی تصویر و دوم، استخراج نواحی که بیشترین ثبات را در تصویر دارند. روش‌هایی که وابسته به اطلاعاتی زمانی هستند در مواردی که از افراد تنها یک تصویر در اختیار داریم، کاربردی ندارند. در رویکردی دیگر می‌توان ابتدا، به کمک مدل‌سازی پس‌زمینه، اشیاء متحرک را استخراج نمود. سپس برای هر شیء، با استفاده از الگوریتم‌های تشخیص انسان، تعلق آن به کلاس انسان و یا غیرانسان را تعیین کرد. درعین حال به کمک اطلاعات مدل پس‌زمینه، می‌توان ناحیه پس‌زمینه را نیز حذف نمود.

۱-۴-۳- حذف سایه

یکی از مشکلاتی که در ردیابی ممکن است با آن روبرو شویم حذف نگریدین قسمت‌های سایه شیء متحرک می‌باشد. به عبارتی روش‌های جداسازی پیش‌زمینه از پس‌زمینه به‌اشتباه نواحی سایه را که با فرد حرکت می‌کند به‌عنوان شیء متحرک در نظر می‌گیرند. لذا روش‌هایی برای حذف سایه ارائه گردیده است. در برخی از این روش‌ها همبستگی بین مقادیر پیکسل‌های پیش‌زمینه با مقادیر رنگی پیکسل مدل پس‌زمینه محاسبه، در صورتی که این فاصله از یک مقدار آستانه کمتر باشد به‌عنوان سایه در نظر گرفته می‌شود [۷]. ایده این روش در این است که نواحی سایه همان نواحی پس‌زمینه هست با این تفاوت که در عددی ضرب گردیده است. در برخی از روش‌ها مقدار پیکسل پس‌زمینه با دو مقدار

^۱ Watershed segmentation

آستانه سنجیده شده و بدین ترتیب ناحیه سایه شناسایی شده که در مرحله بعد در بروزرسانی مدل پس‌زمینه دخالت داده نشود. در برخی از روش‌ها از کانال‌های دیگر رنگی برای تشخیص سایه استفاده می‌گردد. می‌توان چنین فرضی را معتبر دانست که مؤلفه H در مدل رنگی HSV، برای پیکسلی متعلق به سایه، مشابه با مؤلفه H در مدل پس‌زمینه برای آن نقطه است. به عبارتی تفاوت پیکسل سایه و مدل پس‌زمینه در مقدار مؤلفه V آن‌ها است [۷].

۱-۵- هدف رساله

بررسی‌های انجام شده در این رساله نشان می‌دهد که تغییرات زاویه قرارگیری شخص در مقابل دوربین‌های نظارتی، منجر به بروز اختلالاتی در الگوریتم‌های بازشناسی می‌شوند. به عبارتی وضعیت سه‌بعدی قرارگیری شخص در مقابل دوربین باعث می‌شود که بخش‌هایی از پوشش فرد در تصاویر آشکار یا نهان گردند و در نهایت ظاهری متفاوت از یک شخص در زوایای مختلف ایجاد شود. در سیستم‌هایی که ناظر انسانی عمل نظارت را انجام می‌دهد با استفاده از دانش پیشین نسبت به پیکره سه‌بعدی انسان، عمل بازشناسی صورت می‌گیرد. در واقع ذهن انسان تلاش می‌کند تطبیق و یکسان‌سازی بین دو تصویر از زوایای مختلف انجام داده، و سپس عمل بازشناسی را انجام دهد. در واقع هدف از این رساله ارائه راهکاری برای مدل‌سازی این الگوریتم تطبیق و یکسان‌سازی است.

سیستم پیشنهادی در این رساله می‌تواند با بسیاری از الگوریتم‌های بازشناسی تلفیق و مورد استفاده قرار گیرد. بدین ترتیب زمانی که دو تصویر از زوایای مختلف به سیستم وارد می‌گردد ابتدا زاویه شخص در تصاویر تخمین زده می‌شود. سپس با استفاده از اطلاعات زاویه‌ای و ناحیه قرارگیری سر در تصویر، نواحی خاصی که تحت چرخش زاویه‌ای می‌تواند مخفی یا آشکار شود و الگوریتم‌های بازشناسی را با اخلال مواجه کند، استخراج می‌گردد. سپس نواحی مذکور وزن‌دهی یا الگوی رنگی و بافتی مشابه با نواحی مجاور به خود می‌گیرند. از این فرآیند تحت عنوان یکسان‌سازی دو تصویر یاد می‌شود. بنابراین

رویکرد پیشنهادی می‌تواند با تغییرات اندکی متناسب با الگوریتم‌های دیگرِ بازشناسی، مورد استفاده قرار گیرد.

۱-۶- مفروضات مورد نظر برای مسئله بازشناسی انسان

همانطور که در این فصل بدان اشاره شد مسئله بازشناسی انسان می‌تواند در شرایط گوناگونی مورد بررسی قرار گیرد. مفروضاتی که در این رساله برای حل مسئله بازشناسی انسان در نظر گرفته شده عبارتند از:

- دوربین‌ها در محیط ثابت هستند.
- دوربین‌ها دارای میدان دید ناهمپوشان می‌باشند.
- سیستم در حالت مجموعه بسته مورد ارزیابی قرار می‌گیرد.
- همانطور که در فصل بعد خواهیم دید، اغلب روش‌های بازشناسی ارزیابی را در چنین حالتی انجام می‌دهند.
- بازشناسی کوتاه مدت مورد نظر می‌باشد.
- به عبارتی پوشش افراد از یک دوربین به دوربین دیگر تغییر چشمگیری نمی‌کند.
- افراد در سیستم نظارتی در حال تردد هستند.
- در نتیجه حالت کلی شخص به صورت ایستاده می‌باشد و حالاتی نظیر نشسته نخواهیم داشت.
- از آنجایی که چالش اصلی و مورد توجه در این رساله مسئله چرخش زاویه‌ای است، داده‌های مورد استفاده باید شامل این چالش باشند.
- در این راستا از میان مجموعه داده‌های موجود، مجموعه داده‌ای به منظور ارزیابی مورد استفاده قرار گرفته که شامل تغییرات زاویه‌ای شدید می‌باشد. در صورتی که داده‌های

مورد استفاده فاقد چرخش زاویه‌ای باشند رویکرد پیشنهادی همانند روش‌های استاندارد عمل خواهد نمود.

- داده‌ها می‌توانند شامل تغییرات نوری، تغییرات رنگی و نویز باشند.
 - به عبارتی رویکرد پیشنهادی تنها زمانی تاثیرگذار خواهد بود که تغییرات زاویه‌ای شدید داشته باشیم. در بقیه چالش‌ها نظیر تغییرات نوری، تغییرات رنگی و نویز، رویکرد پیشنهادی متناسب و مشابه با نوع الگوریتم بازشناسی انتخابی رفتار خواهد نمود.
- مرحله آشکارسازی انسان با دقت نسبتاً خوبی انجام شده است.
 - همانطور که اشاره شد، یکی از پیش‌نیازهای مسئله بازشناسی آشکارسازی ناحیه شخص در تصویر می‌باشد. لذا فرض می‌کنیم که ناحیه مذکور با دقت نسبتاً مطلوبی استخراج شده است. مجموعه داده‌های استاندارد موجود نیز شامل این فرض می‌باشند.

۱-۷- مجموعه داده‌ها به منظور ارزیابی

به منظور ارزیابی الگوریتم‌های بازشناسی مجموعه داده‌های متنوعی در دسترس می‌باشد که هر کدام چالش‌های خاصی را در مساله بازشناسی مطرح می‌نمایند. تعدادی از این مجموعه داده‌ها عبارتند از:

- **ViPER**: این مجموعه داده‌ای شامل تصویر ۶۳۲ فرد می‌باشد. برای هر فرد دو تصویر مربوط به دو دوربین مختلف تهیه شده است. دو تصویر مربوط به یک فرد از زوایای مختلف می‌باشند. شرایط نوری در دو دوربین متنوع می‌باشد. تصاویر هر فرد به مستطیلی محدود گشته و در اندازه ۱۲۸ در ۴۸ پیکسل است. این مجموعه داده سخت‌ترین مجموعه داده‌ای است که در بازشناسی مورد استفاده قرار گرفته است [۸].

- **ETHZ**: این مجموعه داده شامل تصاویر ثبت شده بوسیله یک دوربین متحرک می‌باشد. از هر شخص چندین تصویر موجود است. تصاویر در اندازه‌های مختلف و شامل همپوشانی نیز می‌باشند [۹].

- **i-LIDS**: این مجموعه داده شامل ۴۷۶ تصویر از ۱۱۹ فرد می‌باشد که از دو دوربین با میدان دید ناهمپوشان بدست آمده است. بطور میانگین برای هر فرد ۴ تصویر موجود می‌باشد. تصاویر شامل شرایط متفاوت نوری و شامل همپوشانی نیز می‌باشند. تصاویر در اندازه ۱۲۸ در ۶۴ تهیه گردیده است [۱۰].

- **CAVIAR4REID**: این مجموعه داده‌ای شامل تصاویری از دو دوربین با میدان دید مشترک می‌باشد که از محیط یک فروشگاه تصویربرداری شده است. تصاویر موجود مربوطه به ۷۲ فرد مختلف می‌باشد که انواع شرایط مشکل‌ساز از قبیل شرایط نوری متفاوت، حالت و پوزیشن و همپوشانی را شامل می‌شود [۱۱].

V-47: این مجموعه داده‌ای شامل ویدئوهایی از ۴۷ شخص از دو دوربین مختلف در فضای داخلی گرفته شده است. برای هر دوربین دو ویدئو در دو زاویه تهیه شده که این دو زاویه دو جهت راه رفتن افراد می‌باشد. تغییرات نوری و اندازه و شلوغی در این مجموعه داده‌ای زیاد نیست [۱۲].

۸-۱- نوآوری‌ها و دستاوردهای رساله

به‌طور خلاصه دستاوردها و نوآوری‌های این رساله عبارتند از:

۱- در این رساله ابتدا مروری بر روش‌های بازشناسی انجام گردید. در این راستا دسته‌بندی جدیدی از روش‌های بازشناسی انسان در سیستم‌های نظارت ویدئویی پیشنهاد و نتیجه این مطالعات در مقاله [الف] گزارش شده است.

[الف] علی سبّتی، حمید حسن پور، "بازشناسی انسان در سیستم‌های نظارت ویدئویی"، مجله محاسبات نرم، دانشگاه کاشان، دوره سوم، شماره اول، صفحات ۶۲-۸۱، بهار ۱۳۹۳.

۲- در این رساله یکی از قدم‌های اصلی و پیش‌نیاز، استخراج زاویه قرارگیری شخص در دوربین می‌باشد. به دلیل کیفیت پایین تصاویر سیستم‌های نظارتی، این تشخیص با دشواری‌های بسیاری همراه است. در این راستا دو روش برای تخمین زاویه ارائه گردید:

- در روش اول با استفاده از اطلاعات کانتوری شخص و تطبیق با ابرکانتورهای استخراج‌شده از یک مدل سه‌بعدی انسان، در ساختاری سلسله‌مراتبی، زاویه قرارگیری شخص تخمین زده می‌شود. نتایج این پژوهش در مقاله [ب] گزارش شده است.
- در روش دوم با استفاده از ویژگی هیستوگرام گرادیان‌های جهتی در تصویر برای دو نوع ناحیه (ایستا و پویا) و درنهایت ترکیب کلاسه‌بندهای رگرسیون لجستیک که بر روی هر ناحیه آموزش‌یافته، عمل تشخیص زاویه با کارآمدی بالا انجام می‌شود. نتایج این پژوهش در [A] گزارش شده است.

[ب] علی سبّتی، حمید حسن پور، "تشخیص زاویه قرارگیری شخص در تصویر با استفاده از اطلاعات کانتوری"، نشریه مهندسی برق و مهندسی کامپیوتر ایران، ب- مهندسی کامپیوتر، سال ۱۴، شماره ۴، صفحات ۳۲۲-۳۲۲، زمستان ۱۳۹۵.

[A] Ali Sebti, Hamid Hassanpour. *Body Orientation Estimation with the Ensemble of Logistic Regression Classifiers Multimedia Tools and Applications*, Springer, 2016, doi: 10.1007/s11042-016-4129-0.

۳- به منظور استخراج نواحی خاصی از تصویر که تحت چرخش زاویه‌ای، منجر به تغییر ظاهر شخص می‌شوند، ناحیه سر شخص به‌عنوان یک ناحیه کلیدی در تصویر استخراج شده است. این موقعیت‌یابی با استفاده از کلاسه‌بندی بر پایه شبکه عصبی کانولوشن انجام شده است. به منظور بهبود کارایی الگوریتم، کلاسه‌بند دیگری برای ناحیه سر و شانه نیز آموزش داده شده و مورد

استفاده قرار گرفته است. در نهایت ترکیب خروجی حاصل از این دو کلاسه‌بند، منجر به موقعیت‌یابی دقیق‌تری از ناحیه سر شده است. به‌منظور کاهش پیچیدگی محاسباتی، دو کلاسه‌بند ذکر شده در بالا، بر روی تصاویری با تفکیک‌پذیری یک‌چهارم آموزش داده شده و تست می‌گردند و در مرحله ترکیب به تفکیک‌پذیری اولیه نگاشت می‌شوند.

۴- با استفاده از اطلاعات زاویه‌ای و موقعیت سر در تصویر، نواحی خاصی از پوشش فرد که در زیر ناحیه سر واقع شده است بر مبنای خوشه‌بندی نواحی رنگی استخراج می‌گردد. نواحی استخراج شده سپس در صورت لزوم در مرحله تطبیق و بازشناسی، مورد وزن‌دهی یا تغییر بر اساس نواحی مجاور قرار می‌گیرند. در نهایت با اعمال تغییرات مذکور در تصاویر به‌عنوان پیش‌پردازش، بهبود در فرآیند بازشناسی را خواهیم داشت. در یکی از پژوهش‌های صورت گرفته با فرض دانستن زاویه شخص در تصویر و البته استخراج نواحی خاص به‌صورت غیرآگاهانه نیز بهبود در بازشناسی را شاهد هستیم که نتایج این پژوهش در مقاله [پ] آمده است. همچنین در پژوهشی دیگر رویکرد نهایی و پیشنهادی در این رساله گزارش شده است [ت].

[پ] علی سبطی، حمید حسن‌پور، "بهبود الگوریتم SDALF در بازشناسی انسان با بهره‌گیری از اطلاعات زاویه شخص"، بیست و سومین کنفرانس مهندسی برق ایران، دانشگاه صنعتی شریف، تهران، ایران، ۱۳۹۴.

[ت] علی سبطی، حمید حسن‌پور، "بازشناسی انسان در سیستم‌های نظارت ویدئویی با بهره‌گیری از اطلاعات زاویه‌ای"، مجله مهندسی برق دانشگاه تبریز، تابستان ۱۳۹۶. (تحت داوری)

۹-۱- ساختار رساله

فصل‌بندی و ساختار این رساله به‌صورت زیر می‌باشد. در فصل دوم روش‌های موجود در بازشناسی، دسته‌بندی و مورد بررسی قرار خواهند گرفت. در فصل سوم رویکرد پیشنهادی در این رساله تشریح می‌گردد. در این فصل جزئیات رویکرد پیشنهادی شامل تشخیص زاویه، موقعیت‌یابی ناحیه سر و یکسان‌سازی تصاویر مورد بررسی قرار خواهد گرفت. در فصل چهارم رویکرد پیشنهادی را مورد ارزیابی

قرار داده و نتایج آزمایش‌های انجام شده و مقایسات صورت گرفته گزارش می‌گردد. در نهایت در فصل پنجم کارهای انجام شده جمع‌بندی و پیشنهاداتی برای کارهای آتی ارائه شده است.

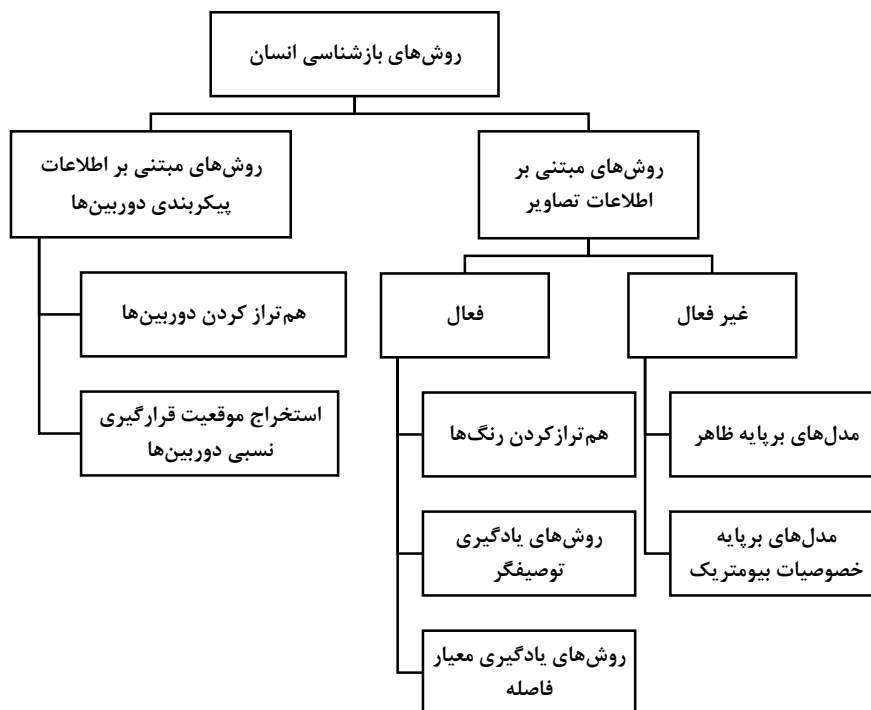
فصل ۲. مروری بر کارهای پیشین

در سیستم‌های بازشناسی یک یا چند تصویر از یک فرد وارد سیستم می‌گردد. در گام بعد، سیستم شباهت تصویر آزمون را با تصاویر مجموعه گالری محاسبه کرده و لیست مرتب‌شده‌ای براساس مقادیر شباهت ایجاد می‌کند. در حالت سیستم بسته، برچسب تصویر آزمون، برابر برچسب بالاترین عنصر در لیست مرتب‌شده خواهد بود. این شباهت بر اساس شباهت ظاهری محاسبه می‌شود. از منظری دیگر یک سیستم بازشناسی می‌تواند در دو حالت کوتاه‌مدت و بلندمدت تعریف گردد. در حالت کوتاه‌مدت، فرض پایداری در ویژگی‌های ظاهری و عدم تغییر پوشش فرد وجود دارد. اکثر کارهای انجام‌شده تاکنون، این فرض را اساس کار خود قرار می‌دهند [۱]. دلیل معتبر بودن این فرض در هدف نهایی سیستم‌های نظارت ویدئویی نهفته است.

بطور معمول هدف در سیستم‌های نظارت ویدئویی، بررسی رفتار غیرعادی و جابجایی‌های افراد است. در حالت بلندمدت، بازشناسی بر اساس خصوصیات غیر از ظاهر صورت می‌گیرد. از جمله خصوصیات مورد استفاده در این سیستم‌ها، ویژگی‌های بیومتریک است که از پایداری لازم در طول زمان برخوردار هستند. از جمله محدودیت‌های سیستم‌های بیومتریک، نیاز به دوربین‌های قوی جهت تصویربرداری است. به‌طور کلی از آنجایی که هدف در اغلب سیستم‌های نظارتی احراز هویت کاربر نیست، بازشناسی بلندمدت کمتر مورد نیاز می‌باشد.

روش‌های مختلفی برپایه مدل‌سازی ظاهری ارائه شده‌اند که همگی سعی در استفاده از ویژگی‌هایی دارند که از طرفی بازگوکننده‌ی خوبی برای ظاهر فرد باشند و از طرف دیگر متمایزکننده خوبی برای افراد مختلف باشند. برخی دیگر از روش‌ها، بر پایه یادگیری ارائه شده‌اند. فرآیند یادگیری می‌تواند در قسمت‌های مختلفی نظیر تعیین فاصله یا شباهت، به‌روزرسانی مدل ظاهری فرد، تعیین تابع انتقال رنگی بین دوربین‌های نصب‌شده، تخمین موقعیت قرارگیری دوربین‌ها و مدل راه رفتن فرد مورد استفاده قرار گیرد.

بر اساس مقالات [۱، ۱۳] روش‌های مختلف بازشناسی انسان در سیستم‌های نظارتی را می‌توان به صورت نمودار شکل ۲-۱ دسته‌بندی نمود.



شکل ۲-۱: دسته‌بندی روش‌های بازشناسی افراد در سیستم‌های نظارتی [۱]

۲-۱- روش‌های مبتنی بر اطلاعات تصاویر

در این دسته از روش‌ها تنها اطلاعات بصری افراد است که در فرآیندهای مختلف بازشناسی مورد استفاده قرار می‌گیرد و اطلاعات بیرونی نظیر موقعیت دوربین‌ها و هم‌ترازی^۱ آن‌ها مورد استفاده نمی‌باشد. اکثر مقالات و کارهای انجام شده در این دسته می‌باشند. این گروه خود به روش‌های فعال و غیرفعال دسته‌بندی می‌شود. همچنین اطلاعات مورد استفاده می‌تواند به صورت تک‌تصویره^۲ یا چندتصویره^۳ باشد. بدین ترتیب که در حالت تک‌تصویره از تصویر آزمون تنها یک تصویر داریم و در حالت چندتصویره دنباله‌ای از تصاویر در اختیار است که این دنباله از تصاویر می‌تواند از طریق

^۱ Calibration

^۳ Multi-Shot

^۲ Single-Shot

ردیابی در مرحله قبلی بدست آید. همچنین می‌توان چندین تصویر را در قالب یک مدل خلاصه نمود و در نهایت، تطبیق بین مدل‌ها صورت گیرد.

۲-۱-۱- روش‌های غیرفعال

در این دسته، هدف طراحی توصیفگری برای انسان است که در عین بازنمایی مطلوب، تمایزپذیری خوبی نیز بین انسان‌های مختلف ایجاد نماید. منظور از غیرفعال بودن، عدم استفاده از یادگیری به صورت بانظارت یا بی‌نظارت، در مرحله استخراج توصیفگر یا در مرحله تطبیق می‌باشد. از آنجایی که روش‌های این دسته نیاز به مرحله یادگیری ندارند لذا بیشتر مورد توجه هستند. همچنین زمان مورد نیاز فرآیند راه‌اندازی چنین سیستم‌هایی کمتر بوده و دارای پیچیدگی محاسباتی کمتری هستند.

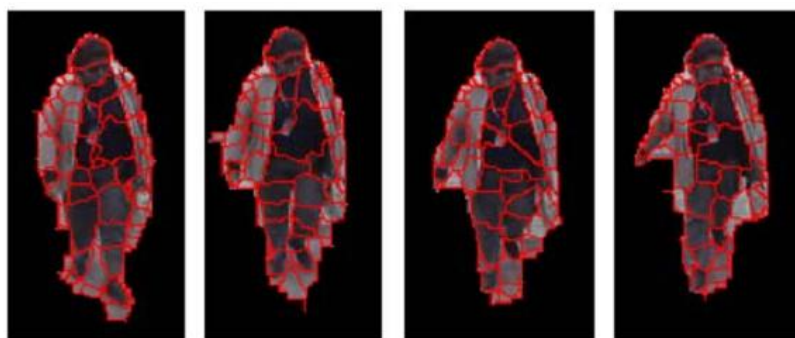
در دسته‌بندی ارائه شده شکل ۱-۲ مدل‌های برپایه خصوصیات بیومتریکی نیز در این دسته قرار داده شده‌اند. زمانی که هدف، بازشناسی در بلندمدت باشد تنها خصوصیات بیومتریکی است که دستخوش تغییرات کمتری می‌شوند و می‌توانند وجه تمایزی بین افراد مختلف ایجاد نمایند. از جمله خصوصیت‌های بیومتریکی مورد توجه، می‌توان به چهره و نحوه راه رفتن اشاره نمود. به دلیل محدودیت‌های سخت‌افزاری سیستم‌های نظارتی، در عمل از این گروه از روش‌ها کمتر استفاده می‌شود. برای مثال به دلیل تفکیک‌پذیری پایین دوربین‌های نصب‌شده در محیط، چهره افراد به خوبی قابل تشخیص نیست. همچنین امکان دارد زاویه شخص طوری باشد که چهره فرد آشکار نباشد. این مشکلات برای تشخیص و مدل‌سازی نحوه راه رفتن یک فرد نیز وجود دارد. محدودیت دیگر دوربین‌ها که تشخیص راه رفتن را مشکل می‌سازد نرخ تصویربرداری پایین دوربین‌ها است که در چنین حالتی مدل راه رفتن فرد به خوبی استخراج نمی‌شود.

۲-۱-۱-۱- مدل ظاهری مکانی-زمانی

یکی از مشکلات در مسئله بازشناسی انسان این است که در طول حرکت و یا بر اساس موقعیتی که فرد در آن قرار دارد ممکن است لباس فرد دستخوش حرکتی گردد. به عبارتی، چین و چروک‌هایی ممکن است بر روی لباس فرد ایجاد شود که کیفیت بازشناسی را تحت تأثیر قرار دهد. در مقاله [۶] سعی شده بر این مشکل غلبه شود. در این مقاله، الگوریتم بخش‌بندی جدیدی تحت عنوان بخش‌بندی مکانی-زمانی ارائه شده است.

این الگوریتم بخش‌بندی سعی دارد تصویر فرد را به بخش‌هایی مرتبط با قسمت‌های رنگی مختلف لباس فرد تقسیم نماید. بدین ترتیب ابتدا لبه‌های بااهمیت تصویر را استخراج می‌کند. این لبه‌ها در واقع مرز بین نواحی مذکور را تشکیل می‌دهند. بدین ترتیب لبه‌های بی‌اهمیت، نظیر چین و چروک‌های لباس نادیده گرفته می‌شود. در این مقاله ابتدا ناحیه بدست آمده برای فرد به دو بخش پیش‌زمینه و پس‌زمینه تقسیم می‌شود. این تقسیم‌بندی بر اساس میزان فراوانی مقدار رنگی پیکسل‌ها در یک پنجره زمانی خاص تعیین می‌گردد.

سپس ناحیه پیش‌زمینه برای هر یک از تصاویر موجود در پنجره زمانی مذکور، قطعه‌بندی می‌گردد. این قطعه‌بندی به کمک الگوریتم قطعه‌بندی حوزه آبریز انجام می‌گیرد. نتیجه این قطعه‌بندی نواحی بهم پیوسته $\mathbb{R} = \{r_i^t\}$ می‌باشد که در آن i اندیس ناحیه در تصویر t ام می‌باشد. بر این اساس گراف $G = \{\mathbb{V}, \mathbb{E}\}$ با مجموعه رئوس $\mathbb{V} = \{v_i^t\}$ و مجموعه یال $\mathbb{E} = \{e_{i,i'}^{t,t'}\}$ تعریف می‌گردد. در این تعریف رئوس گراف (v_i^t) متناظر نواحی r_i^t و یال‌ها $(e_{i,i'}^{t,t'})$ نیز دو راس v_i^t و $v_{i'}^{t'}$ را بهم متصل می‌کند. پس از ساخت چنین گرافی از الگوریتم‌های افزایش‌بندی یا خوشه‌بندی گراف استفاده شده و بخش‌بندی بر روی نواحی مذکور صورت می‌گیرد. شکل ۲-۲ و شکل ۳-۲ نتیجه قطعه‌بندی تصویر را در طول زمان نشان می‌دهند.



شکل ۲-۲: نتیجه قطعه‌بندی به روش حوزه آبریز در ناحیه پیش‌زمینه برای چند فریم متوالی [۶]



شکل ۲-۳: استخراج لبه‌های با اهمیت در تصویر با استفاده از الگوریتم ارائه شده در [۶]. تصویر الف: تصویر ورودی، تصویر ب: عکس فرکانسی تصویر پ: نتیجه خوشه‌بندی گراف و تصویر ت: مقدار رنگی میانه برای هر بخش از خروجی خوشه‌بندی.

در این بازنمایی گرافی، دو نوع لبه خواهیم داشت: لبه مکانی و لبه زمانی. لبه مکانی مرز مشترک بین نواحی است (r_i^t) و لبه زمانی میزان شباهت نواحی را در طول زمان مشخص می‌کند (r_i^{t+1}) . ناحیه r_i^{t+1} شبیه‌ترین ناحیه متناظر با ناحیه r_i^t به لحاظ بافتی و رنگی است. در این مقاله مفهوم تصویر فرکانسی با تعریف ریاضی زیر معرفی شده است.

$$F_{N,t}(x, y) = \sum_{k=0}^N H(I_t(x, y) - I_{t+k}(x, y)) \quad (1-2)$$

در این رابطه F تصویر فرکانسی و $I_t(x, y)$ مقدار رنگ پیکسل در موقعیت (x, y) و لحظه t و N بازه زمانی برای محاسبه تصویر فرکانسی می‌باشد که در این مقاله مقدار ۱۰ در نظر گرفته شده است. و $H(\cdot)$ به صورت زیر تعریف می‌گردد:

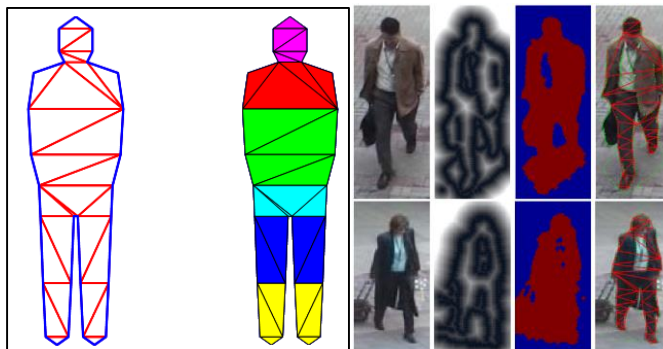
$$H(z) = \begin{cases} 1 & \text{if } |z| < \delta \\ 0 & \text{otherwise} \end{cases} \quad (2-2)$$

بدین ترتیب برای نواحی که مقدار تابع F بزرگتر است به معنی ثبات بیشتر آن نقطه از تصویر در طول زمان می‌باشد. برای تعیین همسایگی دو ناحیه در فریم‌های متوالی مقدار مجموع پیکسل‌های اشتراک دو ناحیه محاسبه و ناحیه‌ای که بزرگ‌ترین مجموع عددی را دارد به‌عنوان یال زمانی انتخاب می‌گردد. در نهایت وزن بین نواحی (w) براساس تفاضل مقدار میانه رنگی (M) در هر ناحیه محاسبه می‌گردد:

$$w_{i,i'}^{t,t} = |M(i, t) - M(i', t)| \quad (3-2)$$

$$w_{i,i'}^{t,t+1} = \frac{1}{3} |M(i, t) - M(i', t+1)|$$

ضریب یک‌سوم اجازه تغییرات بیشتری را در زمان فراهم می‌کند. در این مقاله در نهایت دو الگوریتم مورد مقایسه قرار گرفته است. الگوریتم اول با استفاده از عملگر استخراج نقاط مورد توجه هسین^۱ می‌باشد. بعد از استخراج نقاط مذکور عملیات تطبیق بین این نقاط انجام می‌گیرد. الگوریتم دوم سعی در انطباق مدلی به تصویر انسان و لبه‌های بااهمیت بدست آمده در قسمت بخش‌بندی می‌باشد. این انطباق از طریق بهینه‌سازی یک تابع هزینه به کمک برنامه‌نویسی پویا انجام شده است. مدل مذکور یک مدل مثلثی به صورت شکل ۴-۲ می‌باشد.



شکل ۴-۲: تصویر راست مدل مثلثی، تصویر چپ: دو نمونه انطباق به کمک لبه‌های بااهمیت و قسمت ناحیه پیش‌زمینه بدست آمده از الگوریتم بخش‌بندی مبتنی بر پارتیشن‌بندی گراف [۶]

^۱ Hessian

در نهایت مقایسه‌ای بین سه الگوریتم که به ترتیب بر اساس نقاط هسین، انطباق مدل مثلثی و روش پایه‌ای که تنها کل ناحیه فرد را مورد توجه قرار می‌دهد، انجام می‌گیرد. نتایج بدست آمده، برتری قابل توجهی را در الگوریتم دوم نسبت به دو الگوریتم دیگر نشان می‌دهد. در این مقاله ویژگی مورد استفاده در تطبیق نواحی هیستوگرام مؤلفه توسعه‌یافته H در مدل رنگی HSV می‌باشد. این توسعه از مؤلفه H، نسبت به تغییرات نوری و تغییرات گاما مقاوم می‌باشد.

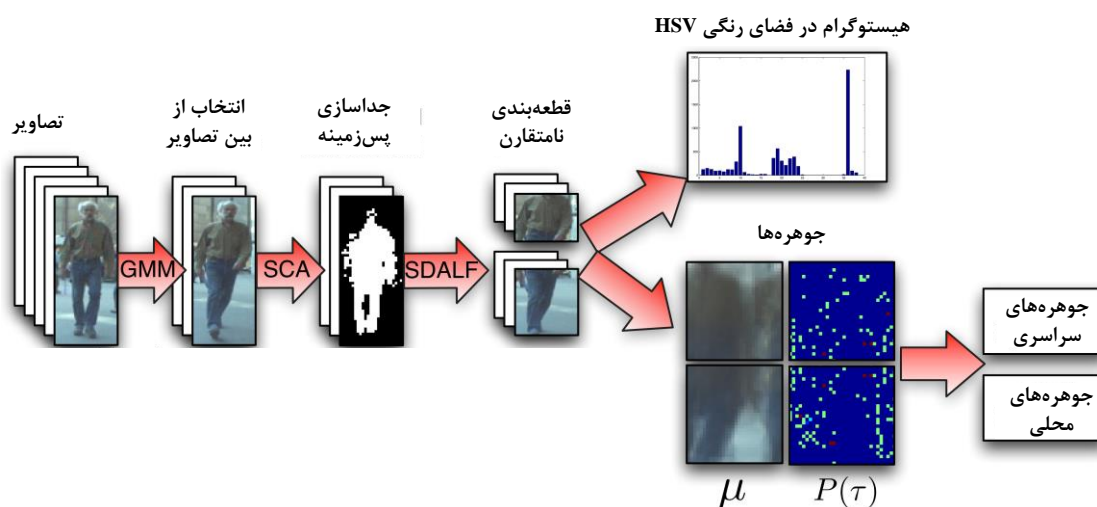
۲-۱-۱-۲- بازشناسی چندتصویره انسان به کمک توصیفگر HPE

در مقاله [۱۴] از یک روش ترکیبی برای ارائه توصیفی از چندین تصویر به‌طور همزمان استفاده شده است. قسمت اصلی این الگوریتم از دو بخش هیستوگرام و جوهره^۱ تشکیل می‌شود (HPE^۲). در این روش ابتدا دنباله‌ای از تصاویر یک فرد را که توسط الگوریتم ردیابی حاصل گردیده در سه خوشه جدا خوشه‌بندی می‌کند. مراکز خوشه‌ها به‌عنوان سه تصویر در گام‌های بعدی الگوریتم مورد استفاده قرار می‌گیرد. استفاده از خوشه‌بندی باعث حذف داده‌های نویزی و حاوی همپوشانی می‌شود. برای سه تصویر بدست آمده ناحیه پیش‌زمینه به کمک الگوریتم ارائه شده در [۵] جداسازی می‌گردد. یکی از اجزای توصیفگر HPE، هیستوگرام در فضای رنگی HSV می‌باشد. به ترتیب برای مؤلفه H شانزده بازه، S شانزده بازه و برای کانال V چهار بازه لحاظ می‌گردد که در مجموع ۳۶ بعد خواهیم داشت. برای سه تصویر یادشده میانگین این سه بردار به‌عنوان قسمت اول توصیفگر محاسبه می‌شود.

قسمت دوم توصیفگر یادشده از آنالیز جوهره استفاده می‌کند. آنالیز جوهره تجمیعی از چند تصویر در قالب یک مدل بازگوکننده فراهم می‌آورد به‌طوری که این مدل دربرگیرنده خصوصیات بافتی، شکلی و ظاهری است. این آنالیز یکبار به‌صورت کلی و بار دیگر به‌صورت محلی بر روی تصاویر حاصل از خوشه‌بندی اعمال می‌گردد. شکل ۲-۵ خلاصه‌ای از این الگوریتم را نشان می‌دهد.

^۱ Epitome

^۲ Histogram Plus Epitome



شکل ۲-۵: خلاصه‌ای از الگوریتم ارائه‌شده در [۱۴]

بدین ترتیب برای محاسبه فاصله دو مجموعه از تصاویر از رابطه سه‌بخشی زیر استفاده می‌شود:

$$d(X^A, X^B) = \beta_1 \cdot (d_c(H(X^A), H(X^B))) + \beta_2 \cdot (d_c(Ge(X^A), Ge(X^B))) \quad (4-2)$$

$$+ \beta_3 \cdot (d_e(Le(X^A), Le(X^B)))$$

که در آن d_c فاصله باتچریا^۱ می‌باشد. همچنین d_e فاصله بین جوهرهای محلی دو فرد می‌باشد که خود از رابطه حاصل می‌شود:

$$d_e = \frac{1}{C} \sum_{b \in Le(X^B)} \min_{a \in Le(X^A)} d_c(H(a), H(b)) \quad (5-2)$$

این رابطه فاصله دو جوهر محلی را بر اساس مجموعی از مینیمم فاصله تکه‌های جوهره B و تکه‌های جوهره A محاسبه می‌کند. در نهایت نسخه ترکیبی این روش با قطعه‌بندی غیر قرینه‌ای ارائه‌شده در [۱۵] منجر به بازشناسی با نرخ دقت بیشتر شده است.

۲-۱-۱-۳- تجمیع ویژگی‌های محلی بر پایه قرینگی

در مقاله [۱۵] از یک بازنمایی سه‌بخشی برای توصیف انسان استفاده شده است. در این بازنمایی ابتدا ناحیه بدن فرد به سه بخش سر، تنه و پاها تقسیم‌بندی می‌شود. برای جداسازی این بخش‌ها دو

^۱ Bhattacharyya

محور افقی که بیشترین اختلاف رنگی در دو سمت آن رخ دهد، استخراج می‌شود. از آنجایی که ناحیه سر، بازگوکننده و متمایزکننده خوبی برای انسان نیست، تنها از دو ناحیه تنه و پاها استفاده شده است. در گام بعد برای هر بخش با روشی مشابه قبل، محوری عمودی که تصویر فرد حول آن دارای قرینگی در رنگ باشد استخراج می‌شود. فرض قرینگی در راستای افقی، در اکثر اوقات برای انسان، فرض منطقی می‌باشد. سپس برای هر یک از دو بخش بالاتنه و پاها، سه ویژگی استخراج می‌گردد.

اولین ویژگی هیستوگرام وزن دار می‌باشد. این هیستوگرام بر اساس مدل رنگی HSV محاسبه می‌شود. طریقه وزن دار کردن آن نیز بدین ترتیب است که هرچه یک پیکسل در تصویر فرد از محور قرینگی عمودی دور باشد وزن کمتری در هیستوگرام خواهد گرفت. دلیل این وزن دهی در این موضوع نهفته است که هرچه از محور عمودی دور می‌شویم احتمال تعلق پیکسل به پس‌زمینه بیشتر می‌گردد. دومین ویژگی مورد استفاده، توصیفگر نواحی رنگی با بیشترین پایداری^۱ (MSCR) می‌باشد. توصیفگر MSCR در واقع خوشه‌بندی بر روی پیکسل‌های هر ناحیه از تصویر شخص است. در حین خوشه‌بندی این نواحی به‌طوری انتخاب می‌شوند که به لحاظ رنگی فاصله بیشتری داشته باشند. هر خوشه با نه بعد نمایش داده می‌شود که شامل مرکز خوشه، میانگین رنگی RGB و ماتریس ممنتوم^۲ دوم می‌باشد. در حالت چندتصویره خوشه‌بندی دیگری بر روی این نواحی برای چند تصویر انجام می‌شود.

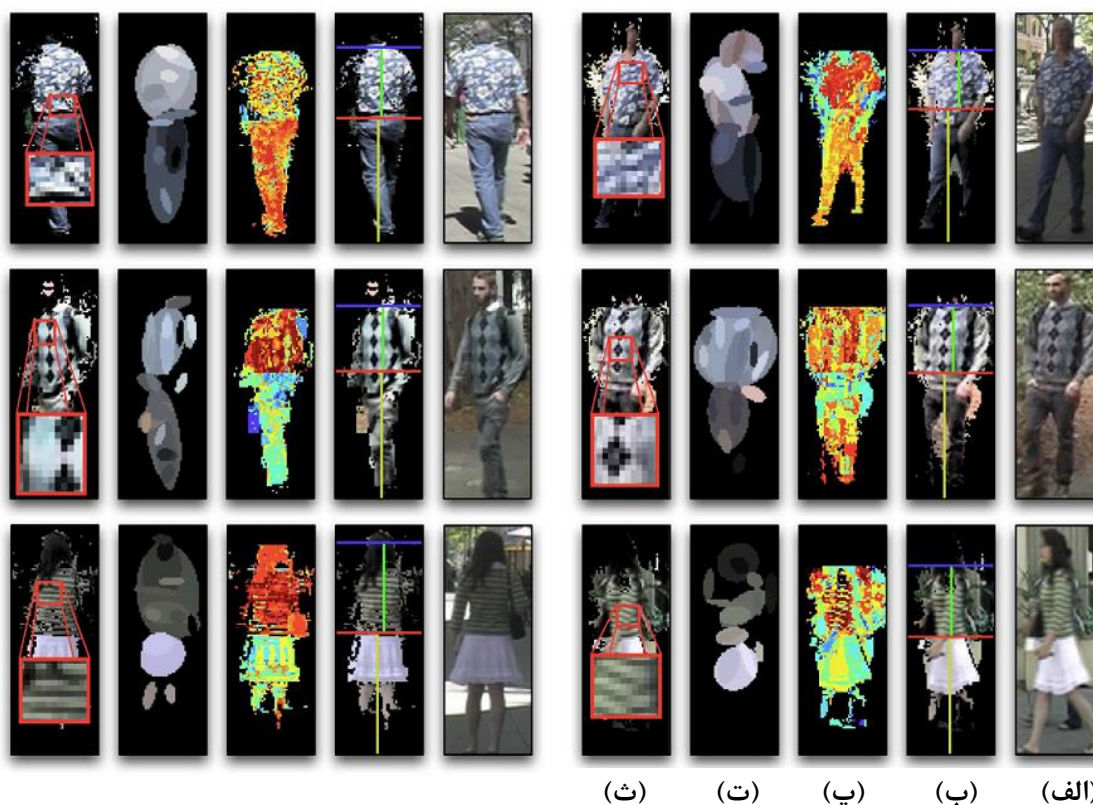
سومین ویژگی مورد استفاده، نواحی با خصوصیت ساختاری بالا و مکرر (RHSP^۳) می‌باشد. در این ویژگی هدف استخراج نواحی از تصویر است که به لحاظ خصوصیات بافتی، بیشتر در تصویر دیده می‌شود. بدین منظور ابتدا تعدادی ناحیه به‌صورت تصادفی در تصویر انتخاب می‌شود. در مرحله بعد نواحی که مقدار آنتروپی کانال‌های رنگی آن‌ها از یک مقدار آستانه پایین‌تر باشند، حذف می‌شوند

^۱ Maximally Stable Color Regions (MSCR)

^۲ Recurrent High-Structured Patches (RHSP)

^۳ Momentum

(نواحی هموار). سپس هر تکه بدست آمده، تحت تبدیلی‌هایی از قبیل چرخش و اندازه قرار می‌گیرد. در نتیجه نواحی بدست آمده بیشترین مقاومت را نسبت به تغییرات ظاهری خواهند داشت. سپس همبستگی متقابل محلی نرمال شده^۱ (LNCC) تکه اصلی و تبدیلی‌های تولیدشده از روی آن با تصویر اولیه محاسبه می‌گردد. مجموع LNCCهای بدست آمده برای یک تکه میزان وقوع این تکه را به‌طور کلی در تصویر نشان می‌دهد. در مرحله بعد تکه‌هایی که نقشه LNCC آنها از یک مقدار آستانه بیشتر باشد انتخاب شده و در نهایت با اعمال خوشه‌بندی بر روی تکه‌های انتخابی، نواحی با ساختار بالا و مکرر انتخاب می‌گردد. شکل ۲-۶ الگوریتم SDALF^۲ و سه ویژگی استخراجی را نشان می‌دهد.



شکل ۲-۶: سه ویژگی از معرفی شده در الگوریتم SDALF برای شش تصویر. ستون الف: تصویر اصلی، ستون ب: استخراج نواحی تنه و پاها و محورهای قرینگی، ستون پ: تصویر back-projection هیستوگرام وزن‌دار، ستون ت: توصیفگر MSCR و ستون ث: توصیفگر RHSP [۱۵]

^۱ Local Normalized Cross-Correlation (LNCC)

^۲ Symmetry-Driven Accumulation of Local Features (SDALF)

در نهایت ترکیب خطی وزن داری از این سه جزء فاصله بین دو فرد را تعیین می کند.

۲-۱-۱-۴- روش های بر پایه توصیفگر کوواریانس

توصیفگر کوواریانس در ابتدا برای تشخیص اشیاء و کلاسه بندی بافت در تصاویر مورد استفاده قرار گرفت [۱۶]. این توصیفگر برای یک ناحیه از تصویر محاسبه می شود. از خصوصیات مطلوب این توصیفگر، مقاومت آن نسبت به چرخش، تغییرات نوری و اندازه می باشد. لذا می توان از آن در کاربرد بازشناسی نیز استفاده نمود. ایده اصلی در این توصیفگر محاسبه ماتریس کوواریانس تعدادی بردار ویژگی هم بعد، برای یک ناحیه از تصویر می باشد. این ویژگی ها می تواند شامل موقعیت، کانال های رنگی و پاسخ های ناحیه مذکور به انواع فیلترها باشد. در واقع توصیفگر کوواریانس همبستگی بین چندین نوع ویژگی را محاسبه می کند که این ماتریس می تواند خصوصیات ساختاری و بافتی آن ناحیه از تصویر را بازنمایی کند.

نحوه محاسبه این توصیفگر بدین ترتیب است که فرض کنید R_1 ، ناحیه ای از تصویر باشد که می خواهیم این توصیفگر را برای آن محاسبه کنیم. برای هر پیکسلی که درون این ناحیه قرار می گیرد ویژگی های مورد نظر از قبیل رنگ، گرادیان و پاسخ های فیلتری محاسبه می شود. بدین ترتیب برای هر پیکسل برداری d بعدی از ویژگی ها را خواهیم داشت. برای مثال در کاربرد بازشناسی انسان، می توان سه کانال رنگی R و G و B و خصوصیت y را به عنوان ویژگی در نظر گرفت. دلیل انتخاب خصوصیت y این است که تصویر انسان در راستای x به دلیل قرینگی متمایز کننده نیست. با فرض اینکه ابعاد تصویر $W \times H$ باشد مجموعه بردارهای ویژگی $\{f_i\}_{i=1 \dots W*H}$ به صورت زیر تعریف می شود:

$$f_i = (y, R(x, y), G(x, y), B(x, y))^T \quad (۶-۲)$$

بر اساس این بردارهای ویژگی، ماتریس کوواریانس به صورت زیر محاسبه می گردد:

$$Cov_{R_1} = \frac{1}{WH} \sum_{i=1}^{WH} (f_i - \mu_{R_1})(f_i - \mu_{R_1})^T \quad \mu_{R_1} = \frac{1}{WH} \sum_{i=1}^{WH} f_i \quad (7-2)$$

که در آن μ_{R_1} بردار میانگین بردار ویژگی‌هاست. بدین ترتیب ماتریس کوواریانس برای هر ناحیه از تصویر با هر اندازه‌ای، ماتریسی مربعی و قرینه و مثبت معین با ابعاد $d \times d$ می‌باشد.

از آنجایی که ماتریس کوواریانس در فضای اقلیدسی قرار ندارد، برای محاسبه شباهت دو ماتریس کوواریانس نمی‌توان از معیارهای فاصله مرسوم نظیر فاصله اقلیدسی استفاده نمود. ماتریس‌های کوواریانس بر روی خمینه^۱ ریمانی قرار گرفته‌اند. خمینه ریمانی شکلی شبیه به کره دارد. بنابراین این ماتریس‌ها نقاطی بر روی یک کره می‌باشند. به صورت محلی بر روی این کره می‌توان از فاصله اقلیدسی استفاده نمود ولی در حالت کلی چنین نیست. در فضای ریمانی فاصله یا عدم شباهت دو ماتریس کوواریانس C_i و C_j از رابطه زیر محاسبه می‌شود.

$$\rho(C_i, C_j) = \sqrt{\sum_{k=1}^d \ln^2 \lambda_k(C_i, C_j)} \quad (8-2)$$

که در آن $\lambda_k(C_i, C_j)_{k=1 \dots d}$ مقادیر ویژه تعمیم‌یافته دو ماتریس C_i و C_j است که از رابطه زیر محاسبه می‌گردد:

$$\lambda_k C_i x_k - C_j x_k = 0 \quad k = 1 \dots d \quad (9-2)$$

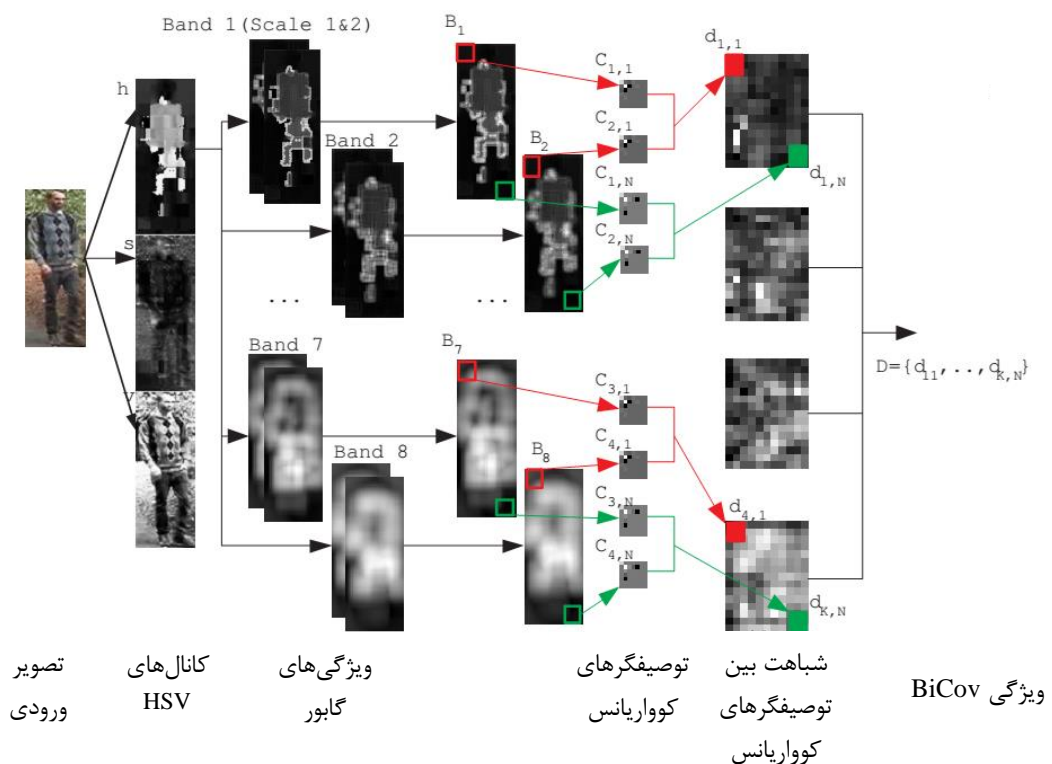
در مقاله [۱۷] بازشناسی را در سیستمی چندتصویره انجام می‌دهد. بدین ترتیب که برای یک فرد دنباله‌ای از تصاویر داریم. برای هر یک از این تصاویر توصیفگر کوواریانس محاسبه می‌شود. سپس گرافی تشکیل می‌دهیم که رئوس آن ماتریس‌های کوواریانس و یال‌ها میزان فاصله ماتریس‌ها از هم می‌باشد که در بالا ذکر شد. سپس بر روی گراف بدست آمده خوشه‌بندی طیفی^۲ صورت گرفته و در نتیجه برای

^۱ Manifold

^۲ Spectral clustering

هر دنباله، k میانگین محاسبه می‌شود و در نهایت در مرحله بازشناسی، این k میانگین با هم مقایسه و شباهت تعیین می‌گردد.

در مقاله [۱۸] تصویر به سه کانال رنگی HSV تجزیه و سپس به کمک فیلتر گابور در هشت باند و در هر باند دو سایز مختلف بر روی هر کانال رنگی اعمال می‌شود. تصاویر بدست آمده از فیلترهای گابور به نواحی کوچک‌تری شکسته می‌شود. این نواحی با هم، دارای همپوشانی جزئی هستند. دلیل اینکه برای کل تصویر از یک ماتریس کوواریانس استفاده ننموده، سرعت بیشتر محاسبات در ناحیه‌های کوچک‌تر و همچنین حفظ بیشتر خصوصیات مکانی است. در نهایت در هر باند برای دو ماتریس کوواریانس متناظر، در دو سایز مختلف فاصله برحسب رابطه ۲-۸ محاسبه و این تفاضل‌ها سرانجام به هم متصل گردیده و برداری را تشکیل می‌دهند. ویژگی‌های مورد استفاده در این مقاله ویژگی x ، y ، شدت رنگی، مقادیر مشتق اول و دوم در راستای x و مقادیر مشتق اول و دوم در راستای y می‌باشد. طرح کلی الگوریتم این مقاله در شکل ۲-۷ نشان داده شده است.



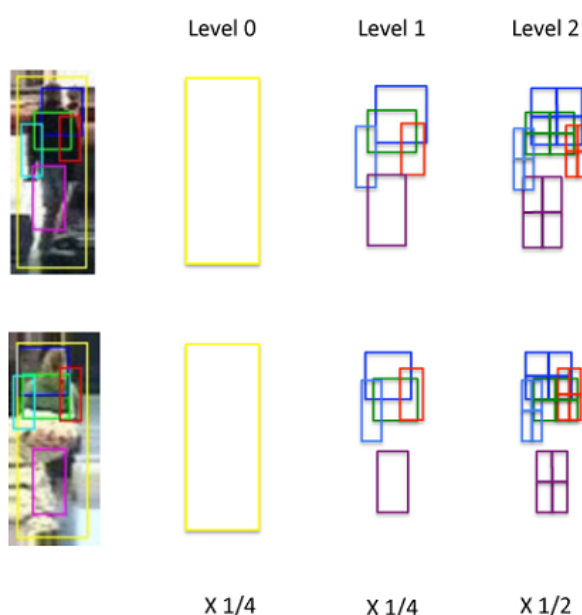
شکل ۲-۷: طرح کلی الگوریتم مقاله [۱۸]

در [۱۹] با استفاده از توصیفگر بدست‌آمده و استفاده در ساختار KISSME [۲۰] نسخه‌ای بهبودیافته برپایه یادگیری معیار فاصله ارائه شده است. در مقاله [۲۱] قبل از محاسبه توصیفگر کوواریانس از یکسان‌سازی هیستوگرام استفاده شده است. دلیل استفاده از یکسان‌سازی، بهبود عملکرد الگوریتم در تغییرات نوری است. در این مقاله تعدادی ویژگی در نظر گرفته شده سپس نواحی مختلفی در تصویر یک فرد، به صورت تصادفی در اندازه‌های مختلف انتخاب می‌گردد. یکی از ویژگی‌های انتخابی، فاصله از مرکز تصویر است که برگرفته از اهمیت اطلاعات این ناحیه می‌باشد. سپس برای نواحی مذکور، با استفاده از روش انتخاب ویژگی مبتنی بر همبستگی^۱ (CFS) سه ویژگی بهتر انتخاب می‌شود. سه ویژگی مذکور به‌گونه‌ای انتخاب می‌شوند که برای داده‌های یک کلاس فاصله کمتر و برای داده‌های غیر یک کلاس فاصله بیشتر شود. پس از انتخاب ویژگی‌ها و ساخت ماتریس کوواریانس برای هر ناحیه، جفت تصاویر افراد بر اساس ماتریس کوواریانس‌های متناظر، قابل‌مقایسه خواهند بود. در حالت چندتصویره نیز با استفاده از تعریف میانگین در بحث ماتریس‌های کوواریانس میانگین هر ناحیه حساب گردیده و در نهایت مشابه با حالت تک تصویره عمل خواهیم کرد. در مقاله [۲۲] نیز از فیلتر گابور و LBP برای بردار ویژگی استفاده نموده و همچنین ناحیه بدن فرد را به چهار قسمت تقسیم‌بندی نموده و بر روی هر ناحیه توصیفگر کوواریانس را براساس ویژگی‌های یادشده محاسبه می‌کند.

مقاله [۲۳] یکی از روش‌های موفق در استفاده از توصیفگر کوواریانس است. در این مقاله از یک ساختار هرمی استفاده شده است. بدین ترتیب که ابتدا به کمک توصیفگر HOG و کلاسه‌بندی که به‌منظور استخراج ناحیه‌ای از تصویر که انسان در آن واقع شده، آموزش‌دیده است، ناحیه فرد را تشخیص می‌دهیم. در مرحله بعد به کمک کلاسه‌بندهای دیگری که باز هم با استفاده از توصیفگر HOG ولی این بار برای تشخیص اعضای بدن آموزش‌دیده‌اند استفاده می‌کنیم. بعد از استخراج ناحیه‌های مستطیلی مربوط به اعضای بدن، ساختاری هرمی خواهیم داشت. شکل ۲-۸ ساختار هرمی سه‌سطحی معرفی شده

^۱ Correlation based Feature Selection (CFS)

در این مقاله را نشان می‌دهد. همان‌طور که در شکل مشاهده می‌شود سطح اول مربوط به ناحیه کل بدن فرد می‌باشد. سطح بعدی مربوط به اعضای بدن فرد و سطح سوم مربوط به زیر ناحیه اعضای بدن فرد می‌باشد. فرآیند تطبیق در سیستم‌هایی به شکل چند سطحی و هرمی بدین صورت است که برای سطوح جزئی‌تر وزن بیشتری اختصاص داده می‌شود و برای سطوح کلی‌تر وزن کمتر.



شکل ۲-۸: ساختار هرمی در سه سطح به منظور استخراج توصیفگر کوواریانس در مقاله [۲۳]

رابطه زیر هسته تطبیق هرمی را نشان می‌دهد. فرض کنید می‌خواهیم دو مدل X و Y را با هم مقایسه کنیم:

$$\kappa^L(X, Y) = \frac{1}{2^L} \mathcal{J}^0 + \sum_{l=1}^L \frac{1}{2^{L-l+1}} \mathcal{J}^l \quad (10-2)$$

که در آن \mathcal{J}^l تابع تطبیق در سطح l هرم است. در این مقاله سطح صفر، معادل کل ناحیه بدن فرد می‌باشد. بدین ترتیب داریم $\mathcal{J}^0 = \frac{1}{\rho(C_i, C_j)}$ و برای سطوح بعدی مقدار تابع \mathcal{J} به صورت زیر خواهد بود:

$$\mathcal{J}^l = \frac{1}{\mathcal{D}^l} \quad \mathcal{D}^l = \frac{\sum_{k=1}^{n(l)} \rho(C_k^i, C_k^j) - \sum_{m \in \mathcal{M}_Z} \rho_m}{n(l) - z} \quad (11-2)$$

که در آن \mathcal{M}_Z مجموعه Z بزرگترین ρ ها بین ماتریس‌های کوواریانس متناظر می‌باشد. مقدار انتخاب‌شده برای $Z = \frac{n(l)}{2}$ می‌باشد. بردار ویژگی‌های استفاده شده در این مقاله نیز به صورت زیر است:

$$(x, y, R_{xy}, G_{xy}, B_{xy}, \|\nabla_{xy}^R\|, \theta_{xy}^R, \|\nabla_{xy}^G\|, \theta_{xy}^G, \|\nabla_{xy}^B\|, \theta_{xy}^B) \quad (12-2)$$

در مقاله [۲۴] ابتدا دنباله‌ای از تصاویر یک فرد، جدول‌بندی می‌شود (با سلول‌هایی که دارای همپوشانی هستند). سپس برای هر سلول توصیفگر کوواریانس حساب می‌گردد. برای هر موقعیت از سلول‌ها با استفاده از الگوریتمی تکراری بر اساس روش نزول در جهت کاهش گرادیان (روش کارچر)، میانگین n ماتریس کوواریانس در دنباله مذکور بر اساس رابطه زیر محاسبه می‌شود.

$$\mu_{t+1} = \exp_{\mu_t} \left[\frac{1}{N} \sum_{i=1}^N \log_{\mu_t}(C_i) \right] \quad (13-2)$$

در این رابطه \exp_{μ_t} و \log_{μ_t} عملگرهای مخصوصی در خمینه ریمانی می‌باشند. خروجی این مرحله MRCG^۱ نامیده می‌شود. بدین ترتیب برای n فرد در سیستم، مجموعه n امضا $\mathfrak{S}^c = \{\mathfrak{s}_i^c\}_{i=1}^n$ به صورت زیر خواهیم داشت (m تعداد سلول‌ها در جدول‌بندی می‌باشد):

$$\mathfrak{s}_i^c = \{\mu_{i,1}^c, \mu_{i,2}^c, \dots, \mu_{i,m}^c\} \quad (14-2)$$

در قدم بعد برای هر فرد، میزان اهمیتی برای هر سلول‌های آن تعریف می‌شود. بدین منظور از تعریف واریانس استفاده می‌کند. بدین ترتیب که سلولی که در مجموع فاصله بیشتری با سلول متناظر آن در بین کل افراد داشته باشد، برای فرد مذکور سلول با اهمیتی است. در نتیجه با این روش سلول‌های پس‌زمینه به دلیل اینکه واریانس زیادی ندارند وزن کمتری گرفته و همچنین تمرکز بر روی تفاوت‌های فرد با بقیه افراد می‌باشد. بر این اساس واریانس هر سلول از رابطه زیر محاسبه گردیده:

$$\sigma_{i,j}^c = \frac{1}{n-1} \sum_{k=1, k \neq i}^N \rho^2(\mu_{i,j}^c, \mu_{k,j}^c) \quad (15-2)$$

^۱ Mean Riemannian Covariance Grid (MRCG)

در نهایت شباهت دو فرد به صورت زیر محاسبه می گردد:

$$S(\xi_A, \xi_B) = \sum_{i \in K} \frac{\sigma_{A,i} + \sigma_{B,i}}{\rho(\mu_{A,i}, \mu_{B,i})} / \|K\| \quad (۱۶-۲)$$

در [۲۵] تصویر شخص به نواحی سطری که با یکدیگر همپوشانی دارند، تقسیم می گردد. سپس نواحی مذکور به زیرنواحی متراکمی تقسیم بندی می گردد. برای هر یک از زیرنواحی بازنمایی مشابه با توصیفگر کوواریانس محاسبه می شود. بدین ترتیب که برای هر پیکسل (i) از زیرنواحی، بردار ویژگی بصورت زیر ساخته می شود.

$$f_i = (y, M_{0^\circ}, M_{90^\circ}, M_{180^\circ}, M_{270^\circ}, R, G, B)^T \quad (۱۷-۲)$$

در این بردار y مختصات در راستای عمودی، M_i° اندازه گرادیان در راستای i° و RGB سه کانال رنگی قرمز، سبز و آبی می باشد. سپس برای یک زیرناحیه (s)، توزیع آماری پیکسل های آن بر اساس مدل گاوسی تک مده بصورت زیر تعیین می گردد.

$$\mathcal{N}(f; \mu_s, \Sigma_s) = \frac{\exp\left(-\frac{1}{2}(f - \mu_s)^T \Sigma_s^{-1} (f - \mu_s)\right)}{(2\pi)^{\frac{d}{2}} |\Sigma_s|} \quad (۱۸-۲)$$

که در آن Σ_s ماتریس کوواریانس و μ_s میانگین پیکسل های زیرناحیه s می باشد. بدین ترتیب برای یک ناحیه از تصویر تعدادی توزیع گاوسی به فرم رابطه ۲-۱۸ خواهیم داشت. از آنجایی که این توزیع ها بر روی یک خمینه ریمانی واقع شده اند بر طبق ریاضیات مربوطه، فضای d بعدی گاوسی به فضای $d + 1$ بعدی Sym_{d+1}^+ که نوعی از فضای مماسی^۱ اقلیدسی است، قابل نگاشت می باشد. تخمینی از این تبدیل بصورت رابطه زیر می باشد.

$$\mathcal{N}(f; \mu_s, \Sigma_s) \sim P_s = |\Sigma_s|^{-\frac{1}{d+1}} \begin{bmatrix} \Sigma_s + \mu_s \mu_s^T & \mu_s \\ \mu_s^T & 1 \end{bmatrix} \quad (۱۹-۲)$$

^۱ Tangent space

در نهایت ماتریس رابطه فوق به برداری با $m = \frac{(d^2+3d)}{2} + 1$ بُعد بصورت زیر ایجاد می‌گردد. تابع vec عناصر بالامثلشی از ماتریس آرگومان خود را بصورت برداری بر می‌گرداند.

$$g_s = vec(log(P_s)) \quad (20-2)$$

$$= [b_{s(1,1)}, \sqrt{2}b_{s(1,1)}, \dots, \sqrt{2}b_{s(1,d+1)}, b_{s(2,2)}, \sqrt{2}b_{s(2,3)}, \dots, b_{s(d+1,d+1)}]^T$$

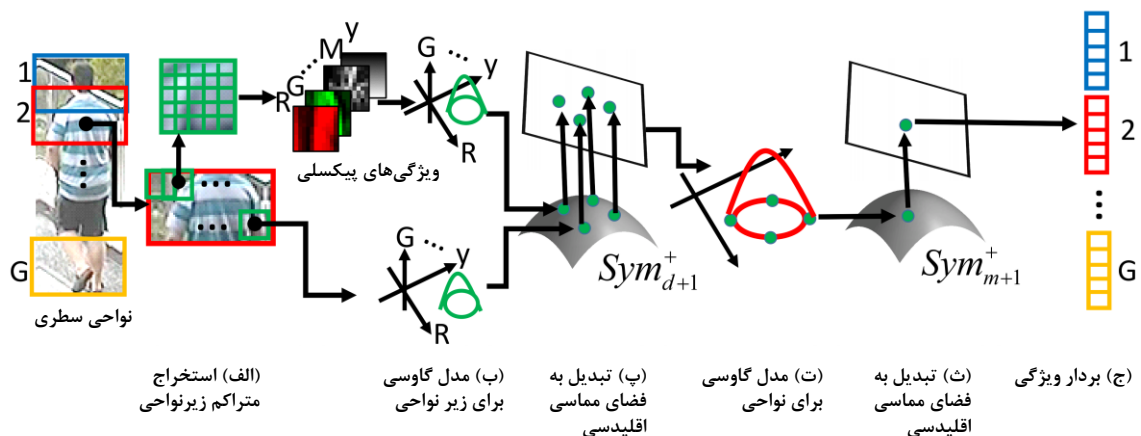
که در آن $b_{s(i,j)}$ درایه (i,j) از ماتریس $B_s = log(P_s)$ می‌باشد. بدین ترتیب فضای ماتریسی تبدیل به برداری شده که عملگرهای فضای اقلیدسی بر روی آن قابل استفاده می‌باشد. در نتیجه در یک ناحیه به تعداد زیرنواحی، بردار m بعدی خواهیم داشت. بردارهای بدست آمده مجدداً به وسیله توزیع گاوسی تک‌مده $\mathcal{N}(g; \mu^G, \Sigma^G)$ به طریق مشابه مدل می‌گردند. که G مشخص‌کننده ناحیه خاصی در تصویر می‌باشد. به منظور کاهش تاثیرات پس‌زمینه، مکانیزم وزن‌دهی برای زیر نواحی برای محاسبه μ^G و Σ^G بصورت زیر استفاده شده است.

$$\mu^G = \frac{1}{\sum_{s \in G} w_s} \sum_{s \in G} w_s g_s \quad (21-2)$$

$$\Sigma_s = \frac{1}{\sum_{s \in G} w_s} \sum_{s \in G} w_s (g_s - \mu^G)(g_s - \mu^G)^T$$

در این روابط w_s متناظر با زیرناحیه s می‌باشد. و در نهایت مشابه با رابطه 20-2، توزیع گاوسی اخیر به فضای برداری $1 + \frac{(m^2+3m)}{2}$ بُعدی نگاشت می‌گردد. بردار ویژگی نهایی از کنار هم قراردادن بردارهای نگاشت نهایی برای ناحیه‌های افقی و دارای همپوشانی کل تصویر حاصل می‌گردد. همچنین به منظور بالابردن دقت، ترکیبی از چهار فضای رنگی RGB, Lab, HSV و nRnG به عنوان ویژگی ترکیبی ارائه شده است. در این مقاله ویژگی نهایی تحت عنوان گاوسی گاوسی^۱ (GoG) می‌باشد که در بردارنده الگوریتم ذکر شده در بالا است. شکل 2-9 خلاصه‌ای از الگوریتم محاسبه توصیفگر گاوسی سلسله مراتبی را نشان می‌دهد.

^۱ Gaussian of Gaussian (GoG)



شکل ۲-۹: خلاصه‌ای از الگوریتم محاسبه توصیفگر گاوسی سلسله مراتبی [۲۵]

۲-۱-۱-۵- مدل مکانی-زمانی بر پایه بخش‌بندی بدن انسان

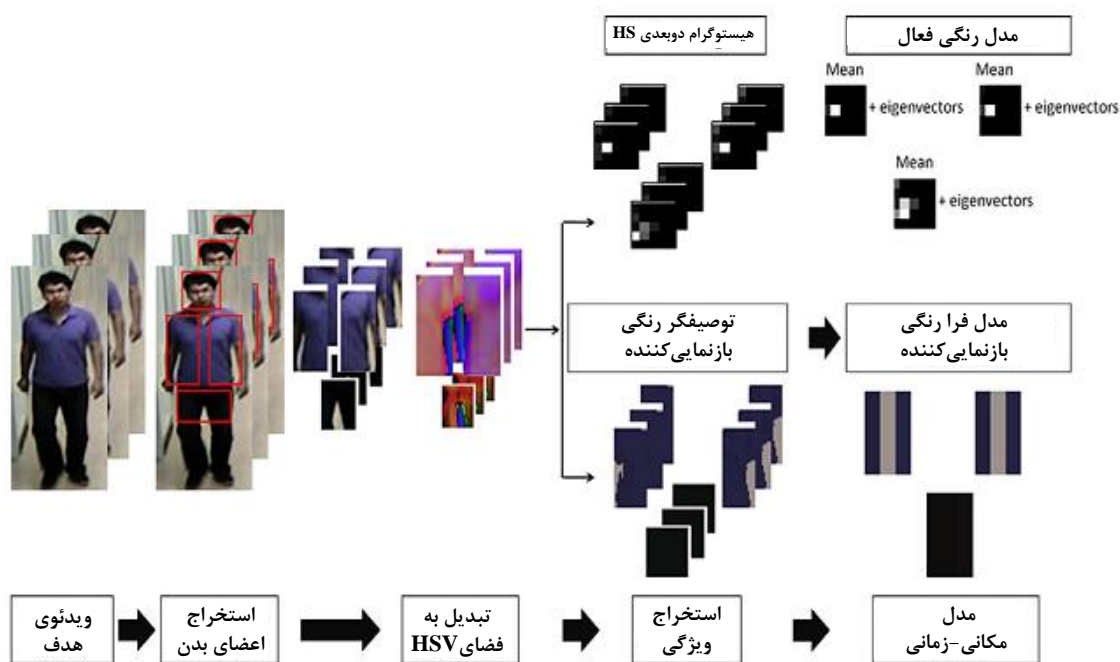
در مقاله [۲۶] ابتدا اعضای بدن در تصویر استخراج می‌شوند. از بین شش عضو استخراج شده تنها چهار عضو پایدار سر، قسمت راست تنه، قسمت چپ تنه و قسمت بالای پاها انتخاب می‌گردند. برای هر قسمت، دو ویژگی مدل رنگی فعال^۱ (ACM) و مدل فرا رنگی بازنمایی کننده^۲ (RMC) محاسبه می‌شود. برای محاسبه مدل ACM با استفاده از دو کانال H و S در مدل رنگی HSV هیستوگرامی دوبعدی تشکیل می‌گردد. تعداد بازه‌های فراوانی برای هر کدام از این دو کانال، مقدار پنج انتخاب شده که هیستوگرامی شامل ۲۵ مؤلفه خواهد بود. برای دنباله‌ای از تصاویر یک فرد، این هیستوگرام دوبعدی به صورت رابطه $g = \bar{g} + A_g \cdot b_g$ خواهد بود که در آن \bar{g} میانگین این ماتریس سه‌بعدی برای دنباله‌ای از فریم‌ها می‌باشد. A_g نیز مدهای این مدل دوبعدی را مشخص می‌کند. برای بدست آوردن RMC، خوشه‌بندی در فضای رنگی RGB برای هر قسمت انجام می‌شود. تعداد خوشه‌ها کوچک‌تر مساوی ۳ در نظر گرفته می‌شود. برای هر خوشه میانگین رنگی، بیانگر رنگ کلی خوشه می‌باشد. در قدم بعد، برای فریم‌های متوالی عمل تطبیق خوشه‌ها صورت می‌گیرد. خوشه‌هایی که برای ده فریم متوالی با هیچ

^۱ Active Color Model

^۲ Representative Meta Colors Model

خوشه دیگری تطبیق نیابد حذف می‌شوند. در نهایت بین خوشه‌های باقی‌مانده، مجدداً خوشه‌بندی انجام می‌گیرد و خوشه‌های نهایی بدست آمده، حاوی اطلاعات مکانی و زمانی خواهند بود.

همچنین در این مقاله برای بخش صورت فرد در صورتی که کنتراست^۱ بیشتر از یک مقدار آستانه را دارا باشد سعی در استخراج ویژگی از جزئیات صورت فرد دارد. در صورتی که کنتراست از مقدار آستانه کمتر باشد بخش صورت در مرحله بازشناسی مورد استفاده قرار نمی‌گیرد. به منظور تطبیق نیز رابطه‌ای خطی و وزن دار از روی سه ویژگی صورت و ACM و RMC تشکیل داده می‌شود. شکل ۲-۱۰ طرحی کلی از الگوریتم ارائه شده را نشان می‌دهد.



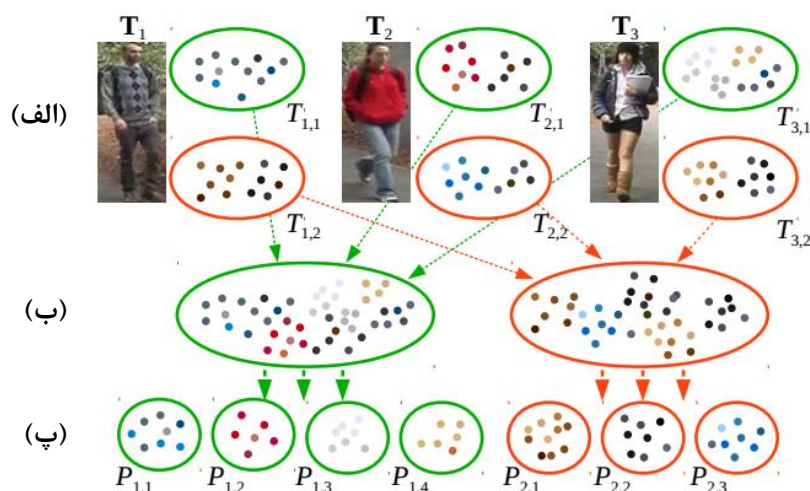
شکل ۲-۱۰: طرح کلی از روش ارائه شده در [۲۶]

۲-۱-۱-۶- بازشناسی سریع بر اساس بازنمایی عدم شباهت

یکی از رویکردهای دیگر در بازشناسی، بازنمایی یک فرد در قالب برداری از مؤلفه‌های عدم شباهت می‌باشد. در مقاله [۲۷] این بازنمایی ارائه شده است. هدف اصلی از این بازنمایی پایین آوردن پیچیدگی

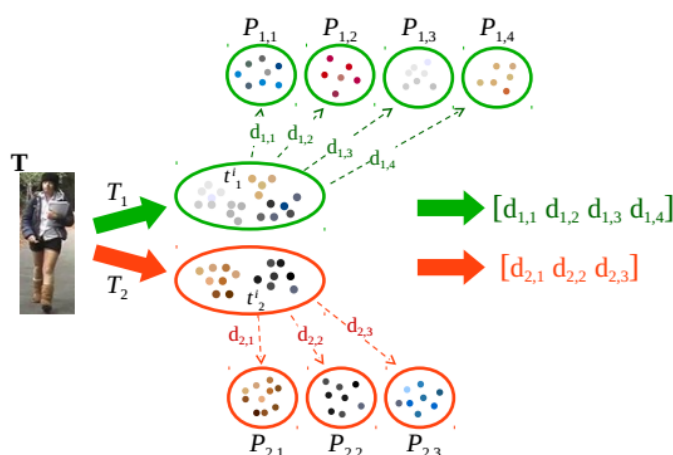
^۱ Contrast

محاسباتی در مرحله بازشناسی می‌باشد. دقت تشخیص در روش ارائه‌شده در حد قابل قبولی است اما پیچیدگی محاسباتی پایینی دارد. در این معماری ابتدا باید ناحیه بدن فرد به چند بخش تقسیم‌بندی گردد. سپس برای هر بخش، بردار ویژگی ایجاد شود. در قدم بعدی برای مجموعه‌ای از تصاویر افراد برای هر بخش به تفکیک، خوشه‌بندی انجام می‌گیرد. برای مثال برای قسمت تنه این خوشه‌بندی منجر به تولید چهار خوشه می‌گردد. برای پایین‌تنه نیز فرض کنید سه خوشه تولید گردد. شکل ۲-۱۱ مراحل تعیین نماینده‌ها را نشان می‌دهد.



شکل ۲-۱۱: سطر الف: بخش‌بندی تصویر فرد به دو ناحیه و استخراج مؤلفه‌های ویژگی. سطر ب: ادغام کلیه مؤلفه‌های مرتبط با یک بخش. سطر پ: خوشه‌بندی هر بخش و استخراج نماینده‌ها [۲۷]

در هنگام بازشناسی یک فرد جدید، مؤلفه‌های هر بخش را با نماینده‌ها مقایسه و فاصله با هر یک محاسبه و در برداری ذخیره می‌شود. بردار ویژگی برای مجموعه گالری نیز به همین ترتیب محاسبه می‌شود. شکل ۲-۱۲ مرحله بازشناسی را نشان می‌دهد.



شکل ۲-۱۲: مرحله بازشناسی در بستر عدم شباهت [۲۷]

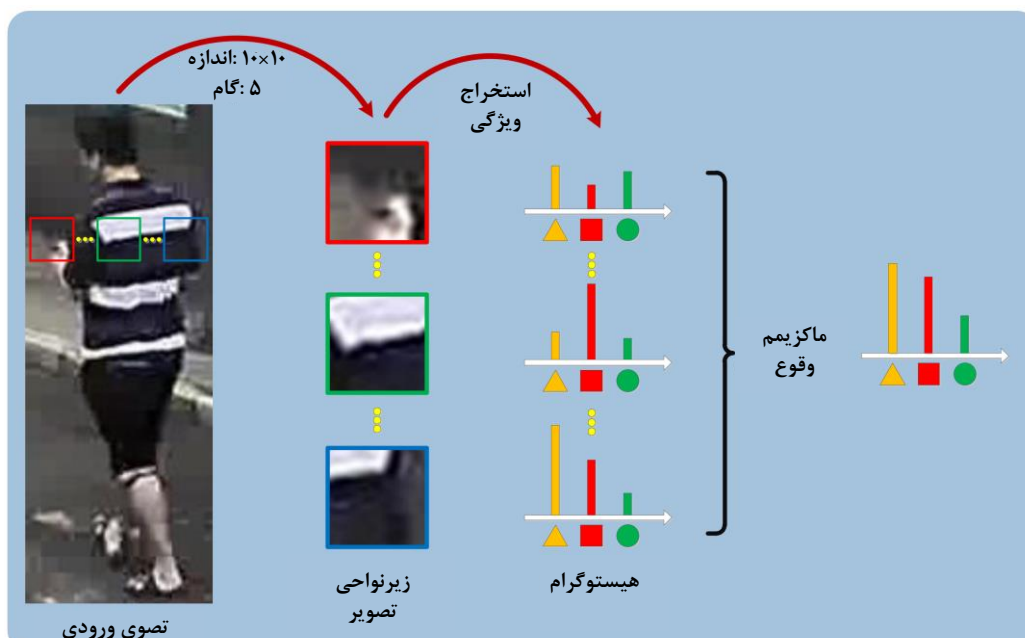
۲-۱-۱-۷- بازشناسی بر پایه هیستوگرام رنگی و بافت

هیستوگرام یکی از ویژگی‌های مناسب به منظور بازنمایی یک تصویر می‌باشد. دلیل استفاده گسترده از این ابزار، سرعت بالا در پردازش‌هایی نظیر محاسبه و مقایسه دو هیستوگرام می‌باشد. از مشکلات اصلی هیستوگرام در نسخه استاندارد، نادیده گرفتن اطلاعات مکانی در تصویر است. برای مثال در [۲۸] بازنمایی جدیدی بر پایه هیستوگرام رنگی در فضای HSV و توصیفگر^۱ SILTP معرفی شده است. توصیفگر SILTP نسخه بهبودیافته‌ای از توصیفگر LBP است که به منظور توصیف بافت‌ها در تصویر استفاده می‌شود. از ویژگی مذکور تحت عنوان ماکزیمم وقوع محلی^۲ یاد می‌شود. ابتدا تصویر ورودی به فضای رنگی HSV نگاشت می‌گردد. همچنین بر روی تصویر با سطوح خاکستری توصیفگر SILTP اعمال می‌شود. خروجی این توصیفگر تصویری هم‌بعد با تصویر ورودی است که هر پیکسل از آن، دربردارنده نسبت مقدار آن پیکسل با مقادیر پیکسل‌های مجاور در یک شعاع همسایگی است. در گام بعد، هیستوگرامی از این مقادیر برای یک ناحیه خاص، بازگو کننده خصوصیات بافتی برای آن ناحیه می‌باشد. بدین ترتیب ابتدا تصویر ورودی به تعدادی سطر بخش‌بندی می‌شود. این سطرها با یکدیگر دارای همپوشانی هستند. در هر سطر پنجره‌ای با ابعاد مشخص و با اندازه گام‌های مشخص لغزنده

^۱ Scale Invariant Local Ternary Pattern (SILTP)

^۲ Local Maximal Occurrence (LOMO)

می‌شود. برای هر پنجره یک هیستوگرام HSV و دو هیستوگرام SILTP برای دو شعاع همسایگی محاسبه می‌شود. در نهایت برای هر سطر یک هیستوگرام HSV تشکیل می‌شود که مقادیر هر مؤلفه از آن، برابر با بیشینه مقدار برای آن مؤلفه در تمامی هیستوگرام‌های HSV سطر مربوطه می‌باشد. به‌طور مشابه هیستوگرام‌های SILTP برای هر سطر محاسبه می‌گردد. بردار ویژگی نهایی از کنار هم قرار دادن این هیستوگرام‌ها در سه مقیاس ۱، ۰.۵ و ۰.۲۵ از تصویر تشکیل می‌گردد. شکل ۲-۱۳ خلاصه‌ای از طریقه محاسبه ویژگی ماکزیمم وقوع محلی برای یک تصویر را نشان می‌دهد.



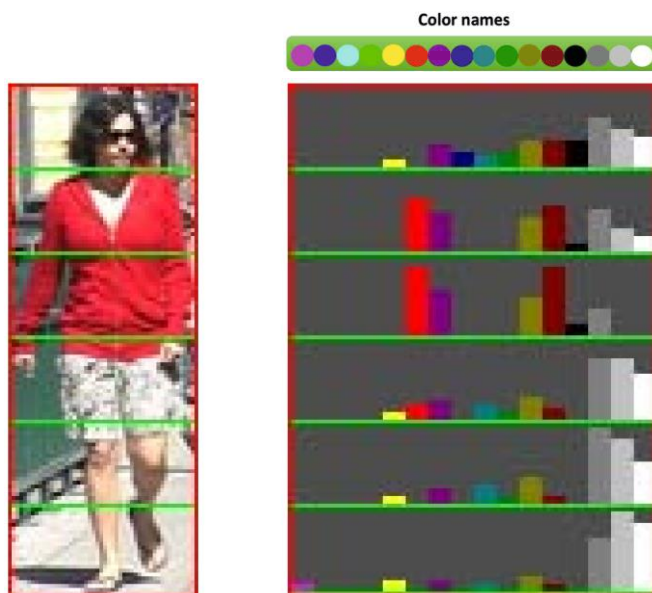
شکل ۲-۱۳: نحوه محاسبه ویژگی ماکزیمم وقوع محلی در [۲۸]

در مقالات [۲۹، ۳۰] سعی شده با در نظر گرفتن حالت فازی و احتمالی، کاربرد آن را در مسئله بازشناسی بهبود بخشد. در [۲۹] تعریف جدیدی از هیستوگرام رنگی تحت عنوان هیستوگرام رنگی احتمالاتی^۱ معرفی شده است. روش کار بدین ترتیب است که برای کلیه رنگ‌های موجود، ۱۱ رنگ مرسوم که در فرهنگ‌های مختلف وجود دارد، به‌عنوان رنگ‌های اصلی انتخاب می‌شود. سپس مجموعه‌ای از تصاویر رنگی مختلف، در شرایط محیطی گوناگون از قبیل روشنایی مختلف و غیره که برچسب رنگ

^۱ Probabilistic color histogram

خورده‌اند، با استفاده از روش خوشه‌بندی فازی خوشه‌بندی می‌شود. بدین ترتیب فضای رنگی Lab اولیه به ۱۱ رنگ به صورت فازی تقسیم‌بندی می‌گردد. در نهایت مقدار یک مؤلفه خاص از هیستوگرام نهایی برابر است با مجموع درجه عضویت تمام پیکسل‌های تصویر به آن رنگ خاص. بدین ترتیب مقدار رنگی یک پیکسل از تصویر، می‌تواند بر روی بیش از یک بازه فراوانی از هیستوگرام تغییر ایجاد نماید.

در [۳۱] نیز توزیع‌های رنگی روی مجموعه‌ای از رنگ‌های شناخته شده شامل ۱۶ رنگ، محاسبه می‌شود. بازنمایی نهایی شامل محاسبه این توزیع برای چندین فضای رنگی و ترکیب آنها می‌باشد. توصیفگر معرفی شده در این مقاله تحت عنوان توصیفگر رنگی برپایه رنگ‌های برجسته^۱ (SCNCD) می‌باشد. بدین منظور تصویر ورودی به شش بخش افقی تقسیم می‌گردد و برای هر بخش هیستوگرام روی ۱۶ رنگ مذکور محاسبه می‌شود. برای محاسبه تعلق رنگ‌ها به هر یک از این ۱۶ رنگ، k نزدیکترین همسایه انتخاب می‌شود. شکل ۲-۱۴ نمونه‌ای از محاسبه توصیفگر فوق را برای یک تصویر ورودی نشان می‌دهد.



شکل ۲-۱۴: نمونه‌ای از بازنمایی تصویر یک فرد، بر اساس توصیفگر SCNCD برای ۱۶ رنگ برجسته [۳۱]

^۱ Salient Color Names based Color Descriptor (SCNCD)

ایده اصلی مقاله [۳۰] نیز استفاده از پنج ویژگی R, G, B, x و y در هیستوگرام می‌باشد. بدین ترتیب که هیستوگرامی پنج‌بعدی بر اساس پنج ویژگی فوق تشکیل داده و سپس برای هر پیکسل با توجه به موقعیت و رنگ، درجه عضویتی به یکی از این بازه‌های فراوانی قائل شده است. بدین ترتیب همانند روش قبلی یک پیکسل بر روی بیش از یک بازه فراوانی از هیستوگرام تأثیر خواهد گذاشت.

در مقاله [۳۲] تعمیمی از هیستوگرام استاندارد ارائه شده است. نقطه قوت این تعمیم، لحاظ کردن خصوصیت مکان در هیستوگرام است. از این رو هیستوگرام جدید را Spatiogram می‌نامد. بر طبق تعریف، هیستوگرام استاندارد یک Spatiogram مرتبه صفر می‌باشد. تصویر I را نگاشتی از فضای X به فضای V به صورت زیر در نظر می‌گیریم:

$$I: \mathbf{x} \rightarrow \mathbf{v} \quad \mathbf{x} = [\mathbf{x}, \mathbf{y}]^T \quad (22-2)$$

بدین ترتیب هیستوگرام مکانی مرتبه دوم به صورت زیر تعریف می‌گردد:

$$h_l^2(b) = (n_b, \mu_b, \Sigma_b) \quad b = 1, \dots, B \quad (23-2)$$

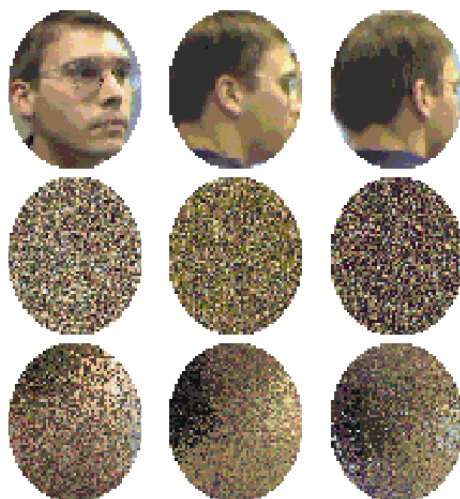
که در آن n_b تعداد پیکسل‌های بازه فراوانی b و μ_b و Σ_b به ترتیب میانگین مکانی و کوواریانس مکانی پیکسل‌هایی هستند که در بازه مربوطه واقع می‌شوند. با این تعریف جدید، شباهت دو هیستوگرام مکانی نیز به صورت زیر تعریف می‌گردد:

$$\rho(h, h') = \sum_{b=1}^B \psi_b \rho_n(n_b, n'_b) \quad (24-2)$$

$$\psi_b = \eta \cdot \exp\left(-\frac{1}{2}(\mu_b - \mu'_b) \hat{\Sigma}_b^{-1} (\mu_b - \mu'_b)\right) \quad \hat{\Sigma}_b^{-1} = \Sigma_b^{-1} + (\Sigma'_b)^{-1}$$

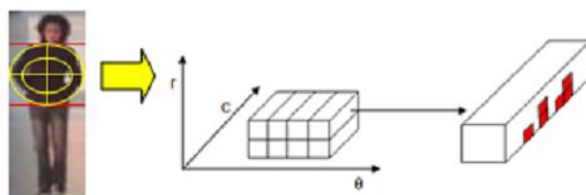
در این رابطه η ثابت نرمال‌سازی گاوسی است. فرض کنید هیستوگرام تصاویر سطر اول شکل ۲-۱۵ را داشته باشیم. سطر دوم از این شکل تصویر بازسازی شده با استفاده از هیستوگرام می‌باشد که کاملاً تصاویری بوده و رنگ‌ها بصورت یکنواخت در تصویر پراکنده شده‌اند. بر اساس هیستوگرام مکانی

مرتبۀ دوم، تصویر حاصله شباهت بیشتری با تصویر اولیه دارد به عبارتی خصوصیات مکانی بهتر حفظ گردیده‌اند.



شکل ۲-۱۵: مقایسه بازسازی یک تصویر از روی هیستوگرام مکانی و هیستوگرام استاندارد [۳۲]

در [۳۳] ابتدا میان تنه تصویر فرد، استخراج شده و چهار هیستوگرام به صورت شعاعی به مرکزیت نیم تنه ایجاد می‌شود. با این طریق اطلاعات مکانی به نوعی بازنمایی می‌گردد. در قسمت دیگری از الگوریتم ارائه شده قسمت پاها نیز به صورت یکنواخت تقسیم‌بندی می‌گردد. شکل ۲-۱۶ هشت هیستوگرام ذکر شده را براساس شعاع، زاویه در دایره و رنگ نشان می‌دهد.



شکل ۲-۱۶: تقسیم‌بندی ناحیه تنه بر اساس روش ارائه شده در [۳۳]

در نهایت این مجموعه از هیستوگرام‌ها به کمک فاصله باتچریا در مرحله بازشناسی مورد مقایسه قرار می‌گیرند.

۲-۱-۲- روش‌های فعال

در این دسته، هدف طراحی توصیفگر و یا معیاری برای تطبیق جفت تصویر انسان است که علاوه بر دارا بودن خاصیت بازگوکنندگی مطلوب، تمایزپذیری خوبی نیز بین انسان‌های مختلف ایجاد نماید. فرآیند یادگیری از ویژگی‌های اصلی این گروه می‌باشد که در نام‌گذاری این دسته بدان اشاره شده است. این یادگیری می‌تواند به صورت بانظارت و یا بی‌نظارت مورد استفاده قرار گیرد. به عبارتی در این دسته از روش‌ها خصوصیات و پارامترهای توصیفگر و یا معیار فاصله برای تطبیق، در یک فرآیند یادگیری تعیین می‌شوند.

۲-۱-۲-۱- هم‌ترازی رنگی

یکی از مشکلاتی که بر دقت بازشناسی تأثیر نامطلوب می‌گذارد تغییرات نوری است که منجر به تغییرات ظاهری فرد از یک دوربین به دوربین دیگر می‌شود. این تغییرات نوری می‌تواند میزان روشنایی و یا حتی تبدیل‌های رنگی گوناگونی را شامل گردد. برای مثال ممکن است دوربینی رنگ سفید را با ته‌مایه زرد، و دوربین دیگری رنگ سفید را با ته‌مایه آبی نشان دهد. روش‌هایی که در این گروه قرار می‌گیرند سعی در تشخیص و تعیین رابطه رنگی بین دو دوربین دارند. به عبارتی در این روش، هدف تعیین رابطه نگاشت رنگی بین دو دوربین است. بنابراین برای هر دو زوج دوربین، این تابع و رابطه، طی یک فرآیند یادگیری محاسبه می‌گردد. در [۳۴] تابع انتقال نوری یا BTF^1 برای یک زوج دوربین مشخص شده و منجر به بهبود در فرآیند بازشناسی گردیده است.

۲-۲-۱-۲- روش‌های یادگیری توصیفگر

در این دسته از روش‌ها، هدف بدست آوردن نگاشتی از ویژگی‌هاست که بیشترین تمایز را بین نمونه‌های مختلف افراد در سیستم بازشناسی ایجاد نماید. برای مثال این نگاشت می‌تواند بصورت انتخاب

¹ Brightness Transfer Function

زیرمجموعه‌ای از ویژگی‌ها یا وزن‌دهی آن‌ها بر اساس میزان اهمیت باشد. در [۳۵] به کمک بوستینگ تطبیقی از ترکیب وزن‌دار ویژگی‌ها به منظور تعیین شباهت بین دو فرد استفاده شده است. داده‌هایی که در این مقاله مورد استفاده قرار گرفته در زوایای مختلف بوده است. در الگوریتم بوستینگ تطبیقی هر ویژگی به‌عنوان یک کلاسه‌بند ضعیف در نظر گرفته می‌شود. در مرحله آموزش زوج تصاویر افراد به الگوریتم یادگیر داده می‌شود. بنابراین بردار ورودی شامل ویژگی‌های مربوط به دو فرد و خروجی شامل برچسب ۱ یا ۰-۱ می‌باشد. در صورتی که دو تصویر بردار ورودی، مربوط به یک شخص باشد خروجی، مقدار ۱ و در غیر این صورت دارای مقدار ۰-۱ است. در الگوریتم بوستینگ با استفاده از ویژگی‌هایی که خود به تنهایی خاصیت متمایزکنندگی قابل توجهی ندارند فضای ورودی کلاسه‌بندی می‌شود. در مرحله بعد بر اساس داده‌هایی که به‌غلط کلاسه‌بندی شده‌اند بهترین ویژگی که کلاسه‌بندی بهتری انجام دهد انتخاب شده و بر اساس میزان خطای کلاسه‌بندی، وزن‌دهی می‌گردد. از ترکیب این ویژگی‌های ضعیف، کلاسه‌بندی قوی ایجاد می‌شود. ویژگی‌های مورد استفاده در این مقاله عبارت‌اند از کانال‌های مدل‌های رنگی RGB، YCbCr و HSV و پاسخ فیلترهای بافتی نظیر اشمیت^۱ و گابور^۲. ویژگی‌های انتخابی وابسته به ناحیه در تصویر نیز می‌باشند.

در [۳۶] دو خاصیت بازگوکنندگی و متمایزکنندگی به‌صورت ترکیبی مورد استفاده قرار گرفته است. بدین ترتیب که به کمک توصیفگر کوواریانس بر روی مجموعه‌ای از ویژگی‌ها، براساس درخواست کاربر ۵۰ مورد ابتدایی را انتخاب و به کاربر جهت انتخاب نمایش می‌دهد. اگر فرد مذکور در ۵۰ نفر ابتدایی نبود از لیست مرتب‌شده افراد، تعدادی از ابتدای لیست را به‌عنوان مورد مثبت و تعدادی از انتهای لیست را به‌عنوان مورد منفی برچسب می‌زند. در ادامه به کمک الگوریتم بوستینگ، ۲۰ ویژگی متمایزکننده برای فرد استخراج شده و در نهایت لیست مذکور، مجدداً به‌وسیله مقدار اطمینان^۳ کلاسه‌بند

^۱ Schmid filter

^۲ Confidence

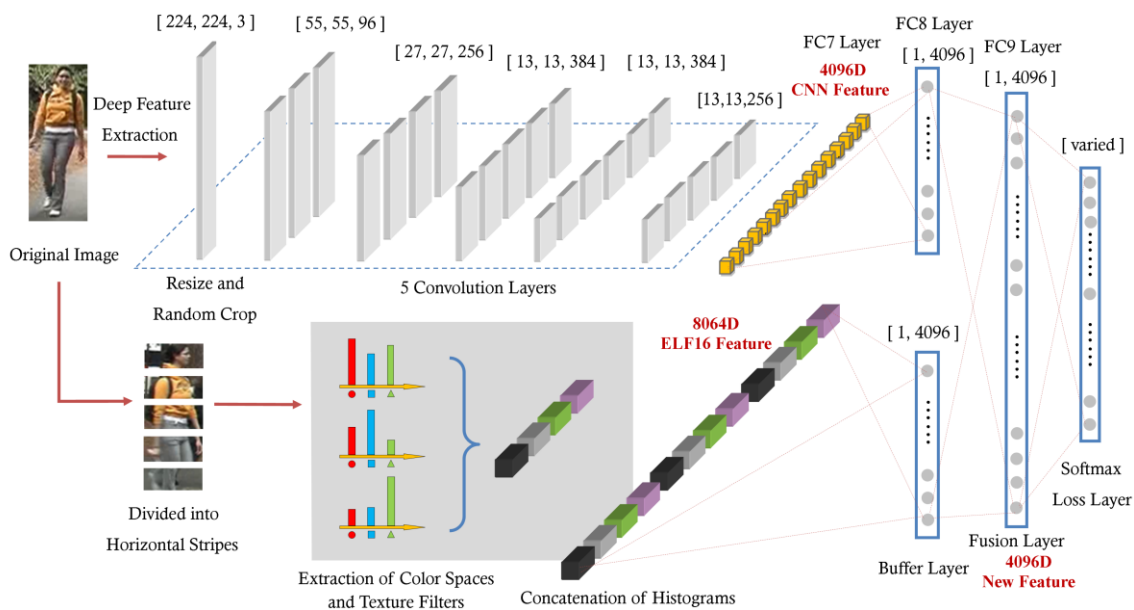
^۲ Gabor filter

مرتب می‌گردد. در [۳۷] نیز برای دنباله‌ای از تصاویر یک فرد، توصیفگر کوواریانس برای نواحی مختلف تصویر محاسبه می‌شود. سپس برای فریم‌های متوالی، مقدار میانگین ماتریس‌های کوواریانس متناظر با هر ناحیه محاسبه می‌گردد. در قدم بعدی به کمک دو روش، تکه‌هایی از یک فرد خاص، که قابلیت متمایزکنندگی بیشتری دارد مشخص می‌شود. در روش اول به کمک تعیین قابلیت اعتماد برای یک تکه، برپایه تعریف جدیدی از واریانس در فضای ریمانی و روش دوم به کمک الگوریتم بوستینگ تکه‌های متمایزکننده‌تر انتخاب می‌شود. همچنین در مقاله‌ای دیگر به کمک بوستینگ تطبیقی از بین ویژگی‌ها و ضرایب توصیفگر رنگ غالب^۱ (DCD) زیرمجموعه‌ای از ویژگی‌ها که هر فرد را از بقیه بهتر متمایز می‌کند، انتخاب می‌گردد [۳۸].

در [۳۹] توصیفگر جدیدی با استفاده از ترکیب شبکه عصبی کانولوشن و ویژگی ELF16^۲ [۳۵] معرفی شده است. نحوه ترکیب این دو بخش با استفاده از شبکه عصبی می‌باشد. در این ویژگی، تصویر شخص به ۱۶ سطر افقی بخش‌بندی می‌گردد. برای هر بخش هیستوگرام فضاهای رنگی RGB، HSV، Lab، XYZ، YCbCr و NTSC و ویژگی‌های بافتی شامل گابور و اشمیت و LBP محاسبه و سپس نرمال می‌گردند. ساختار شبکه مذکور بصورت شکل ۲-۱۷ می‌باشد.

^۱ Dominant Color Descriptor (DCD)

^۲ Ensemble of Localized Feature (ELF)

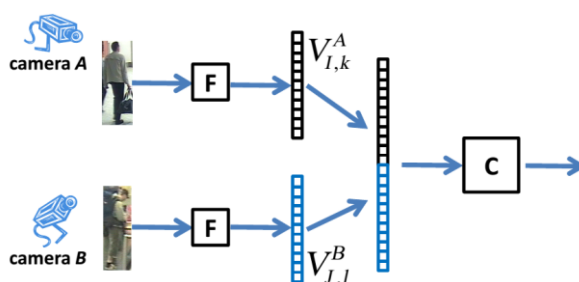


شکل ۲-۱۷: ساختار شبکه عصبی مورد استفاده در [۳۹] (شامل ویژگی ELF16 و ویژگی‌های حاصل از شبکه عصبی کانولوشن)

تعداد خروجی‌های این شبکه به تعداد افراد در سیستم است که برای یک شخص خاص، یکی از این خروجی‌ها مقدار یک به خود می‌گیرد. با استفاده از یک مجموعه آموزشی، بخش شبکه عصبی کانولوشن آموزش یافته و بدین ترتیب فرآیند استخراج ویژگی بصورت ضمنی در این بخش از شبکه صورت می‌گیرد. همچنین در لایه‌های پایانی، خروجی ویژگی ELF16 و شبکه عصبی کانولوشن با هم ترکیب می‌شوند. بعد از آموزش شبکه مذکور بر روی نمونه‌های آموزشی، لایه آخر از شبکه حذف می‌شود. در واقع توصیفگر نهایی، خروجی لایه ماقبل آخر از شبکه مذکور می‌باشد. در مرحله تطبیق از فرآیند بازشناسی، تصویر شخص و ویژگی ELF16 متناظر با آن، به شبکه اعمال گشته و خروجی لایه ماقبل آخر را به عنوان بردار ویژگی نهایی استفاده می‌کنیم. با این روش لایه‌های کانولوشن در بخش شبکه عصبی کانولوشن، ویژگی‌های مطلوب را استخراج نموده و در ویژگی نهایی قرار می‌گیرد.

یکی دیگر از رویکردهای استفاده شده، تبدیل مسئله بازشناسی در قالب یک مسئله کلاسه‌بندی می‌باشد. بدین ترتیب که برای جفت داده‌ها برای دو دوربین از یک فرد، مجموعه داده‌های آموزشی مثبت و جفت‌هایی از دو فرد در دو دوربین که مربوط به یک شخص نمی‌شوند مجموعه آموزشی

منفی ساخته می‌شود. در [۴۰] از این رویکرد استفاده شده و برای مجموعه ویژگی‌ها از هیستوگرام مدل رنگی HSV برای پنج ناحیه افقی در تصویر انسان استفاده شده است. بدین ترتیب کلاسه‌بندی که این چنین آموزش داده می‌شود به نوعی رابطه تابع انتقال بین یک جفت دوربین را مدل کرده و یاد می‌گیرد. شکل ۱۸-۲ نحوه آموزش کلاسه‌بند با استفاده از تصاویر ورودی را نشان می‌دهد.



شکل ۱۸-۲: ویژگی F بدست آمده از دو فرد برای دو دوربین A و B بهم متصل شده و بردار ساخته شده به کلاسه بند C ارسال می‌گردد [۴۰]

گروهی دیگر از روش‌ها بر پایه خصیصه‌های سطح بالاتر و معنایی‌تر ارائه شده است. این خصیصه‌ها در واقع نوعی از ویژگی‌های سطح میانی هستند که ویژگی‌های سطح پایین را به ویژگی‌های معنایی سطح بالاتر مرتبط می‌سازند. در واقع زمانی که ویژگی‌های کلی و تصویر کلی دو فرد مشابه هم باشند این خصیصه‌ها می‌توانند بین این چنین مواردی تمایز ایجاد کنند. این خصیصه‌ها می‌توانند به صورت بانظارت و یا به صورت بی‌نظارت مدل‌سازی شوند و از تصاویر افراد استخراج گردند. در حالت بانظارت برای مثال مجموعه‌ای از خصیصه‌ها از قبیل بلندقد بودن، کوتاه‌قد بودن، مو بلند بودن، جین‌پوش بودن و غیره می‌تواند مورد استفاده باشد. در حالت بی‌نظارت بر روی مجموعه بزرگی از تصاویر خوشه‌بندی صورت گرفته که بدین ترتیب هر خوشه به نوعی بیانگر خصیصه‌ای است. در مرحله بعدی به کمک این خصیصه‌ها نواحی از تصویر فرد که تمایز بیشتری را با بقیه افراد ایجاد می‌کند به عنوان نواحی بااهمیت در نظر گرفته می‌شوند.

در [۴۱] سعی در وزن‌دهی و اهمیت‌دهی در ویژگی‌های سطح پایین با توجه به خصیصه‌های استخراج شده، دارد. مراحل انجام الگوریتم ارائه شده بدین ترتیب است که ابتدا تصاویر برچسب نخورده

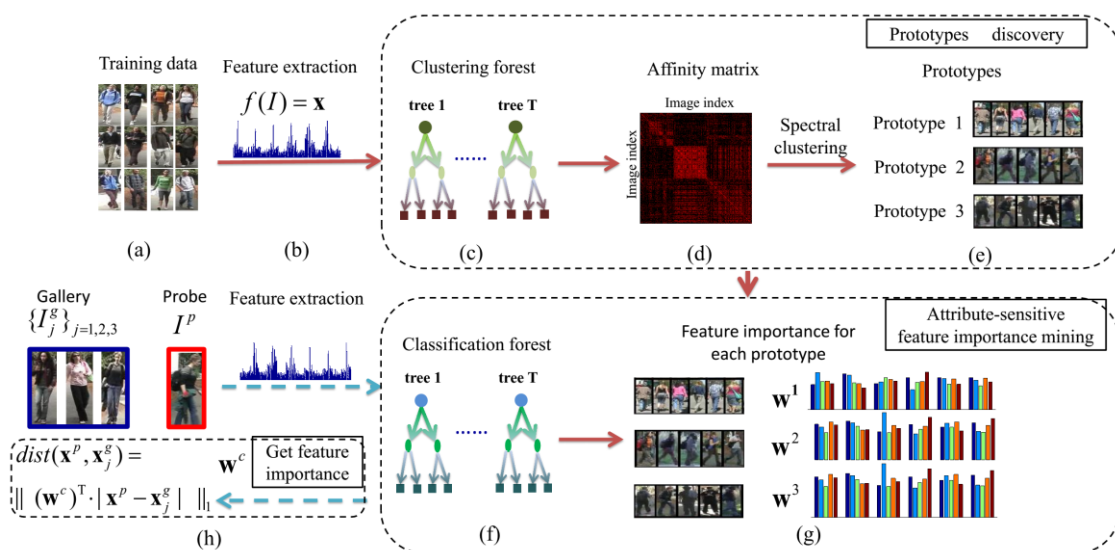
را خوشه‌بندی می‌کند. بدین ترتیب خروجی خوشه‌بندی، مجموعه‌ای از نماینده‌ها خواهد بود که هر کدام بازگوکننده خصیصه‌ای هستند. روش خوشه‌بندی صورت گرفته با استفاده از پارتیشن‌بندی گراف می‌باشد. بدین ترتیب که ابتدا شباهت جفت تصاویر را بدست آورده و به دنبال آن از خوشه‌بندی درختی استفاده می‌شود. سپس برای هر درخت یک ماتریس مجاورت ایجاد می‌شود که عناصر این ماتریس مجاورت بر اساس رابطه زیر محاسبه می‌گردند:

$$A_{ij}^t = \exp^{-d^t(x_i, x_j)} \quad (25-2)$$

$$d^t(x_i, x_j) = \begin{cases} 0 & \text{if } l(x_i) = l(x_j) \\ \infty & \text{otherwise} \end{cases}$$

که در آن $l(x_i)$ اندیس برگی از درخت است که فرد x_i در آن قرار گرفته است. عمل خوشه‌بندی درختی T بار انجام شده که در نهایت ماتریس شباهت کلی از میانگین T ماتریس شباهت حاصل می‌شود ($A = \frac{1}{T_{Cluster}} \sum_{t=1}^{T_{Cluster}} A^t$). در مرحله بعد ماتریس A به وسیله الگوریتم برش‌های نرمال^۱ به K پارتیشن تقسیم می‌گردد. در قدم بعدی الگوریتم باید برای هر نماینده یک بردار وزن متناظر با ویژگی‌های سطح پایین محاسبه گردد. وزن هر ویژگی بدین صورت محاسبه می‌شود که به کمک یک ویژگی داده‌های نماینده‌های مختلف را دسته‌بندی می‌کنیم. هرچه یک ویژگی بهتر بتواند یک نماینده خاص را از دیگر نماینده‌ها مجزا کند این ویژگی برای نماینده مذکور از اهمیت بالاتری برخوردار است. در مرحله بازشناسی زمانی که تصویر یک فرد جدید وارد سیستم می‌شود ابتدا خوشه یا نماینده تصویر ورودی محاسبه می‌گردد. سپس با استفاده از بردار وزنی محاسبه شده برای خوشه مذکور، تصویر ورودی وزن‌دهی می‌گردد. پس از اعمال این وزن‌دهی، تصاویر موجود در مجموعه گالری رتبه‌بندی می‌شوند. شکل ۱۹-۲ مراحل انجام این الگوریتم را نشان می‌دهد.

^۱ Normalized cuts algorithm



شکل ۲-۱۹: مراحل الگوریتم ارائه شده در مقاله [۴۱]

در [۴۲] با در نظر گرفتن ویژگی‌های سطح میانی که برای انسان نیز قابل فهم می‌باشند کیفیت بازشناسی را بهبود داده است. بدین ترتیب که ابتدا ۱۲ ویژگی توسط یک خبره تعیین شده و سپس با استفاده از یک پایگاه داده برای هر ویژگی یک کلاسه‌بند طراحی می‌گردد. برخی از ویژگی‌های مذکور عبارت‌اند از شلوار کوتاه، دامن پوش، موبلند، موکوتاه، قدبلند، قدکوتاه، حمل‌کننده یک شیء و کوله‌پشتی‌دار. سپس برای بازشناسی یک فرد جدید، با استفاده از هر یک از این کلاسه‌بندها مشخص می‌شود که فرد جدید کدام یک از خصیصه‌های مذکور را دارد. بدین ترتیب برای هر فرد برداری از وجود یا عدم وجود خصیصه‌ها را خواهیم داشت که با ترکیب برداری از مقادیر این کلاسه‌بندها و روش‌های گذشته کیفیت بازشناسی بهبود می‌یابد. از مزایای این روش این است که با استفاده از این کلاسه‌بندها در صورتی که کاربر انسانی در مورد فردی با ویژگی خاصی درخواستی مطرح کند سیستم قادر به ارائه پیشنهاداتی برای فرد مورد نظر می‌باشد.

در [۴۳، ۴۴] سعی می‌شود نواحی بااهمیت برای یک فرد استخراج گردد. روش کار بدین صورت است که با استفاده از مفهوم انطباق متراکم^۱، به کمک توصیفگر SIFT^۲ ابتدا با مقایسه تصویر با مجموعه

^۱ Dense correspondence

^۲ Scale-Invariant Feature Transform (SIFT)

تصادفی از تصاویر انسان، تکه‌هایی که در بیشتر از نیمی از افراد دیده نشده است را به‌عنوان نواحی بااهمیت مشخص کرده و درنهایت در رابطه‌ای ترکیبی و وزن‌دار، این نواحی بااهمیت را وزن‌دهی بیشتری کرده و نشان می‌دهد که با این روش نرخ تشخیص بالاتر خواهد رفت.

در [۴۵] یک تصویر بصورت افقی به چند سطح تقسیم می‌گردد. برای هر سطح نواحی با ابعاد 10×10 و اندازه گام ۵ استخراج می‌شود. در این مقاله هدف یادگیری نواحی هست که خاصیت بازنمایی و تمایز را بصورت همزمان داشته باشند. بدین علت نام نواحی مذکور را فیلترهای سطح میانی نام‌گذاری می‌کند. این نواحی الگوهای تصویری را شامل می‌شوند که می‌تواند بین افراد مختلف تمایز ایجاد کند. بردار ویژگی ساخته شده برای هر ناحیه یا فیلتر، ترکیبی از توصیفگر SIFT برای سه کانال فضای رنگی Lab و هیستوگرام در سه مقیاس می‌باشد. برای استخراج فیلترهایی با خاصیت متمایزکنندگی بیشتر، برای هر فیلتر میزان نزدیکترین فاصله تا فیلترهای متناظر در تصاویر مختلف محاسبه می‌گردد. بدین ترتیب فیلتری که به لحاظ فراوانی در مجموعه داده‌ها، دارای مقدار متوسطی باشد انتخاب می‌شود. بنابراین در مرحله تطبیق برای فیلترهای متناظر از دو تصویر، وزن‌دهی براساس میزان اهمیت فیلتر مذکور صورت می‌گیرد.

۲-۱-۲- روش‌های یادگیری معیار فاصله

در این دسته از روش‌ها بجای تمرکز بر روی انتخاب ویژگی مناسب، تمرکز بر روی یادگیری معیار فاصله‌ای است که دقت تطبیق را بدون در نظر گرفتن نحوه بازنمایی ظاهری، حداکثر نماید. درواقع نگاشتی از فضای ویژگی‌های اولیه حاصل می‌شود که در آن فاصله نمونه‌های یک کلاس تا حد امکان کم و فاصله نمونه‌های کلاس‌های متفاوت زیاد می‌باشد. به عبارتی هدف یادگیری و آموزش این نگاشت می‌باشد به‌نحوی که شرایط ذکرشده در بالا هرچه بیشتر برآورده گردد. این یادگیری می‌تواند به‌صورت بانظارت صورت گیرد. در مرحله آموزش زوج داده‌های مرتبط با یک فرد را به‌عنوان داده‌های کلاس مثبت و زوج داده‌های مرتبط با دو فرد متفاوت را به‌عنوان نمونه کلاس منفی در نظر می‌گیریم. فرض

کنید داده‌های مربوط به هر فرد با بردار m بعدی x نمایش داده شود. بدین ترتیب هدف یادگیری معیار فاصله‌ای است که با ماتریس $D \in R^{m \times m}$ نمایش داده می‌شود. با استفاده از این ماتریس فاصله دو نمونه x_i و x_j به صورت زیر محاسبه می‌گردد:

$$d(x_i, x_j) = (x_i - x_j)^T D (x_i - x_j) \quad (26-2)$$

به کمک تعریف فاصله به صورت بالا هدف کمینه کردن رابطه پایین به کمک برنامه‌نویسی محدب^۱ می‌باشد:

$$\min \sum_{(x_i, x_j) \in Pos} \|(x_i, x_j)\|_D^2 \quad s.t. D \succcurlyeq 0, \text{ and } \sum_{(x_i, x_j) \in Neg} \|(x_i, x_j)\|_D^2 \geq 1 \quad (27-2)$$

در این رابطه $\|\cdot\|_D^2$ برابر با رابطه ۲۶-۲ است. به عبارت دیگر در این رابطه می‌خواهیم ماتریس D را طوری بدست آوریم که مجموع فاصله زوج نمونه‌های مرتبط با یک فرد (Pos) کمینه و درعین حال مجموع فاصله زوج نمونه‌های غیر مرتبط (Neg) بزرگ‌تر مساوی یک گردد. نسخه دیگری از ایده مطرح‌شده در بالا در [۴۶] ارائه شده است. در این مقاله مجموعه بردارهای x را به صورت بردار تفاضل بین دو نمونه Z و Z' که بردارهای ویژگی متناظر برای دو شخص است، به صورت زیر تعریف می‌کند:

$$x = d(z, z') \quad z, z' \in R^q \quad (28-2)$$

تابع فاصله f را به صورت زیر بر اساس فاصله ماهالانوبیس^۲ داریم:

$$f(x) = x^T M x, \quad M \succcurlyeq 0 \quad (29-2)$$

هدف مینیم کردن رابطه زیر معرفی گردیده است:

$$\min_f r(f, \mathbb{O}) \quad r(f, \mathbb{O}) = -\log\left(\prod_{\mathbb{O}_i} C_f(x_i^p, x_i^n)\right) \quad (30-2)$$

^۱ Convex programming

^۲ Mahalanobis distance

که در آن مجموعه \mathbb{O} شامل دو مقدار x_i^p, x_i^n می باشد که اولی برای نمونه های غیر مرتبط و دومی برای دو نمونه مرتبط می باشد. در واقع هر عضو از مجموعه \mathbb{O} حاوی اطلاعات سه فرد است: فاصله فرد اول با فرد دوم که برچسب یکسانی دارند و فاصله فرد اول با فرد سوم که برچسب غیر یکسانی دارند. تابع C_f به صورت تابع لجستیک می باشد به صورت زیر تعریف شده است:

$$C_f(x_i^p, x_i^n) = P(f(x_i^p) < f(x_i^n)) = (1 + \exp\{f(x_i^p) - f(x_i^n)\})^{-1} \quad (31-2)$$

در [47] به نوعی از فاصله بین نمونه ها در فرآیند یادگیری استفاده شده است. در این مقاله سعی شده از داده های برچسب نخورده برای بازشناسی استفاده شود. از آنجایی که مجموعه افراد مشابه، بر روی یک خمینه قرار دارند، با استفاده از گراف و همسایگی داده ها، می توان برچسب را در فضای گراف منتشر کرده و نزدیک ترین فرد را در مجموعه گالری استخراج نمود. در واقع داده های برچسب نخورده فضای خمینه را پر کرده و در نتیجه مزیت منتشر شدن برچسب از طریق آن ها را برای ما فراهم می سازد. نسخه دیگری از این گروه در قالب یک مسئله رتبه بندی است. در [48] فرض شده مجموعه آموزشی، شامل تصاویر تهیه شده از یک فرد در دو دوربین مختلف به صورت $X = \{(x_i, y_i)\}_{i=1}^m$ باشد که در آن x_i بردار d بعدی ویژگی ها و y_i برچسب نمونه i می باشد. دو مجموعه d_i^+ و d_i^- به ترتیب نمونه های هم برچسب با نمونه x_i و نمونه های ناهم برچسب با نمونه x_i می باشند.

$$d_i^+ = \{x_{i,1}^+, x_{i,2}^+, \dots, x_{i,m^+(x_i)}^+\} \quad d_i^- = \{x_{i,1}^-, x_{i,2}^-, \dots, x_{i,m^-(x_i)}^-\} \quad (32-2)$$

که برای مقادیر $m^+(x_i)$ و $m^-(x_i)$ داریم $m^-(x_i) = m - m^+(x_i) + 1$ که در آن $m^+(x_i) \ll m$ می باشد. دلیل آن این است که نمونه های هم برچسب تعداد کمتری نسبت به نمونه های ناهم برچسب دارند. هدف در این رویکرد، طراحی تابع رتبه بندی δ است به طوری که برای نمونه های مرتبط مقدار بزرگ تری نسبت به نمونه های غیر مرتبط نتیجه دهد. تابع δ نیز به صورت زیر تعریف می شود که حاصل ضرب داخلی بردار وزن w در قدر مطلق بردار تفاضل x_i و $x_{i,j}$ می باشد.

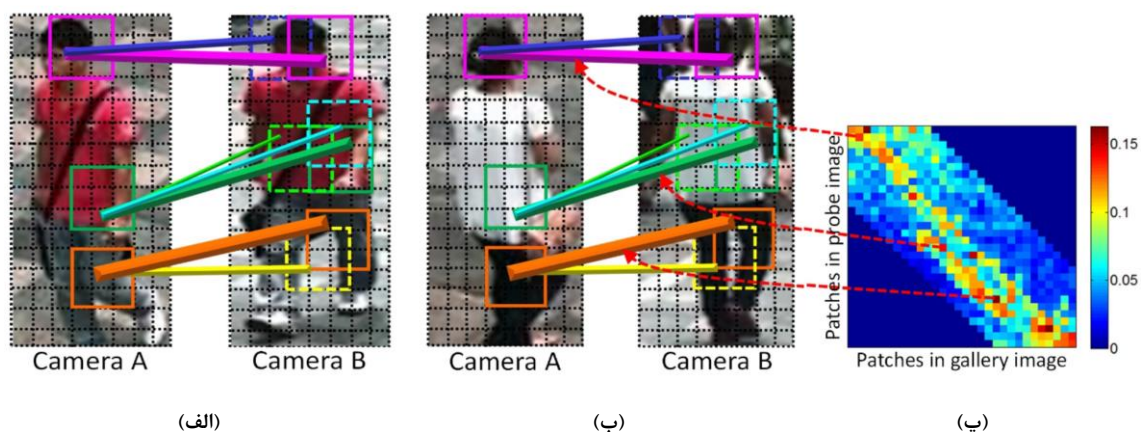
$$\delta(x_i, x_{i,j}^+) > \delta(x_i, x_{i,j}^-) \quad \delta(x_i, x_{i,j}) = w^T |x_i - x_{i,j}| \quad (33-2)$$

برقراری نامساوی ذکرشده در بالا منجر به حل مسئله‌ای مشابه ماشین بردار پشتیبان خواهد گردید که در آن بردار W برای کلیه نمونه‌های آموزشی طوری تعیین می‌شود که نامساوی بالا برقرار شود. رابطه بهینه‌سازی در قالب ماشین بردار پشتیبان به صورت زیر خواهد بود:

$$w = \arg \min_w \frac{1}{2} \|w\|^2 + C \sum_{s=1}^{|P|} \ell(0, 1 - w^T (\hat{x}_s^+ - \hat{x}_s^-))^2 \quad (34-2)$$

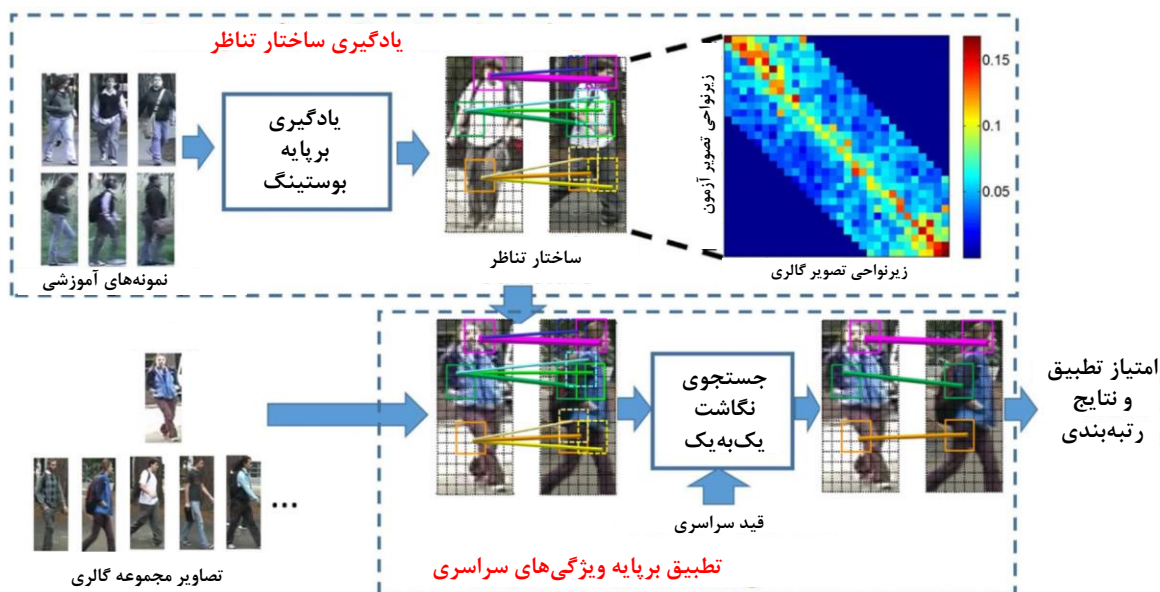
که در آن $\hat{x}_s^+ = |x_i - x_{i,j}^+|$ و $\hat{x}_s^- = |x_i - x_{i,j}^-|$ می‌باشد. در ادامه برای حل مشکل پیچیدگی زمانی داده‌های آموزشی را به چند مجموعه مجزا تقسیم نموده برای هر یک رتبه‌بند ضعیف طراحی کرده و در نهایت از ترکیبی از چند رتبه‌بند ضعیف و تشکیل یک رتبه‌بند قوی، استفاده نموده است. بدین ترتیب برای یک زوج دوربین معیار فاصله‌ای آموزش می‌یابد که می‌تواند نگاهی از فاصله زوج تصاویر افراد ایجاد کند که در آن مقدار بیشتر، فرد مرتبط‌تری را پیشنهاد خواهد داد.

در [49] تناظر مکانی بخش‌های تصویر برای جفت دوربین‌ها طی یک فرآیند یادگیری مدل می‌گردد. به دلیل پیکربندی و موقعیت قرارگیری دوربین‌ها در محیط، تشابهی بین تصاویر یک دوربین خاص از منظر نحوه قرارگیری شخص در تصویر وجود دارد. اگر تصاویر مربوط به هر دوربین را به زیرنواحی کوچکتری بخش‌بندی کنیم خواهیم دید که برای یک جفت دوربین، تناظری بین این نواحی به لحاظ شباهت وجود دارد که تقریباً برای تمامی تصاویر ثبت‌شده یکسان می‌باشد. برای مثال در شکل ۲-۲۰ تناظر بین زیرنواحی برای یک جفت دوربین نشان داده شده است.



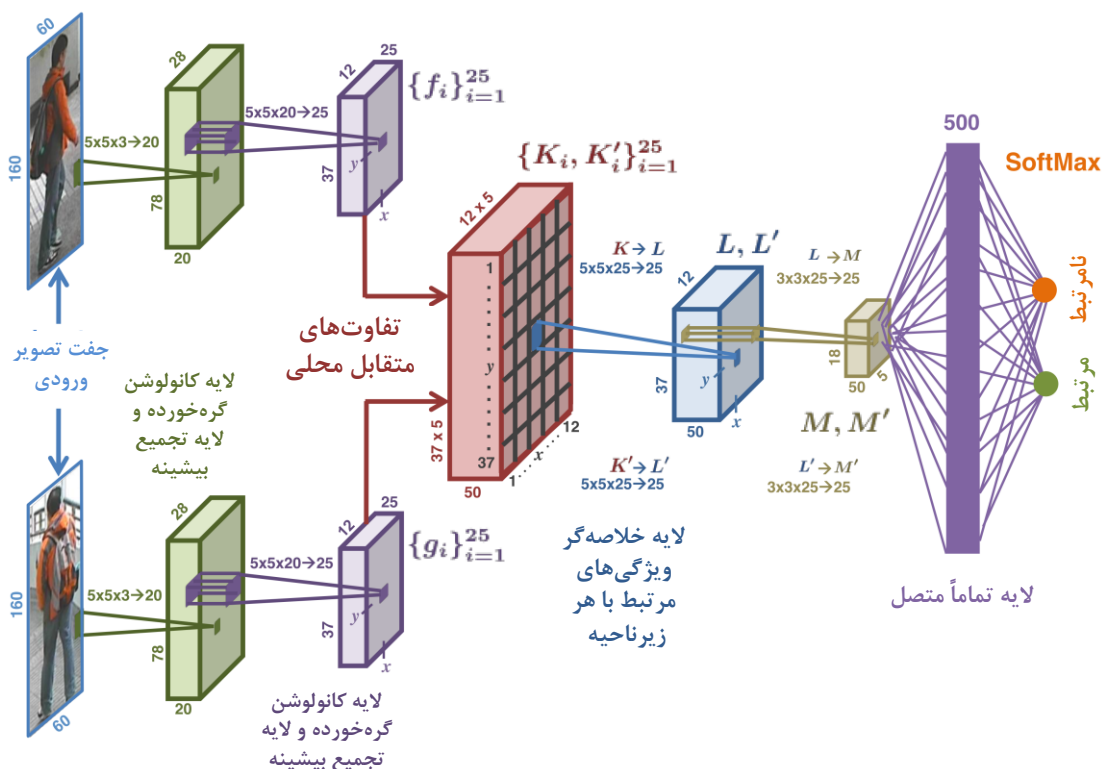
شکل ۲-۲: تناظر مکانی زیرنواحی تصاویر ثبت شده برای یک جفت دوربین، الف و ب: تناظر زیرنواحی برای تصاویر ثبت شده از دو شخص در دو دوربین، پ: ماتریس تناظر ساختاری برای یک جفت دوربین [۴۹]

در این مقاله ماتریسی تحت عنوان ماتریس تناظر ساختاری برای استخراج احتمال تناظر و تطبیق بین هر دو زیرناحیه برای یک جفت دوربین طی یک فرایند یادگیری ساخته می‌شود. در مرحله تست، برای یک جفت تصویر از یک جفت دوربین، تطبیق زیرنواحی دو تصویر بر اساس میزان تناظر بدست آمده از ماتریس تناظر صورت می‌گیرد. شکل ۲-۲۱ خلاصه‌ای از الگوریتم پیشنهادی در [۴۹] را نشان می‌دهد.



شکل ۲-۲۱: خلاصه‌ای از مراحل آموزش و تست در روش پیشنهادی [۴۹]

در [۵۰] یک معماری جدید کانولوشنی عمیق معرفی شده است. در این معماری دو تصویر به عنوان ورودی شبکه داده می‌شود و خروجی شبکه ارتباط بین دو تصویر را در دو کلاس نشان می‌دهد. کلاس اول یکی بودن دو تصویر و کلاس دوم متفاوت بودن تصاویر را نشان می‌دهد. از آنجایی که خروجی شبکه مقدار حقیقی است می‌توان از آن به عنوان معیار فاصله نیز استفاده نمود. ساختار شبکه کانولوشنی مذکور به صورت شکل ۲-۲۲ می‌باشد.



شکل ۲-۲۲: ساختار شبکه کانولوشنی عمیق پیشنهادی در [۵۰]

ورودی این شبکه دو تصویر با ابعاد $3 \times 160 \times 60$ در فضای رنگی RGB می‌باشد. در اولین لایه که لایه کانولوشن است تعداد ۲۰ فیلتر با ابعاد $3 \times 5 \times 5$ آموزش می‌یابد. منظور از کانولوشن گرہ خورده^۱ این است که برای دو تصویر ورودی، فیلترهای مشترکی آموزش یافته و در مرحله تست اعمال می‌گردند. در ادامه با اعمال لایه جمع بیشینه^۲، ابعاد نقشه ویژگی‌ها^۳ به نصف کاهش می‌یابد. در لایه بعد

^۱ Tied convolution layer

^۲ Feature map

^۳ Max pooling layer

کانولوشن گره خورده‌ای با ۲۵ فیلتر در ابعاد $5 \times 5 \times 20$ آموزش می‌یابد. مجدداً لایه تجمیع بیشینه ابعاد نقشه ویژگی‌ها را به نصف کاهش می‌دهد. در این مرحله برای هر تصویر ورودی ۲۵ نقشه ویژگی با ابعاد 12×37 داریم. در بخش بعد تفاضل هر درایه از هر نقشه ویژگی، با مقادیر همسایگی با ابعاد 5×5 در درایه متناظر در تصویر متقابل محاسبه می‌گردد. همین عمل بصورت معکوس محاسبه شده و خروجی لایه از کنار هم قرار دادن دو آرایه K_i و K'_i حاصل می‌شود. لایه بعدی لایه کانولوشن با اندازه گام ۵ می‌باشد که در واقع اطلاعات تفاضل دو درایه نقشه ویژگی‌ها را بصورت خلاصه در یک درایه تجمیع می‌کند. در لایه کانولوشن بعدی ارتباط مکانی بین مقادیر تفاضل بوسیله ۲۵ فیلتر یاد گرفته می‌شود. در نهایت خروجی این لایه بصورت تماماً متصل به دو نورون خروجی متصل می‌شود. تابع اتلاف در لایه خروجی تضمین می‌کند که مجموع مقادیر دو خروجی برابر با یک باشد. بدین ترتیب در طول این ساختار، از ویژگی‌های سطح پایین به سمت ویژگی‌های انتزاعی‌تر پیش می‌رویم. و در لایه تفاضل ارتباط بین ویژگی‌های استخراج شده برای دو تصویر و نوع کلاس این زوج تصویر، مشخص می‌گردد.

۲-۲- روش‌های مبتنی بر اطلاعات پیکربندی دوربین‌ها

این روش‌ها سعی دارند با استفاده از منابع اطلاعاتی غیر از ظاهر افراد، فضای جستجو را محدود نمایند. منظور از فضای جستجو، تصاویر افرادی است که در مجموعه گالری وجود دارد و در نتیجه هر یک کاندیدی برای بازشناسی می‌باشند. برای مثال با در دست داشتن ارتباط مکانی دوربین‌های نصب شده و بر اساس یک توزیع احتمالاتی، می‌توان حدس زد فرد خاصی که به میدان دید یک دوربین وارد شده احتمالاً در لحظاتی قبل در کدام دوربین‌ها می‌توانسته ظاهر شود. بدین ترتیب تعداد افراد کاندید با توجه به قید جدید محدودتر شده و دقت بازشناسی بیشتر می‌گردد.

۲-۲-۱- استخراج موقعیت قرارگیری نسبی دوربین‌ها

یکی از مواردی که منجر به پیچیدگی بازشناسی می‌شود وجود دوربین‌های متعدد با دامنه دید ناهمپوشان است. وجود چنین شرایطی گستردگی مشاهدات را در زمان و مکان‌های گوناگون در دوربین‌های مختلف دربر خواهد داشت. قاعده‌ای که به‌طور معمول می‌توان فرض کرد این است که اشیاء متحرک در محیط‌هایی مانند بزرگراه، راهروهای اماکن و غیره تمایل به عبور از مسیرهای مشترکی در طول زمان دارند. با استفاده از این فرض می‌توان فضای جستجو را در مسئله بازشناسی محدود نمود. الگوریتم‌های ارائه شده در این دسته، سعی در یادگیری ارتباط زمانی و مکانی بین دوربین‌ها در قالب تخمین یک تابع چگالی احتمال چند متغیره را دارند.

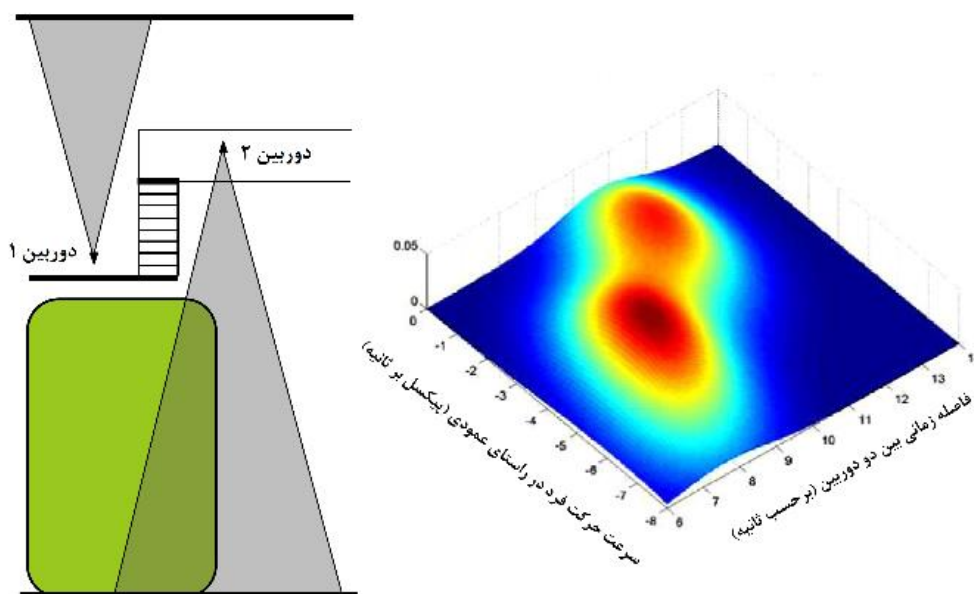
در [۵۱] فرض‌های مذکور را در قالب توابع توزیع احتمالاتی بیان نموده و در مرحله بعد به کمک داده‌های آموزشی و با استفاده از روش پنجره پارزن^۱ توابع توزیع مذکور را تخمین زده است. این توابع توزیع احتمالاتی چندمتغیره، ارتباطات زیر را مدل می‌کنند:

- ارتباط بین زمان انتقال و اندازه و جهت حرکت اشیاء برای زوج دوربین‌ها.
- ارتباط بین زمان انتقال و موقعیت خروج از یک دوربین و موقعیت ورود به میدان دید دوربین دیگر.
- وابستگی بین موقعیت ورود و موقعیت خروج برای زوج دوربین‌ها.

بدین ترتیب با تخمین این توابع توزیع، می‌توان فضای جستجوی مسئله بازشناسی را محدود نمود و تصاویر کاندید کمتری را در نظر گرفت. شکل ۲-۲۳ تابع توزیع احتمال دومتغیره برای ارتباط بین سرعت خروج فرد از دوربین ۲ و زمانی که طول می‌کشد که همان فرد به دوربین ۱ وارد شود را نشان می‌دهد. نواحی قرمز رنگ در این نمودار مناطقی که احتمال وقوع بیشتری دارد را نشان می‌دهد. این

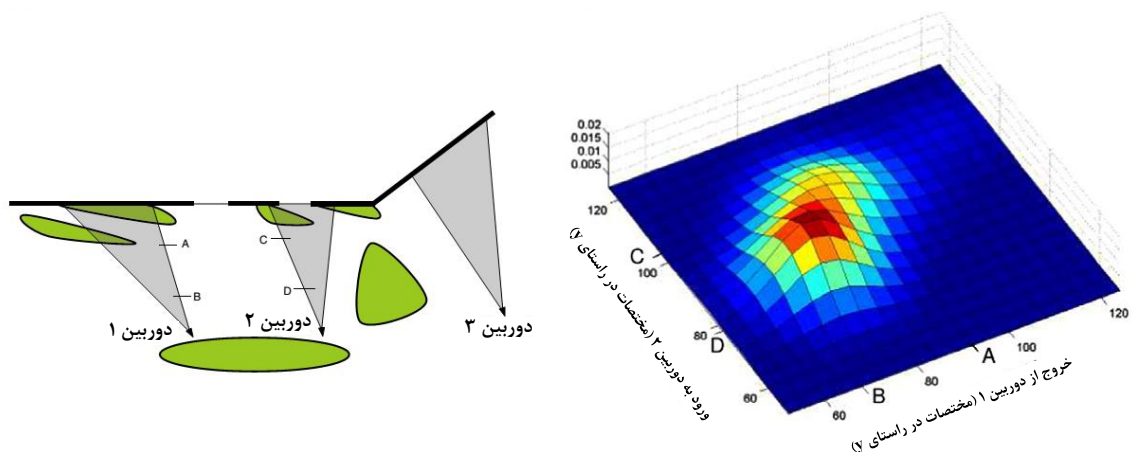
^۱ Parzen window

نواحی نشان می‌دهد که هرچه سرعت خروج فرد از دوربین ۲ بیشتر باشد زمانی که طول می‌کشد که در دوربین ۱ ظاهر شود کمتر خواهد بود یا به عبارتی برای سرعت‌های زیاد و زمان انتقال کم، تابع چگالی، مقدار بیشتری را نشان می‌دهد.



شکل ۲-۲۳: تابع توزیع تخمین شده سمت راست ارتباط بین سرعت خروج فرد از دوربین ۲ و زمانی که طول می‌کشد که همان فرد به دوربین ۱ وارد شود (مستطیل سبز فضای سبزی را نشان می‌دهد که افراد کمتر وارد آن می‌شوند) [۵۱]

شکل ۲-۲۴ نیز ارتباط بین موقعیت در راستای y برای خروج از دوربین ۱ و ورود به دوربین ۲ را نشان می‌دهد. همان‌طور که مشاهده می‌کنیم برای نقطه خروج A از دوربین ۱ نقطه C محتمل‌ترین نقطه برای ورود به دوربین دوم می‌باشد و همچنین برای نقطه خروج B از دوربین ۱ نقطه D محتمل‌ترین نقطه برای ورود به دوربین دوم می‌باشد که نشان‌دهنده تمایل افراد به حرکت مستقیم در محیط می‌باشد.



شکل ۲-۲۴: ارتباط بین موقعیت در راستای γ برای خروج از دوربین ۱ و ورود به دوربین ۲ (نواحی سبزرنگ فضای سبز را در محیط مذکور نشان می‌دهد) [۵۱]

۲-۲-۲- هم‌تراز کردن دوربین‌ها

در این دسته از روش‌ها معمولاً دوربین‌های نصب‌شده در محیط دارای میدان دید همپوشان هستند. بدین ترتیب می‌توان هموگرافی^۱ بین دوربین‌ها را استخراج نمود. منظور از هموگرافی استخراج نگاهی است که تحت نگاهت مذکور، یک صفحه رؤیت شده در یک دوربین به تصویر همان صفحه در دوربین دوم تبدیل می‌شود. از این عمل تحت عنوان هم‌ترازی^۲ نیز یاد می‌شود. پس از هم‌ترازی دوربین‌ها می‌توان خصوصیات یکتایی از اشیاء رؤیت شده استخراج نمود. این خصوصیات می‌تواند موقعیت سه‌بعدی شیء در محیط، مدل سه‌بعدی شیء و ارتفاع شیء باشد.

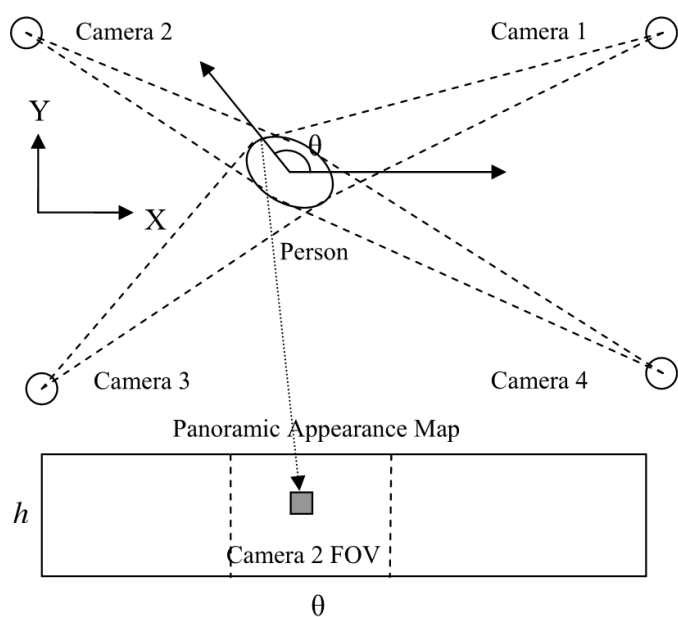
در [۵۲] با در دست داشتن دنباله‌ای از تصاویر فرد از چند زاویه دید، مدلی تمام‌نما^۳ ساخته می‌شود. به عبارتی نقشه تمام‌نمای ظاهری بر اساس اطلاعات بدست آمده از دوربین‌هایی که دارای همپوشانی هستند به‌عنوان ویژگی فرد در نظر گرفته شده است. بدین ترتیب که ابتدا مدل سه‌بعدی فرد ایجاد و از روی آن تصویری بر روی یک استوانه نگاهت می‌شود که در واقع اطلاعات کلیه دوربین‌ها را به‌طور توأمان دارا است. درنهایت برای مقایسه تصویر جدید، مقایسه بر روی هیستوگرام مدل و

^۱ Homography

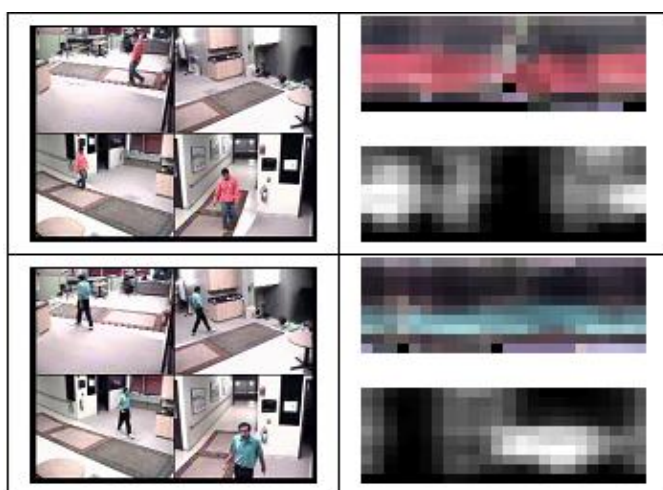
^۲ Panoramic model

^۳ Calibration

هیستوگرام مورد آزمون صورت می گیرد. شکل ۲-۲۵ مثالی از شخصی است که بین چهار دوربین هم‌تراز شده قرار گرفته است. با استفاده از اطلاعات زاویه برای هر دوربین، مدل پانورامیک استوانه‌ای شخص استخراج می‌شود. شکل ۲-۲۶ نیز نقشه پانورامیک بدست آمده را با الگوریتم پیشنهادی در مقاله مذکور برای دو شخص خاص نشان می‌دهد.

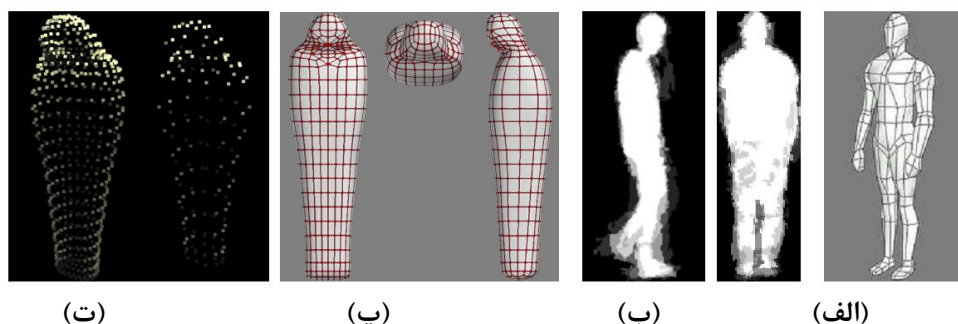


شکل ۲-۲۵: نقشه پانورامیک استخراج شده از تصاویر بدست آمده از چند دوربین با میدان دید مشترک [۵۲]



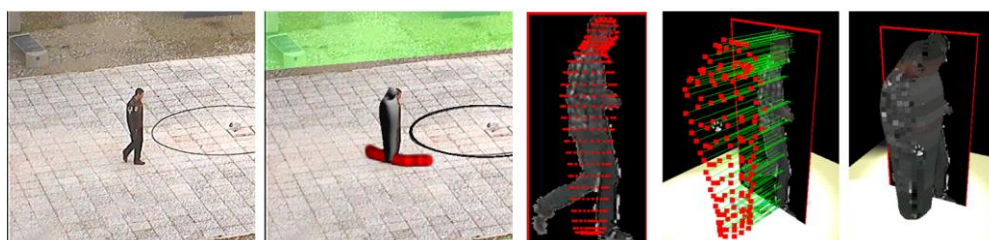
شکل ۲-۲۶: نقشه پانورامیک بدست آمده برای دو شخص خاص [۵۲]

در [۵۳] نیز با استفاده از دوربین‌های هم‌تراز شده اقدام به تشکیل مدلی سه‌بعدی از انسان می‌نماید. پس از تشخیص موقعیت افراد در دوربین‌ها با توجه به مسیر حرکت افراد، جهت، اندازه و موقعیت هر یک مشخص می‌شود. یکی از ایده‌های مطرح‌شده در این مقاله ارائه مدلی سه‌بعدی و ساده‌شده از پیکر انسان می‌باشد که در عین سادگی بسیاری از خصوصیات سه‌بعدی انسان را داراست. شکل ۲۷-۲ مدل پیشنهادی را نشان می‌دهد.



شکل ۲۷-۲: مدل سه‌بعدی استخراجی برای انسان در مرجع [۵۳]. تصویر الف: مدل سه‌بعدی انسان به همراه جزئیات، تصویر ب: میانگین سیلوئت انسان در دو زاویه مختلف، تصویر پ: مدل سه‌بعدی ساده‌شده پیشنهادی، تصویر ت: نمونه‌برداری از رئوس مورد نیاز در مدل ساده‌سازی شده

پس از تعیین جهت شخص، مدل سه‌بعدی بر روی آن قرار گرفته و سپس پیکسل‌های تصویر فرد در جهت مناسب بر روی رئوس مدل، تصویر و انعکاس داده می‌شود. مراحل این فرآیند در شکل ۲۸-۲ نشان داده شده است.



شکل ۲۸-۲: فرآیند انطباق تصویر فرد بر روی مدل سه‌بعدی پیشنهادی [۵۳]

از ویژگی‌های دیگر این الگوریتم، به‌روزرسانی تدریجی مدل سه‌بعدی است. این به‌روزرسانی می‌تواند منجر به تکمیل مدل از زوایای مختلف گردد و یا منجر به وضوح بیشتر در ناحیه خاصی از مدل

شود. همچنین در مدل پیشنهادی برای هر رأس از مدل، میزان قابلیت اعتمادی تعریف می‌گردد که در مرحله بازشناسی از آن استفاده می‌شود. شکل ۲-۲۹ مدل بدست آمده را برای سه نمونه انسان نشان می‌دهد.



شکل ۲-۲۹: مدل ساخته شده برای چند نمونه انسان با استفاده از مدل سه‌بعدی پیشنهادی در [۵۳]

در مرحله بازشناسی برای دو فرد، مدل‌های دو فرد مورد مقایسه و تطبیق قرار می‌گیرد. بدین ترتیب برای دنباله‌ای از تصاویر یک فرد، یک مدل خواهیم داشت که برآیندی از کلیه فریم‌های مرتبط با فرد مذکور می‌باشد. از معایب این روش نیازمندی روش مذکور به تصاویری با کیفیت بالاست. همچنین انطباق بر مدل سه‌بعدی مذکور، فرآیندی زمان‌بر می‌باشد که در نتیجه استفاده از آن را در کاربردهای بلادرنگ محدود می‌سازد.

۲-۳- جمع‌بندی

در این فصل مروری بر انواع روش‌های بازشناسی در سیستم‌های نظارت ویدئویی داشتیم. در دسته بندی مطرح شده دیدیم که اغلب روش‌های بازشناسی در دسته روش‌های مبتنی بر اطلاعات تصاویر قرار می‌گیرند. از آنجایی که رویکرد پیشنهادی در این رساله نیز در این دسته قرار دارد، تمرکز اصلی مطالعات خود را بر این روش‌ها بنا نهادیم. در سیستم بازشناسی همانند هر سیستم شناسایی، مرحله استخراج ویژگی از اهمیت ویژه‌ای برخوردار است. در این راستا تحقیقات بسیار گوناگونی در جهت استخراج ویژگی‌های مناسب برای این کاربرد، انجام شده است. بخش دیگری از تحقیقات علاوه بر استخراج ویژگی مناسب، سعی در یادگیری و مدل‌سازی نگاشت بین دوربین‌ها دارند. دستیابی به یک

نگاشت مناسب می‌تواند بر دقت بازشناسی بیافزاید. رویکردهای ترکیبی نیز مورد توجه محققین در این حوزه می‌باشد. با بررسی مزایا و معایب این تحقیقات در می‌یابیم که مسئله چرخش زاویه‌ای شخص در تصویر، به صورت خاص مورد توجه قرار نگرفته است. به عبارتی تحقیقات مختلف بیشتر سعی دارند ویژگی‌هایی ارائه نمایند که نسبت به انواع تغییرات ظاهری از جمله تغییرات زاویه‌ای مقاوم باشند. یا با یادگیری نگاشت بین دو دوربین مسئله چرخش زاویه‌ای نیز تا حدی برطرف گردد. در واقع روش‌های یادگیری نگاشت بین دوربین‌ها به صورت ضمنی فرض می‌کنند که زاویه قرارگیری اشخاص برای یک دوربین خاص مشابه می‌باشد. این فرض در دنیای واقعی همیشه برقرار نیست و در یک سیستم نظارت ویدئویی، زاویه قرارگیری شخص در برابر یک دوربین خاص می‌تواند بیش از یک حالت باشد. برای برطرف نمودن این فقدان می‌توان از اطلاعات زاویه‌ای در هر عکس استفاده نمود. این اطلاعات به سادگی از تصویر قابل استخراج بوده و با استفاده از این خصیصه سه‌بعدی می‌توان مشکلات ایجاد شده از چرخش زاویه‌ای را به شکلی کاملاً آگاهانه برطرف نمود.

فصل ۳. روش پیشنهادی

۳-۱- مقدمه

در فصل اول انواع مشکلاتی که فرآیند بازشناسی را دشوار می‌سازد برشمردیم. در میان انواع مشکلات، چرخش زاویه‌ای شخص نسبت به دوربین از مواردی است که کمتر مورد توجه قرار گرفته است. اغلب مجموعه داده‌های استاندارد که به هدف بازشناسی تهیه شده‌اند و در دسترس محققین می‌باشند شامل چرخش زاویه‌ای شدیدی نیستند. درحالی‌که وجود تنوع زاویه‌ای در تصاویر نسبت به دوربین‌ها، در یک محیط واقعی، کاملاً بدیهی و امکان‌پذیر می‌باشد. در واقع محل نصب دوربین‌های نظارتی می‌تواند به گونه‌ای باشد که ثبت تصاویر در زوایای مختلف امکان‌پذیر شود. اغلب روش‌های کنونی سعی در ارائه توصیفگر یا یادگیری معیار فاصله‌ای دارند که بر انواع مشکلات غلبه نمایند و تمرکز ویژه‌ای بر روی مسئله چرخش زاویه‌ای ندارند. توصیفگرهایی از قبیل SIFT یا توصیفگر کوواریانس گرچه به چرخش سطحی مقاوم هستند اما در حالت چرخش عمقی هیچ‌یک از این توصیفگرها قادر به بازنمایی تغییرات نیستند. در فصل مروری بر کارهای پیشین به روش‌هایی اشاره کردیم که اجزای بدن را استخراج می‌کنند. این قبیل روش‌ها در صورتی که اجزای بدن را بصورت کاملاً دقیق استخراج نمایند می‌توانند بر مشکل زاویه دید غلبه نمایند. استخراج دقیق اندام مستلزم وجود تصاویر با کیفیت بالا بوده و همچنین سربار محاسباتی بالایی نیز دربر خواهد داشت. لذا مصالحه‌ای بین سرعت اجرا و تحلیل اجزای تصویر وجود خواهد داشت. از آنجایی که استخراج اندام بدن نیز با خطاهای فراوانی همراه است، مشکل چرخش زاویه همچنان به قوت خود باقی خواهد ماند. روش پیشنهادی در این رساله ضمن در نظر گرفتن اطلاعات زاویه‌ای از پردازش‌های اضافی که سعی در تحلیل اندام شخص دارند دوری نموده و بازنمایی از ظاهر شخص در زوایای مختلف ارائه خواهد نمود.

یکی از ویژگی‌های پایه‌ای در بازشناسی، ویژگی‌های مرتبط با رنگ و بافت در تصویر می‌باشد. ویژگی‌های سه‌بعدی بدن انسان به گونه‌ای است که تصویر یک شخص در زوایای مختلف، ترکیب‌های ظاهری متفاوتی ایجاد می‌کند. در واقع یک شخص زمانی که در زاویه‌های مختلف مقابل دوربین قرار

می‌گیرد ممکن است بخشی از پوشش آن آشکار یا نهان گردد. برای روشن تر شدن این مسئله شکل ۱-۳ سه جفت تصویر مرتبط با سه شخص در دو دوربین مختلف را نشان می‌دهد. تصاویر مذکور از مجموعه داده‌های ViPER انتخاب شده است [۸].

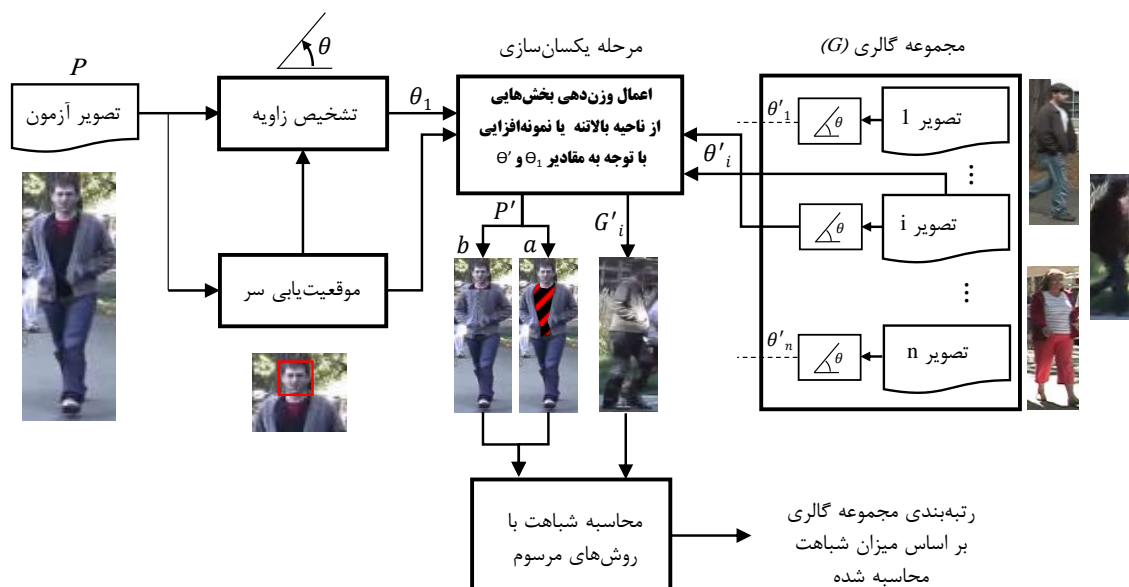


شکل ۱-۳: شش تصویر مربوط به سه شخص از مجموعه داده‌های ViPER

تصاویر (ث) و (ج) مربوط به یک فرد از دو زاویه مختلف می‌باشد. با توجه به این دو تصویر می‌بینیم که پیراهن سفیدرنگ شخص مذکور در حالت پهلو، به دلیل شکل سه‌بعدی بدن انسان، کمتر رؤیت می‌گردد. همین مشکل در تصویر فرد بعدی که پیراهن قرمز به تن دارد مشاهده می‌گردد. در تصویر از زاویه پهلو، پیراهن قرمز فرد، مخفی شده است. در مورد فرد سوم نیز زاویه پشت شخص، بعضی از خصوصیات زاویه روبرو را در خود ندارد. اکثر روش‌های موجود در چنین حالاتی قادر به تطبیق دو نمونه نیستند یا با عدم دقت همراه می‌شوند. حتی روش‌هایی که سعی در استخراج نقاط کلیدی دارند در چنین مواردی دچار مشکل می‌گردند.

بررسی‌های صورت گرفته نشان می‌دهد که این مشکل بیشتر در پوشش بالاتنه شخص ایجاد می‌شود. فرض کنید که تصویر آزمون، تصویری از روبرو همانند مورد ت و تصاویر موجود در سیستم برای بازشناسی موارد ث، پ و الف باشند. در چنین حالتی ناظر انسانی با توجه به دانش پیشینی که از شکل سه‌بعدی بدن انسان در زوایای مختلف دارد به خوبی عمل تطبیق و بازشناسی را انجام می‌دهد.

بدین ترتیب که این ناظر حدس خواهد زد که این شخص در تصویر از زاویه پهلو به احتمال زیاد فاقد بخشی با طیف رنگی قرمز خواهد بود. همچنین در پوشش اغلب انسان‌ها نظم و قاعده مشترکی وجود دارد که پایه قضاوت ناظر انسانی خواهد بود. از این قواعد می‌توان به امتداد رنگی پوشش پشت فرد به نواحی جلو و آستین‌های فرد اشاره نمود. سیستم پیشنهادی در این رساله نیز سعی در مدل‌سازی این دانش پیشین دارد. شکل ۲-۳ نمای کلی سیستم پیشنهادی را نشان می‌دهد.



شکل ۲-۳: نمایی کلی از سیستم پیشنهادی به همراه زیر سیستم‌ها

همان‌طور که پیش‌تر بیان شد ورودی یک سیستم بازشناسی، تصویری است (تصویر آزمون) که قصد داریم مجموعه تصاویر موجود در پایگاه (گالری) را مورد جستجو قرار دهیم و تصویر متناظر ورودی را بازشناسی نماییم. بدین منظور بردار ویژگی برای کلیه تصاویر استخراج شده و در فضای ویژگی‌ها نزدیک‌ترین نمونه یا تصویر را از مجموعه گالری انتخاب می‌کنیم. بر این اساس گام‌های سیستم پیشنهادی در قالب یک الگوریتم به‌صورت زیر می‌باشد:

۱- با استفاده از کلاسه‌بند CNN^1 موقعیت سر در تصویر مشخص می‌گردد.

¹ Convolutional Neural Network (CNN)

۲- زاویه قرارگیری شخص در تصویر آزمون محاسبه می‌شود. در این بخش به منظور تخمین زاویه، ترکیبی از نتایج کلاسه‌بند رگرسیون لجستیک بر روی توصیفگر HOG و کلاسه‌بند CNN برای تخمین زاویه بر روی تصویر ناحیه سر، استفاده می‌شود.

۳- در مرحله یکسان‌سازی در صورتی که زاویه قرارگیری شخص در تصویر آزمون، تصویری از روبرو باشد (زاویه بین -90° تا $+90^\circ$)، به کمک الگوریتم خوشه‌بندی در فضای رنگی Lab، ناحیه‌ای از پوشش فرد که در زیر ناحیه سر قرار گرفته، در صورتی که شرایط رنگی متفاوتی از نواحی اطراف داشته باشد، استخراج می‌گردد. خروجی این مرحله به دو شکل می‌تواند باشد. در حالت اول شامل ماسکی با ابعاد تصویر اولیه است که نواحی خاصی از پوشش را مشخص می‌کند. در حالت دوم برای هر سطر از تصویر جدید با نمونه‌افزایی^۱ نواحی مجاور ماسک، تصویر کاملی خواهیم داشت که فاقد ناحیه استخراج شده در مرحله قبل می‌باشد. ماسک مذکور با رنگ قرمز هاشورخورده و تصویر نمونه‌افزایی شده در شکل ۳-۲ به ترتیب با برچسب‌های a و b مشخص شده‌اند. برای تصاویر موجود در مجموعه گالری نیز فرآیند مشابهی انجام می‌گیرد.

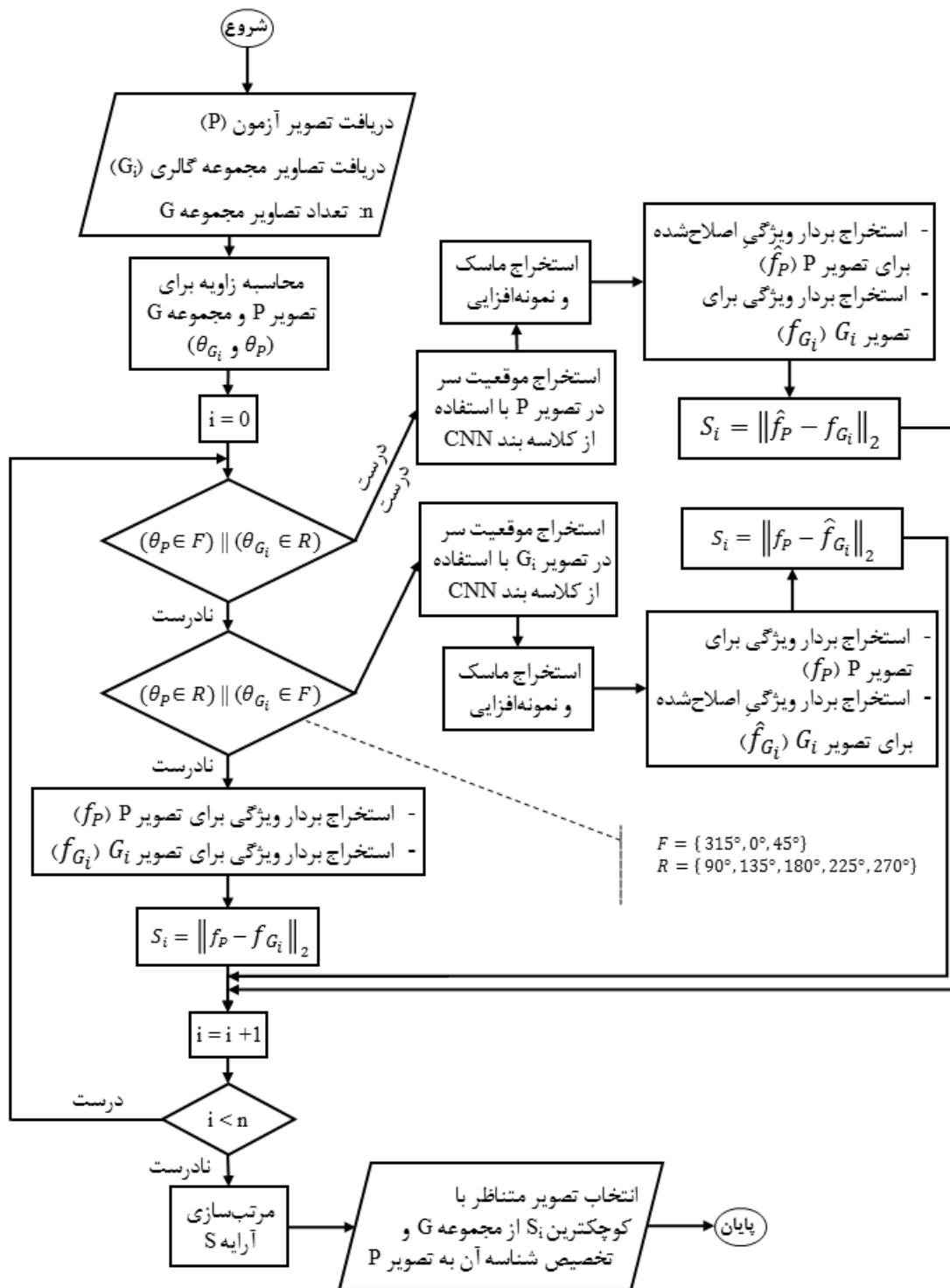
۴- در مرحله تطبیق و محاسبه شباهت، در دو حالت از ماسک یا تصویر نمونه‌افزایی شده استفاده می‌کنیم: حالت اول زمانی که تصویر آزمون (P) از روبرو و تصویر مورد بررسی در مجموعه گالری (G_i)، در زاویه‌ای غیر از روبرو باشد. این حالت بدین معنی است ناحیه استخراج شده از پوشش فرد در تصویر P ، در تصویر G_i قابل‌رؤیت نیست و تغییرات مذکور بر روی P صورت می‌گیرد. حالت دوم زمانی است که تصویر G_i از روبرو و تصویر P در زاویه روبرو نباشد. در این حالت، تصویر G_i تغییر خواهد نمود.

۵- در مرحله استخراج ویژگی در الگوریتم بازشناسی، می‌توان از هر یک از حالات ماسک‌گذاری یا نمونه‌افزایی استفاده نمود. روش نمونه‌افزایی به سادگی و با تغییرات اندک در بسیاری از

^۱ Upsampling

الگوریتم‌ها قابل استفاده می‌باشد. اعمال روش ماسک‌گذاری مستلزم انجام تغییرت بیشتری در گام استخراج ویژگی است. برای مثال در الگوریتم‌هایی که استخراج ویژگی آنها توسط شبکه عصبی عمیق صورت می‌گیرد، استفاده از روش ماسک‌گذاری بسیار دشوار است. درحالی‌که در روش دوم، تصویر نمونه‌افزایی شده به‌سادگی به‌عنوان ورودی شبکه قابل استفاده می‌باشد. بدین ترتیب برای محاسبه میزان شباهت تصویر آزمون و هر یک از تصاویر موجود در مجموعه گالری، بردار ویژگی اصلاح‌شده‌ای بر طبق یکی از دو حالت ذکر شده در بالا خواهیم داشت. در این رساله از دو الگوریتم معرفی شده در [۲۸] و [۲۵] که جزو روش‌های برپایه یادگیری هستند به‌عنوان مرحله نهایی استفاده و نتایج حاصله برای روش‌های ماسک‌گذاری و نمونه‌افزایی بر روی این دو الگوریتم گزارش شده است.

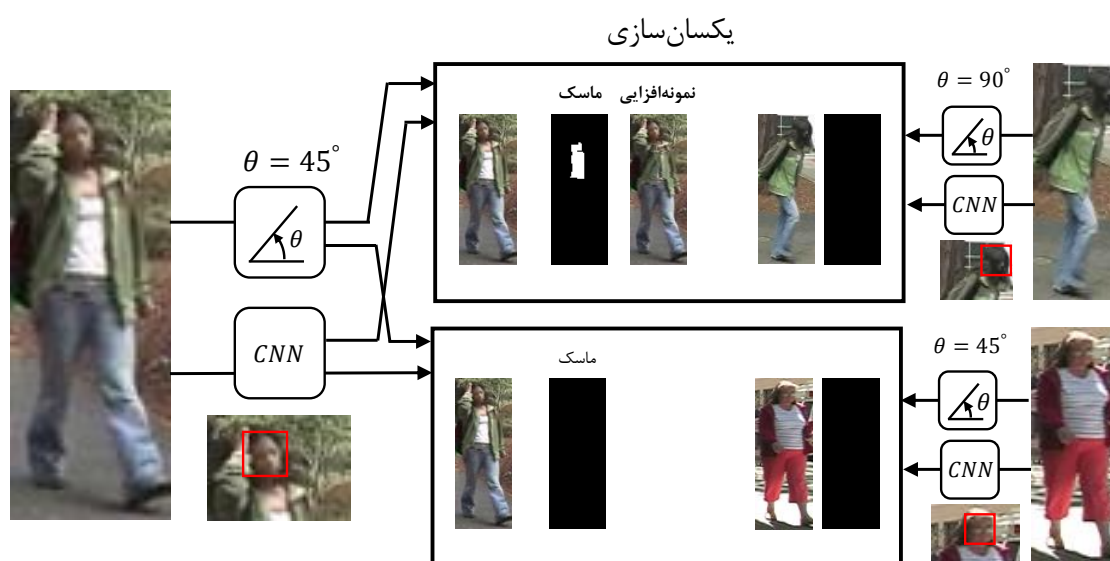
شکل ۳-۳ الگوریتم پیشنهادی را در قالب یک فلوچارت نمایش می‌دهد.



شکل ۳-۳: فلوچارت رویکرد پیشنهادی در این رساله

شکل ۳-۴ مثالی از گام یکسان‌سازی تصویر آزمون و دو تصویر نمونه از مجموعه گالری را نشان می‌دهد. در این مثال تنها زمانی که تصویر آزمون با تصویر اول از گالری مقایسه می‌گردد ماسک موردنظر

استخراج یا نمونه‌افزایی انجام شده و در واقع فرآیند یکسان‌سازی اعمال می‌شود. ماسک و تصویر نمونه‌افزایی شده در این مرحله می‌تواند در گام استخراج ویژگی از هر الگوریتم بازشناسی مورد استفاده قرارگیرد. در آزمایش‌های انجام شده نشان خواهیم داد که استفاده از سیستم پیشنهادی منجر به بهبود در دقت بازشناسی می‌شود.



شکل ۳-۴: مرحله یکسان‌سازی و استخراج ماسک در رویکرد پیشنهادی

در ادامه این فصل هر یک از زیر سیستم‌های رویکرد پیشنهادی به تفصیل مورد بررسی قرار خواهد گرفت. بدین ترتیب، ابتدا به معرفی زیرسیستم تشخیص زاویه می‌پردازیم. در این بخش دو راهکار پیشنهاد شده که هر یک تشریح می‌گردد. سپس زیرسیستم استخراج ناحیه سر در تصویر را خواهیم داشت. در بخش بعد زیرسیستم یکسان‌سازی را با استفاده از اطلاعات بدست آمده از دو بخش قبل شرح خواهیم داد. در بخش نهایی نحوه استفاده از خروجی مرحله یکسان‌سازی و تلفیق آن با یکی از الگوریتم‌های پیشگام در بازشناسی مورد بررسی قرار می‌گیرد.

۳-۲- تشخیص زاویه قرارگیری شخص در تصویر

زاویه قرارگیری شخص در تصویر یکی از ویژگی‌های مهم در تحلیل رفتاری سیستم‌های نظارتی است. این ویژگی شامل اطلاعات کاربردی، نظیر جهت توجه یا جهت حرکت شخص می‌باشد. کیفیت پایین تصاویر، تداخل رنگی و بافتی تصویر شخص با پیش‌زمینه و انسدادهای جزئی، تخمین زاویه را با دشواری همراه می‌سازد. در عین حال از آنجایی که این ویژگی، بخشی از سیستم نظارتی ما است باید با پیچیدگی زمانی پایین و دقت بالا عمل تخمین زاویه انجام گیرد. لذا ارائه روشی که این دو هدف را توأمان برآورده سازد از اهمیت خاصی برخوردار است. یکی از مراحل پیش‌نیاز در تخمین زاویه، استخراج ناحیه شخص در تصویر می‌باشد. دقت در موقعیت‌یابی و تشخیص شخص در تصویر، بر دقت تخمین زاویه نسبت به دوربین تأثیر مستقیمی خواهد گذاشت. روش‌های بسیار زیادی برای استخراج ناحیه شخص در تصویر ارائه شده‌اند که از ویژگی‌هایی نظیر HOG، LBP، توصیفگر کوواریانس و ویژگی‌های شبه‌هار^۱ [۵۴-۵۶] به همراه مدل‌های یادگیری نظیر ماشین بردار پشتیبان و مدل‌های آبخاری استفاده کرده‌اند [۲، ۳].

در زمینه تخمین زاویه، انواع گسترده‌ای از ویژگی‌ها و کلاس‌بندها قابل استفاده است. در اغلب این روش‌ها مسئله تخمین زاویه تبدیل به مسئله کلاس‌بندی زاویه شده است. بدین ترتیب که محدوده تخمین که عددی حقیقی در بازه ۰ تا ۳۶۰ درجه است به ۸ کلاس گسسته شامل $(0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ)$ تقسیم می‌گردد. برای مثال در [۵۷] توصیفگر HOG به‌عنوان بردار ویژگی و ماشین بردار پشتیبان به‌عنوان کلاس‌بند استفاده شده است. در [۵۸] ضرایب تبدیل موجک به‌عنوان ویژگی استفاده شده است. در اغلب روش‌ها مرحله تشخیص ناحیه شخص در تصویر و

^۱ Haar-like features

تخمین زاویه در دو مرحله مجزا انجام می‌شود اما در برخی از روش‌ها این دو مرحله به‌صورت توأمان تولید می‌گردند [۵۹، ۶۰].

در [۶۱] از این توصیفگر کوواریانس به‌منظور تخمین زاویه سر و کل بدن استفاده شده است. بدین منظور کل ناحیه بدن شخص به نواحی کوچک‌تری تقسیم شده و برای هر ناحیه توصیفگر کوواریانس برپایه مجموعه‌ای از ویژگی‌ها محاسبه می‌شود. این توصیفگر که به‌صورت ماتریس است در فضای ریمانی واقع شده و بدین ترتیب معیارهایی نظیر فاصله اقلیدسی برای مقایسه دو نمونه از این توصیفگر قابل استفاده نیست. در [۶۱] معیار فاصله‌ای در فضای ماتریسی ارائه شده است. سپس برای هر یک از زیر ناحیه‌ها یک کلاسه‌بند SVM آموزش می‌یابد. به‌منظور استخراج ویژگی از هر زیر ناحیه، ابتدا تعدادی از تصاویر از مجموعه آموزشی به‌عنوان نقاط مرجع انتخاب می‌گردند. سپس برای هر زیرناحیه، فاصله آن زیرناحیه با زیرناحیه متناظر در تصاویر مرجع محاسبه می‌شود. این سنجش فاصله براساس معیار پیشنهادی در [۶۱] است. بنابراین بردار ویژگی برای هر زیرناحیه، از کنار هم قرار دادن این فواصل تشکیل می‌شود. تخمین نهایی، ترکیبی از تمامی ماشین‌های بردار پشتیبان آموزش یافته برای کلیه زیرناحیه‌ها است.

در [۶۲] چندین کلاسه‌بند برای چند حالت مرسوم از نحوه قرارگیری انسان، آموزش داده شده است. در مرحله تست، ابتدا کلاس کلی شخص در تصویر به کمک کلاسه‌بندهای ذکر شده مشخص می‌گردد. سپس با انطباق کانتوری در هر کلاس، حالت قرارگیری شخص و از جمله زاویه قرارگیری آن، مشخص می‌شود. در برخی از روش‌ها به‌منظور بهبود نتایج، عملیات پس‌پردازشی بر روی خروجی کلاسه‌بندها انجام می‌گیرد. در چنین مواقعی خروجی کلاسه‌بندها عددی حقیقی است که بیان‌کننده درجه اطمینان است [۵۹، ۶۳]. در این روش‌ها از ترکیبی از توابع توزیع گاوسی برای عملیات پس‌پردازشی استفاده می‌کنند. در [۶۳] مجموعه‌ای از درخت‌های تصادفی به‌عنوان کلاسه‌بند و توصیفگر HOG به‌عنوان ویژگی، مورد استفاده قرار گرفته و خروجی نهایی کلاسه‌بند با استفاده از روش ذکر شده

در بالا پس‌پردازش و بهبود می‌یابد. در ادامه دو رویکرد جدید در تخمین زاویه پیشنهاد می‌شود که در این رساله مورد استفاده قرار گرفته است.

۳-۲-۱- تشخیص زاویه با استفاده از اطلاعات کانتوری

در این رویکرد از اطلاعات کانتوری^۱ برای تشخیص زاویه شخص در تصویر استفاده می‌شود. منظور از کانتور، منحنی بسته‌ای در اطراف شیء است که بر روی لبه‌های بیرونی شیء قرار می‌گیرد. در واقع استفاده از کانتور، دید یکپارچه‌تری از تصویر شیء هدف و ساختار آن، نسبت به توصیفگرهای محلی (مانند HOG) فراهم می‌سازد. بدین ترتیب مسئله تشخیص زاویه به مسئله یافتن کانتوری با بیشترین انطباق با تصویر ورودی تبدیل خواهد شد. در این رویکرد دو نوع کانتور سراسری و ناحیه‌ای مورد استفاده قرار گرفته است. منظور از کانتور سراسری کانتوری از کل اندام شخص می‌باشد. منظور از کانتور ناحیه‌ای کانتوری مرتبط با بالاتنه یا پایین‌تنه است.

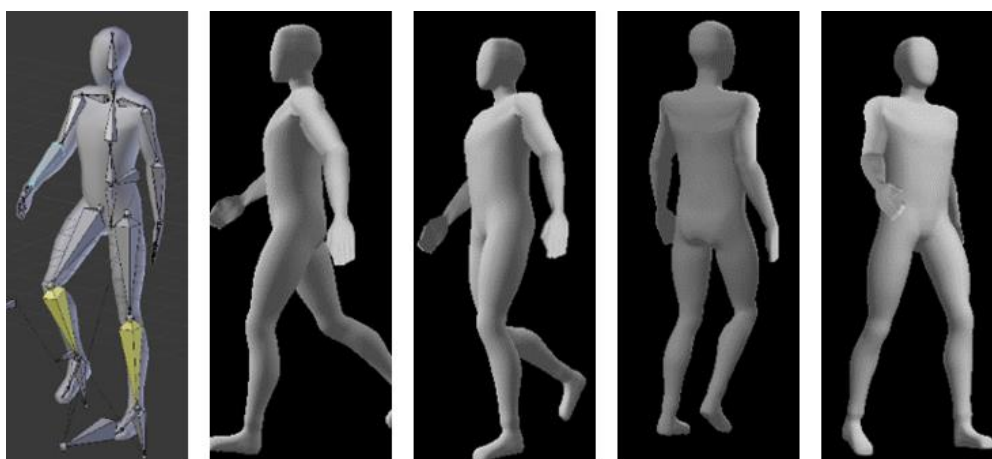
کانتورهای ناحیه‌ای، این امکان را فراهم می‌سازند که تطبیق دقیق‌تری به تصویر ورودی حاصل گردد. برای مثال ممکن است که به دلیل وجود نویز یا وجود پس‌زمینه‌ی شلوغ، ناحیه پایین‌تنه از وضوح بیشتر و لبه‌های مشخص‌تری برخوردار باشد. در نتیجه کانتور مربوط به پایین‌تنه انطباق بهتری نسبت به کانتور سراسری ایجاد نماید. لذا در روش پیشنهادی ترکیبی از کانتورهای بخش‌های مختلف، مورد استفاده قرار گرفته است. تهیه کانتورهای مجزا با استفاده از روش‌های موجود بسیار دشوار است. تولید داده‌های کانتور برای کل بدن، به دو روش ممکن می‌باشد. در روش اول مجموعه‌ای از تصاویر انسان که از زوایای مختلف تهیه شده است به صورت دستی علامت‌گذاری می‌شود و محل کانتور دور فرد استخراج می‌گردد. در روش دوم می‌توان با استفاده از تصاویر ویدئویی و تکنیک‌های جداسازی شیء از پس‌زمینه، کانتور محاط شده از فرد را استخراج نمود که نسبت به روش اول به عملیات دستی کمتری نیاز دارد. در این روش دقت کانتورهای به‌دست آمده به دقت الگوریتم‌های جداسازی پیش‌زمینه از پس‌زمینه

^۱ Contour

وابسته است. علاوه بر مشکلاتی که دو روش فوق در تولید کانتورهای کلی دارند، در تولید کانتورهای ناحیه‌ای نیز ناتوان، یا با دشواری بسیاری همراه می‌باشند.

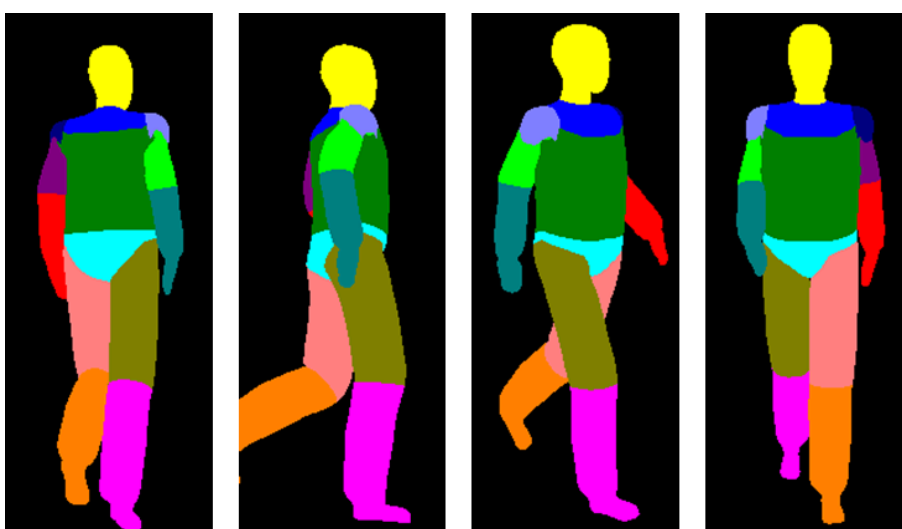
بدین منظور در روش پیشنهادی از یک مدل سه‌بعدی انسان استفاده می‌شود. استفاده از مدل سه‌بعدی امکان ایجاد حالت‌های گوناگون از انسان را ممکن می‌سازد. همچنین از آنجایی که موقعیت اندام‌ها در این مدل مشخص می‌باشد می‌توان به کمک رنگ‌آمیزی، نواحی مختلف بدن را استخراج نمود. علاوه بر آن، استفاده از مدل سه‌بعدی، امکان ایجاد حالات مختلف بدن، انواع نورپردازی، انواع رزولوشن و بسیاری موارد دیگر را فراهم می‌سازد.

مدل سه‌بعدی پیشنهادی در نرم‌افزار طراحی سه‌بعدی Blender ایجاد شده است. در این مدل، استخوان‌بندی مشابه با استخوان‌بندی انسان ایجاد شده است. این استخوان‌بندی امکان جابجایی هر یک از مفاصل را در اختیار ما قرار می‌دهد. مدل سه‌بعدی طراحی شده با استفاده از موتور سه‌بعدی Ogre3D در یک صحنه فرضی با نورپردازی خاصی بارگذاری گردیده است. از این موتور گرافیکی در محیط برنامه‌سازی C++ استفاده شده است. این موتور گرافیکی امکان ایجاد تغییرات در کلیه مفاصل را فراهم می‌سازد. سپس انواع حالات ممکن که یک انسان در هنگام راه رفتن به خود می‌گیرد را روی این مدل مصنوعی اعمال نمودیم که در مجموع ۳۰ حالت در نظر گرفته شده است [۶۴]. در مرحله بعد برای هر حالت، ۷۲ زاویه قرارگیری نسبت به دوربین و سه زاویه نسبت به افق در نظر گرفته می‌شود که در مجموع $۳۰ \times ۷۲ \times ۳ = ۶۴۸۰$ تصویر به‌عنوان مجموعه داده به‌دست آمده است. شکل ۳-۵ تعدادی از این حالات به همراه مدل استخوان‌بندی را نشان می‌دهد.



شکل ۳-۵: مدل سه بعدی طراحی شده از انسان در زوایا و حالت های مختلف راه رفتن

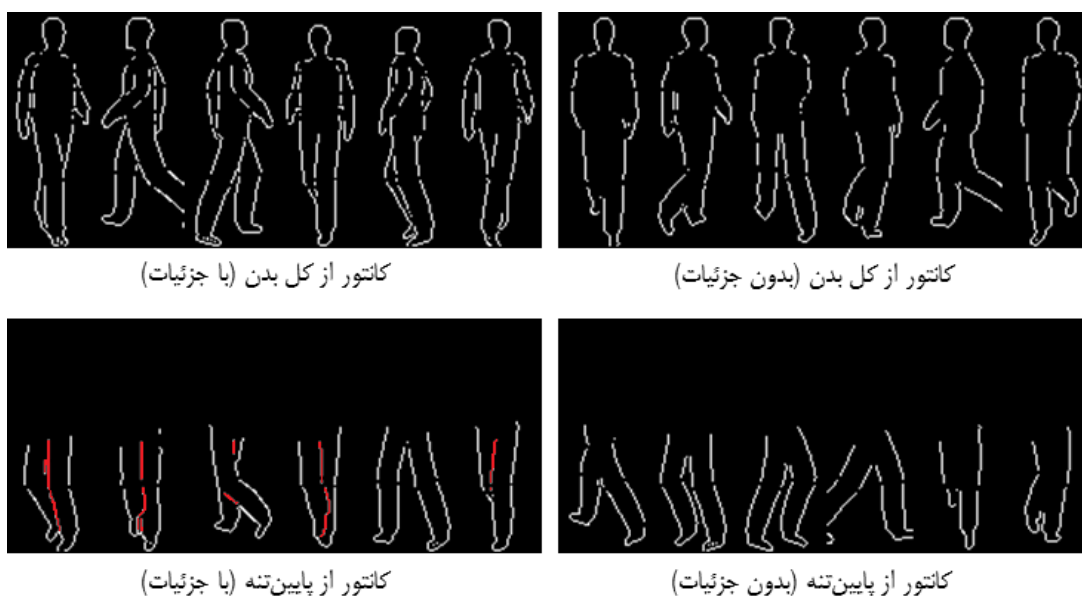
در گام بعد به منظور استخراج کانتور از تصاویر به دست آمده، مدل سه بعدی رنگ آمیزی می شود. رنگ آمیزی صورت گرفته، استخراج کانتورهای مربوط به بخش های مختلف بدن را ممکن می سازد. شکل ۳-۶ نمونه ای از این رنگ آمیزی را نشان می دهد.



شکل ۳-۶: رنگ آمیزی مدل سه بعدی به منظور تشخیص اندام

در انتخاب این رنگ ها سعی می شود که مقادیر سه مؤلفه رنگی متفاوت باشند تا در مرحله بعد با پردازشی بسیار ساده بتوان کانتورهای مورد نظر را استخراج نمود. در گام بعد با عملیات لبه یابی در تصاویر رنگی کانتورهای متنوعی استخراج می شود. برای مثال مجموعه ای از کانتورهای پایین تنه، بالاتنه، کانتور کلی و کانتورهایی با جزئیات تهیه شده است.

استفاده از تکنیک پیشنهادی در این پژوهش، این امکان را در اختیار ما قرار می‌دهد که کانتورهایی با جزئیات را نیز استخراج نماییم. تولید این چنین کانتورهایی با استفاده از روش جداسازی پس‌زمینه امکان‌پذیر نمی‌باشد. کانتورهای با جزئیات در برخی موارد انطباق‌های بسیار بهتری را در تصویر ایجاد می‌کنند که در بخش‌های بعد به آن اشاره خواهیم نمود. شکل ۳-۷ چند نمونه از کانتورهای به‌دست‌آمده را نشان می‌دهد.



شکل ۳-۷: نمونه‌ای از کانتورهای استخراج‌شده از مدل سه‌بعدی برای بخش‌های مختلف بدن

همان‌طور که در شکل ۳-۷ مشاهده می‌کنید در حالت با جزئیات، خطوطی از کانتور ظاهر می‌شود (با رنگ قرمز مشخص شده است) که می‌تواند در تشخیص جهت فرد مفید باشد. همچنین در کاربردهای دیگری نظیر استخراج اعضای بدن، می‌تواند مورد استفاده قرار گیرد.

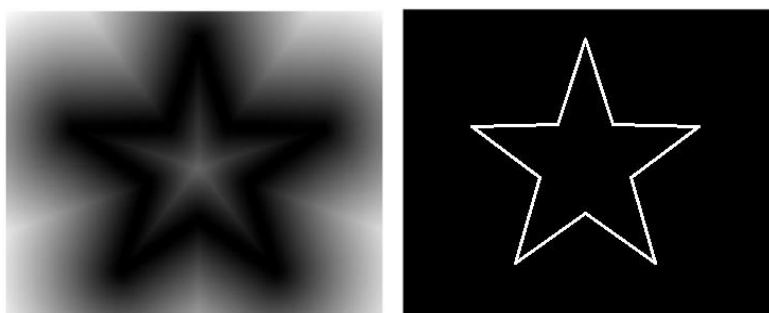
به‌منظور انطباق بهترین کانتور به تصویر ورودی، ابتدا لبه‌های تصویر استخراج می‌شود. در مرحله بعد فاصله بین تصویر باینری لبه و کانتورها محاسبه می‌شود. طریقه محاسبه این فاصله، با استفاده از تبدیل فاصله^۱ برای دو تصویر باینری است که در ادامه شرح داده شده است. یافتن شبیه‌ترین و

^۱ Distance Transform

نزدیک‌ترین کانتور به تصویر لبه می‌تواند به صورت یک جستجوی خطی صورت گیرد. در این حالت پیچیدگی محاسباتی، بسیار بالا می‌باشد. همچنین اگر در تصویر ورودی جابجایی رخ داده باشد ممکن است کانتور مناسبی استخراج نگردد. در روش پیشنهادی ابتدا عمل خوشه‌بندی بر روی کانتورها انجام می‌شود. انجام خوشه‌بندی تعداد مقایسات را به مقدار قابل توجهی کاهش می‌دهد. همچنین عمل پیمایش درخت به نحوی انجام می‌شود که مشکل جابجایی تصویر ورودی تا حدی رفع می‌گردد. در زیربخش ۲-۳ به تشریح نحوه این خوشه‌بندی پرداخته می‌شود.

۳-۲-۱-۱- تبدیل فاصله

تبدیل فاصله یکی از ابزارهای مهم و پرکاربرد در بینایی ماشین، پردازش تصویر و شناسایی الگو می‌باشد. تبدیل فاصله برای یک تصویر باینری، ماتریسی هم‌بعد با تصویر باینری ورودی است. این تصویر باینری معمولاً یک تصویر لبه (برای مثال خروجی یک الگوریتم لبه‌یاب) می‌باشد. درایه‌های این ماتریس بیانگر فاصله هر نقطه از تصویر تا نزدیک‌ترین نقطه روشن در تصویر باینری می‌باشد [۶۵]. شکل ۳-۸ یک نمونه تصویر لبه و تبدیل فاصله متناظر را نشان می‌دهد.



شکل ۳-۸: راست: تصویر لبه، چپ: تبدیل فاصله متناظر با تصویر سمت راست

در شکل ۳-۸ تصویر سمت چپ یک ماتریس تبدیل فاصله را نشان می‌دهد. در این تصویر مقادیر کاملاً تیره بیانگر نقاطی در تصویر اصلی است که روی لبه قرار گرفته‌اند و در واقع فاصله این نقاط تا پیکسل‌های لبه صفر است؛ و بالعکس نقاط روشن بیانگر مکان‌هایی هستند که از لبه‌ها فاصله دارند. هرچه از لبه‌ها فاصله می‌گیریم تبدیل فاصله مقدار عددی بزرگ‌تری به خود می‌گیرد، همان‌گونه که در تصویر چپ این

شکل مقدار روشن تری به آن‌ها اختصاص یافته است. یکی از کاربردهای تبدیل فاصله تعیین شباهت بین دو مجموعه پیکسل می‌باشد. این دو مجموعه پیکسل می‌تواند دو تصویر باینری باشد که پیکسل‌های غیر صفر اعضای مجموعه هستند. فرض کنید برای دو تک‌نقطه a و b ، معیار فاصله $d(\cdot)$ ، به صورت فاصله اقلیدسی باشد. بدین ترتیب به صورت رابطه ۱-۳ داریم.

$$d(a, b) = \|a - b\| \quad (1-3)$$

همچنین فاصله بین تک‌نقطه a و مجموعه نقاط $\mathcal{B} = \{b_1, \dots, b_{N_b}\}$ به صورت رابطه ۲-۳ تعریف شود.

$$d(a, \mathcal{B}) = \min_{b \in \mathcal{B}} \|a - b\| \quad (2-3)$$

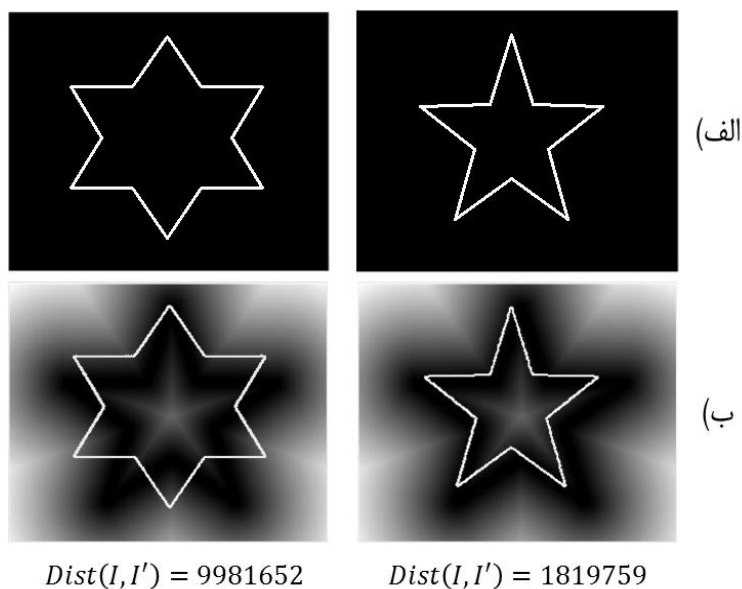
برای محاسبه فاصله بین دو مجموعه از نقاط، روابط متفاوتی قابل تعریف می‌باشد. در روش پیشنهادی ما از رابطه ۳-۳ برای محاسبه این فاصله استفاده نموده‌ایم [۶۶]. یکی از ویژگی‌های این رابطه، مقاوم بودن به مجموعه‌های حاوی نویز می‌باشد. بدین ترتیب فاصله دو مجموعه نقاط \mathcal{A} و \mathcal{B} $\{a_1, \dots, a_{N_a}\}$ به صورت رابطه زیر محاسبه می‌شود:

$$d(\mathcal{A}, \mathcal{B}) = \frac{1}{N_a} \sum_{a \in \mathcal{A}} d(a, \mathcal{B}) \quad (3-3)$$

حال فرض کنید دو تصویر لبه داریم و قصد داریم فاصله دو تصویر را محاسبه کنیم. هر یک از دو تصویر لبه را می‌توان مجموعه‌ای از نقاط در نظر گرفت. با این فرض می‌توانیم از رابطه ۳-۳ برای محاسبه فاصله استفاده کنیم. تبدیل فاصله، ماتریسی از مجموعه \mathcal{B} فراهم می‌کند که ما را از محاسبه‌های چندباره رابطه ۲-۳ بی‌نیاز می‌سازد. در واقع برای هر نقطه از تصویر، کمینه فاصله تا پیکسل لبه محاسبه شده است. با این فرض‌ها فاصله دو تصویر باینری I و I' به صورت رابطه ۴-۳ خواهد بود.

$$Dist(I, I') = Dt(I) \circ I' \quad (4-3)$$

در این رابطه، $Dt()$ تبدیل فاصله می‌باشد. عملگر \circ مجموع ضرب درایه‌های متناظر دو ماتریس می‌باشد. به‌عنوان مثال فرض کنید می‌خواهیم تعیین کنیم که از بین دو چندضلعی سطر الف در شکل ۳-۹ کدامیک به چندضلعی در شکل ۳-۸ شبیه‌تر است.



شکل ۳-۹: محاسبه شباهت تصویر دو چندضلعی و شکل ۳-۸ براساس تبدیل فاصله. الف) تصاویر دو چندضلعی، ب) انطباق تصاویر متناظر با سطر الف بر روی تبدیل فاصله‌ی شکل ۳-۸

در شکل ۳-۹ فاصله دو تصویر چندضلعی با تصویر شکل ۳-۸ براساس رابطه ۳-۴ محاسبه شده است. بر این اساس تصویری که مقدار فاصله کمتری دارد به‌عنوان تصویر شبیه‌تر انتخاب می‌گردد که در مثال بالا تصویر سمت راست خواهد بود. در واقع تصویر سمت راست، همان تصویر شکل ۳-۸ با مقداری جابجایی می‌باشد. کلیه تصاویر در ابعاد یکسان می‌باشند.

۳-۲-۱-۲- خوشه‌بندی کانتورها

همان‌طور که در بخش‌های قبل اشاره شد فاصله دو تصویر باینری برپایه تبدیل فاصله، معیاری غیر اقلیدسی می‌باشد. با توجه به این موضوع خوشه‌بندی سلسله مراتبی^۱ مناسب می‌باشد. در خوشه‌بندی سلسله مراتبی ابتدا فاصله هر زوج از نمونه داده‌ها محاسبه می‌شود و حاصل آن در ماتریسی

^۱ Hierarchical Clustering

به نام ماتریس فاصله ذخیره می‌گردد. به صورت مشابه، در روش پیشنهادی ماتریس فاصله‌ای به ابعاد 6480×6480 برای کلیه کانتورها محاسبه می‌شود. مقدار عددی درایه‌های این ماتریس، بیانگر فاصله هر زوج کانتور (به روش تبدیل فاصله) می‌باشد. در پژوهش صورت گرفته از پیوند میانگین^۱ برای خوشه‌بندی استفاده شده است [۶۷]. نتیجه خوشه‌بندی سلسله مراتبی، درختی است که گره‌های میانی نمایانگر خوشه‌ها هستند. بدین ترتیب براساس تعداد خوشه‌های مورد نیاز می‌توان از ریشه تا سطح دلخواه از درخت را پیمود. گره ریشه در این درخت شامل خوشه‌ای است که دربردارنده کلیه داده‌ها می‌باشد. برای مثال ممکن است در یک گره میانی، تعداد ۱۰۰۰ کانتور درون یک خوشه قرار گیرند. در پژوهش صورت گرفته به گره‌های میانی درخت خوشه‌بندی، کلمه ابرکانتور^۲ اطلاق می‌شود. بدین معنی که این گره مجموعه‌ای از کانتورها را درون خود دارد.

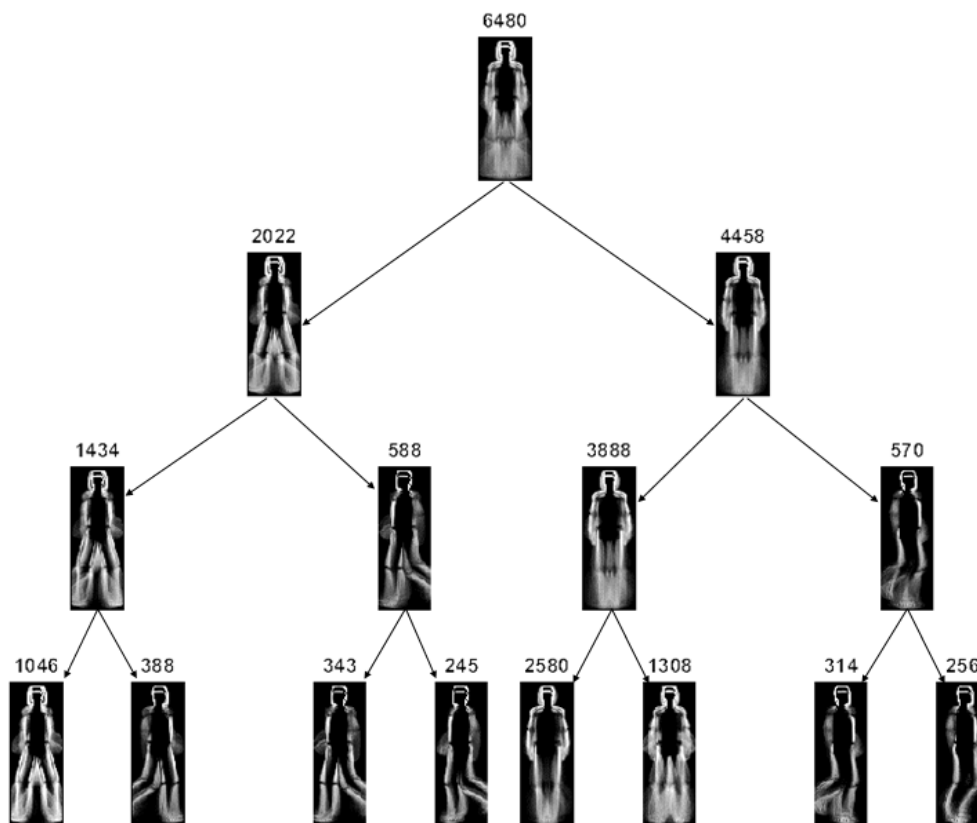
نحوه ساخت یک ابرکانتور بدین صورت است که فرض کنید لبه‌ها در کانتور، مقدار ۱ و در نواحی زمینه، مقدار ۰ را داشته باشد. یک ابرکانتور از مجموع پیکسل‌های نظیر به نظیر یک کانتور با کانتور دیگر حاصل می‌شود. درنهایت، مجموع به‌دست‌آمده بر تعداد تصاویر تقسیم می‌گردد. رابطه ۳-۵ نحوه محاسبه یک ابرکانتور با استفاده از چندین کانتور دیگر را نشان می‌دهد:

$$C = \frac{1}{N} \sum_{i=1}^N C_i \quad (5-3)$$

در این رابطه C_i نمایانگر کانتورها و C نمایانگر ابرکانتور می‌باشد. در شکل ۳-۱۰ قسمتی از درخت حاصل از خوشه‌بندی یکی از انواع کانتورهای به‌دست‌آمده و ابرکانتورهای متناظر با هر گره نمایش داده شده است. در این شکل اعداد بالای هر گره نمایانگر تعداد کانتورهای تشکیل‌دهنده آن ابرکانتور می‌باشد.

^۱ Average linkage

^۲ Hyper contour



شکل ۳-۱۰: خوشه‌بندی سلسله مراتبی کانتورها و تشکیل ابرکانتورها

۳-۲-۱-۳- پیمایش خوشه‌بندی

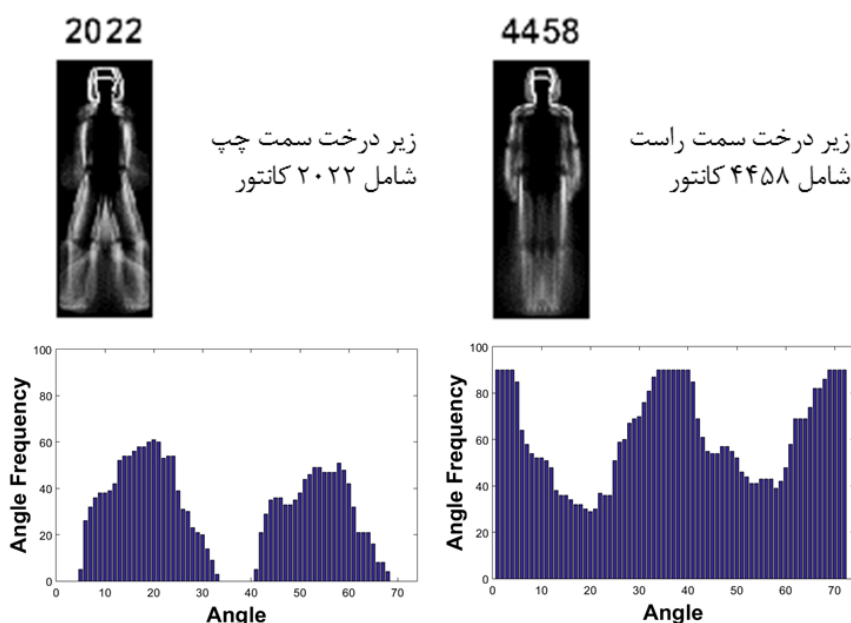
از آنجایی که کانتورهای ساخته شده دارای ابعاد ۱۲۸×۴۸ پیکسل می‌باشند، تصویر ورودی نیز به این ابعاد تغییر داده می‌شود. سپس با استفاده از الگوریتم‌های لبه‌یابی، لبه‌های تصویر استخراج می‌گردد. در این پژوهش با استفاده از الگوریتم لبه‌یابی canny عمل لبه‌یابی انجام می‌شود. دلیل استفاده از این الگوریتم مقاوم بودن آن نسبت به تصاویر نویزی می‌باشد. سپس تبدیل فاصله برای تصویر لبه محاسبه می‌شود. در هر مرحله از پیمایش، فاصله بین ابرکانتور و تصویر تبدیل‌یافته بر اساس رابطه ۳-۴ محاسبه می‌گردد. گره فرزندی که فاصله کمتری دارد برای پیمایش بعدی انتخاب می‌شود. یکی از مشکلاتی که می‌تواند منجر به انطباق نادرست گردد جابجایی در تصویر ورودی است. برای رفع این مشکل مکانیزمی طراحی گردیده که با پایین رفتن در درخت این موقعیت به‌مرور اصلاح می‌شود و در نتیجه کانتوری که

کمترین فاصله را با تصویر لبه دارد استخراج می‌شود. گام‌های الگوریتم پیمایش درخت به صورت زیر می‌باشد:

۱. محاسبه تبدیل فاصله برای لبه‌های تصویر ورودی.
۲. تخصیص بردار جابجایی به صورت $[x=0, y=0]$.
۳. انتخاب ریشه درخت خوشه‌بندی به عنوان گره ریشه.
۴. محاسبه چهار مقدار فاصله برای هر یک از دو گره فرزند (ابركانتور) پیش رو، که عبارت‌اند از:
 - a. ابرکانتور جابجا شده به اندازه +1 پیکسل در راستای x.
 - b. ابرکانتور جابجا شده به اندازه -1 پیکسل در راستای x.
 - c. ابرکانتور جابجا شده به اندازه +1 پیکسل در راستای y.
 - d. ابرکانتور جابجا شده به اندازه -1 پیکسل در راستای y.
۵. انتخاب کمترین فاصله در مرحله ۴ و تخصیص آن به عنوان فاصله فرزند موردنظر.
۶. انتخاب فرزند با فاصله کمتر و تخصیص آن به عنوان مقدار جدید ریشه.
۷. به‌روزرسانی بردار جابجایی کل براساس مقدار جابجایی بدست آمده در گام ۴.
۸. رفتن به قدم ۱۰ در صورتی که تعداد کانتورهای موجود در گره منتخب، کمتر از یک مقدار آستانه باشد.
۹. رفتن به قدم ۴.
۱۰. اتمام.

ساختمان داده مورد استفاده برای ذخیره‌سازی این ابرکانتورها درخت می‌باشد. یکی از اطلاعات مفیدی که در ادامه مورد استفاده قرار می‌گیرد اطلاعات زاویه‌ای در هر گره از درخت می‌باشد. همان‌طور که قبلاً گفته شد ۶۴۸۰ کانتوری که در گره ریشه درخت قرار گرفته شامل ۷۲ زاویه مختلف می‌باشند. بدین ترتیب اگر فراوانی زاویه را برای گره ریشه محاسبه کنیم به صورت یک توزیع یکنواخت بوده که

برای هر زاویه ۹۰ کانتور را شامل می‌شود ($۹۰ \times ۷۲ = ۶۴۸۰$). به تدریج که درخت پیمایش می‌شود، این توزیع از حالت یکنواخت خارج گردیده و به سمت تمرکز بر روی یک زاویه خاص پیش می‌رود. بدین ترتیب برای هر گره از درخت، هیستوگرامی ذخیره می‌شود که شامل فراوانی در ۷۲ زاویه می‌باشد. برای مثال برای دو زیرگره از گره ریشه در شکل ۳-۱۰ هیستوگرام زاویه‌ای به صورت شکل ۳-۱۱ خواهد بود.



شکل ۳-۱۱: هیستوگرام زاویه‌ای برای دو گره از درخت شکل ۳-۱۰

همان‌طور که در شکل ۳-۱۱ مشاهده می‌شود مجموع دو هیستوگرام برای گره‌های فرزند ریشه، برابر فراوانی گره والد خواهد بود که یک توزیع یکنواخت می‌باشد. نکته‌ای که باید بدان اشاره کرد نقطه اتمام پیمایش درخت می‌باشد. به منظور جلوگیری از بیش‌برازش^۱ شدن پیمایش، زمانی که گره‌ای با تعداد کانتور ۲۰۰ ملاقات می‌شود عمل پیمایش خاتمه یافته و هیستوگرام زاویه‌ای مربوط به آن گره به عنوان تخمینی از زاویه قرارگیری فرد نسبت به دوربین در نظر گرفته می‌شود. نحوه انتخاب مقدار حد آستانه در زیر بخش نهایی تشریح شده است.

^۱ Overfitting

۳-۲-۱-۴- محاسبه زاویه قرارگیری شخص به کمک ترکیبی از درخت‌های کانتوری

همان‌طور که در بخش‌های قبل اشاره شد با کمک جداسازی قسمت‌های مختلف بدن، کانتورهایی برای هر بخش از بدن تهیه شده است. در مرحله بعد و با کمک خوشه‌بندی برای هر گروه از کانتورها درخت مجزایی ساخته می‌شود. در مرحله تعیین زاویه قرارگیری شخص در تصویر، تصویر ورودی به چندین نوع درخت داده می‌شود. خروجی هر یک از این درخت‌ها، هیستوگرام زاویه‌ای می‌باشد. به عبارت دیگر خروجی پیمایش هر درخت تخمینی از زاویه قرارگیری شخص در تصویر را فراهم می‌کند. برای مثال برای یک تصویر ورودی فرضی، در صورتی که پیمایش درخت به ابرکانتوری مشابه ابرکانتور سمت چپ در تصویر شکل ۳-۱۱ ختم شود، با توجه به هیستوگرام متناظر با آن، زاویه قرارگیری شخص در تصویر با احتمال بیشتر در زوایای 90° درجه یا 270° درجه می‌باشد. در مرحله بعد هیستوگرام‌های به‌دست‌آمده از طریق رابطه ۳-۶ با هم ترکیب شده و در نهایت یک هیستوگرام زاویه‌ای واحد خواهیم داشت که قسمتی از بردار ویژگی نهایی را تشکیل می‌دهد.

$$H(i) = \sum_{k \in \text{treeTypes}} \left(H^k(i) \times |C^k|_{\oplus} \times \left(1 - \frac{D^k}{\sum_{j \in k} D^j} \right) \right) \quad (6-3)$$

رابطه ۳-۶ ترکیب خطی هیستوگرام‌های زاویه‌ای به‌دست‌آمده از پیمایش درخت‌ها را نشان می‌دهد. عبارت $H(i)$ مؤلفه i ام از هیستوگرام ترکیبی نهایی است. عبارت $h^k(i)$ مؤلفه i ام از هیستوگرام خروجی مربوط به درخت k ام می‌باشد. عملگر $|\cdot|_{\oplus}$ مجموع پیکسل‌های یک ابرکانتور را محاسبه می‌کند. بدین ترتیب هیستوگرام زاویه‌ای که متعلق به ابرکانتوری با لبه‌ها و مساحت بیشتر باشد، وزن بیشتری را در هیستوگرام نهایی خواهد داشت. همان‌طور که در بخش ۳-۳ ذکر شد، تصویر ورودی با استفاده از تبدیل فاصله در درخت خوشه‌بندی (برای مثال نوع k) که شامل ابرکانتورها است پیمایش شده و در نهایت به یک گره برگ ختم می‌گردد. در رابطه فوق فاصله تصویر ورودی و گره برگ یافت شده با نماد D^k نشان داده شده است. در نتیجه عبارت انتهای رابطه ۳-۶ نشان می‌دهد، ابرکانتوری که فاصله کمتری تا تصویر ورودی دارد وزن بیشتری در هیستوگرام نهایی خواهد داشت. در نهایت نرمال‌سازی بر روی هیستوگرام

نهایی انجام می‌شود. در این نرمال‌سازی کلیه مقادیر مؤلفه‌های هیستوگرام بر مقدار ماکزیمم تقسیم می‌گردد که به دلیل پرهیز از پیچیدگی در رابطه ۳-۶ لحاظ نشده است. علت این نرمال‌سازی در بخش بعد مورد بررسی قرار خواهد گرفت.

۳-۲-۱-۵- ترکیب هیستوگرام زاویه‌ای و HOG به‌عنوان ویژگی نهایی

در این پژوهش ترکیبی از هیستوگرام زاویه‌ای به‌دست‌آمده در بخش قبل، به همراه ویژگی ارائه‌شده در مقاله [۵۳] استفاده شده است. ویژگی اصلی مورد استفاده در مقاله مذکور، هیستوگرام گرادیان‌های جهتی است. به‌منظور استخراج ویژگی، هیستوگرام گرادیان‌های جهتی برای سه اندازه از تصویر (به ترتیب با مقیاس ۱، ۰٫۵ و ۰٫۲۵) محاسبه و ضرایب HOG این سه مقیاس، در کنار هم قرار داده شده که برداری با ۲۲۶۸ مؤلفه می‌باشد.

در گام بعد اطلاعات کانتوری پیشنهادی در بخش قبل به بردار ویژگی ارائه شده در مقاله [۵۳] اضافه می‌گردند. همان‌طور که در بخش قبل بررسی شد، پیمایش چهار نوع درخت کانتور، و ترکیب هیستوگرام‌های زاویه‌ای حاصله با استفاده از رابطه ۳-۶، برداری هشت مؤلفه‌ای تولید می‌کند. سپس این هشت مؤلفه در کنار ۲۲۶۸ مؤلفه ذکر شده در بالا قرار می‌گیرد. در نتیجه برای هر تصویر، بردار ویژگی ۲۲۷۶ مؤلفه‌ای خواهیم داشت. همچنین در تولید بردار هشت مؤلفه‌ای، نرمال‌سازی صورت می‌گیرد. این نرمال‌سازی بزرگی بردار تولیدشده را در مقایسه با بردار پایه (۲۲۶۸ مؤلفه‌ای) بیشتر نموده و منجر به تأثیرگذاری در نتایج و درنهایت بهبود می‌گردد. به‌عنوان مثال در بردار نرمال شده همیشه مؤلفه‌ای با مقدار ۱ داریم. به عبارتی این نرمال‌سازی بخشی از نحوه ترکیب دو بردار ویژگی است. البته به‌منظور ترکیب دو روش می‌توان از روش‌های دیگر ترکیب نیز استفاده نمود. در این مقاله سعی شده از ساده‌ترین طریقه ترکیب استفاده شود و تمرکز اصلی بر روی رویکرد تولید ویژگی پیشنهادی باشد. با استفاده از این بردار، ماشین بردار پشتیبان آموزش داده می‌شود. به‌منظور کاهش بار محاسباتی، با

استفاده از آنالیز مؤلفه‌های اصلی^۱، بردار ۲۲۷۶ بعدی به ۱۰۰ بعد کاهش داده می‌شود. آزمایش‌های صورت گرفته کلیه حالات ذکرشده را در بردارد.

۳-۲-۱-۶- داده‌های مورد استفاده و آموزش

مجموعه داده‌های مورد استفاده جهت آموزش، داده‌های TUD است که شامل ۴۷۳۰ تصویر آموزشی و ۴۹۶ تصویر تست می‌باشد [۶۸]. تمامی تصاویر با اطلاعات مفیدی نظیر: محدوده شخص، زاویه قرارگیری و نه نقطه مناظر با بخش‌های مختلف بدن، حاشیه‌نویسی^۲ شده است. این مجموعه داده شامل تصاویری از افراد در جهت‌های مختلف می‌باشد. تفاوت دو زاویه مجاور ۴۵ درجه می‌باشد که بدین ترتیب هشت کلاس مختلف خواهیم داشت. پراکندگی داده‌ها در این هشت کلاس تقریباً یکسان بوده و در حدود ۶۰۰ تصویر برای هر کلاس می‌باشد. اندازه تصاویر در برخی موارد باهم متفاوت بوده و همچنین برخی از تصاویر رنگی هستند. لذا قبل از آموزش، پیش‌پردازشی جهت یکسان‌سازی تصاویر صورت گرفته است. پیش‌پردازشی که بر روی این داده‌ها صورت گرفته شامل برش ناحیه شخص در تصویر می‌باشد. این مجموعه داده‌ها شامل فایل‌ی با فرمت xml بوده که برچسب تصویر و همچنین موقعیت قرارگیری شخص در آن مشخص شده است. شکل ۳-۱۲ تعدادی از این تصاویر برای هشت زاویه مختلف را نشان می‌دهد.



شکل ۳-۱۲: تصاویر مجموعه داده مورد استفاده جهت آموزش (TUD) [۶۸]

^۱ Principle Component Analysis

^۲ Annotation

به منظور ارزیابی دقت کلاسه‌بند، مجموعه داده‌های ViPER مورد استفاده قرار گرفته است. این مجموعه داده شامل ۱۲۶۴ تصویر رنگی است که اشخاصی در جهت‌های مختلف را در برمی‌گیرد [۸]. جهات این مجموعه داده مشابه با مجموعه آموزشی به صورت گسسته در هشت جهت می‌باشد. شرایط نوری گوناگون، کیفیت پایین و وجود نویز از ویژگی‌های این مجموعه داده می‌باشد. در مرحله تست این مجموعه تصاویر به صورت سطح خاکستری تبدیل شده‌اند. دلیل انتخاب این مجموعه تصاویر، دشواری تشخیص زاویه و استفاده از این مجموعه در کاربردهای دیگر از جمله بازشناسی می‌باشد. در این مجموعه داده پراکندگی تصاویر در زوایای مختلف یکسان نیست. شکل ۳-۱۳ تعدادی از این تصاویر را نشان می‌دهد.



شکل ۳-۱۳: تصاویر نمونه از مجموعه داده‌های استفاده‌شده در تست (ViPER) [۸]

۳-۲-۱-۷- آزمایش‌ها و نتایج

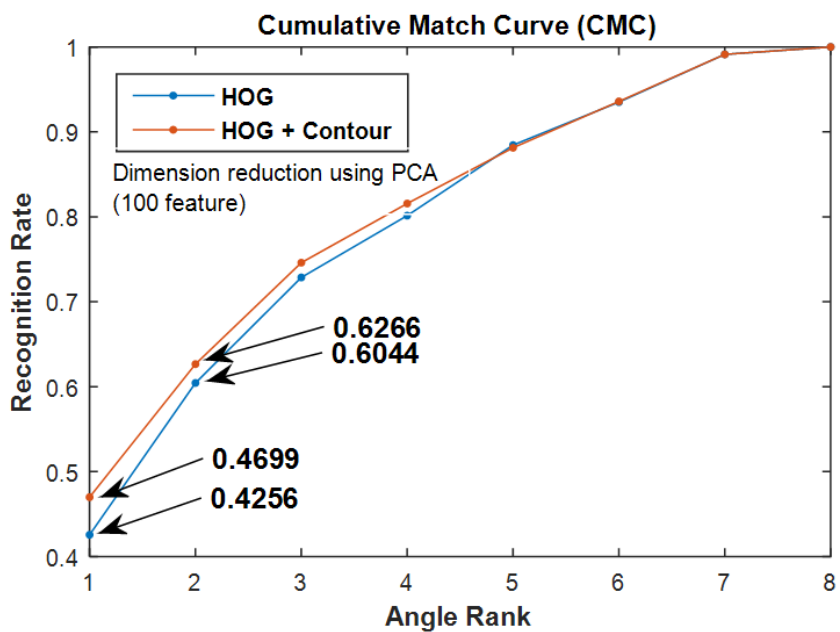
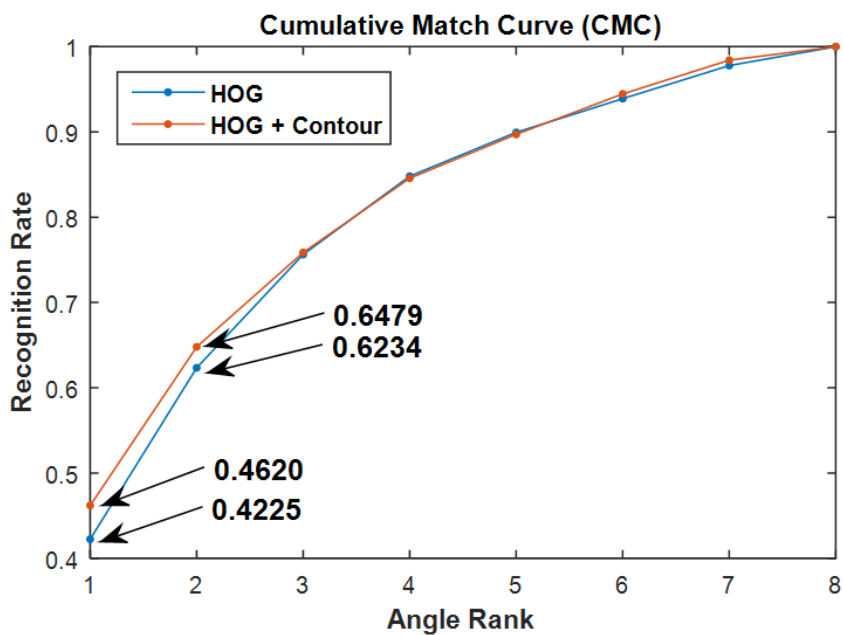
آزمایش‌های صورت گرفته در چهار حالت انجام شده است. در حالت اول تنها با استفاده از ویژگی هیستوگرام گرادیان‌های جهتی و بدون کاهش بعد و با استفاده از کلاسه‌بند SVM، داده‌های آزمون کلاسه‌بندی شده‌اند. حالت اول مشابه با مقالات [۵۳] و [۶۸] می‌باشد. در حالت دوم ویژگی پیشنهادی که برگرفته از اطلاعات کانتوری است به ویژگی HOG اضافه گردیده و بدون کاهش بعد با استفاده از کلاسه‌بند SVM، کلاسه‌بندی انجام شده است. در دو حالت دیگر با استفاده از آنالیز مؤلفه‌های اصلی، ابعاد بردارهای آموزشی ورودی کاهش داده می‌شود (۱۰۰ ویژگی). در دو حالت اخیر به دلیل کاهش قابل توجه تعداد ویژگی‌ها سرعت اجرای الگوریتم در مرحله آموزش و در مرحله تست به صورت

چشمگیری افزایش می‌یابد. نتایج آزمایشات در قالب دو نمودار نشان داده شده است. در این پژوهش از کلاسه‌بند SVM با مقدار خروجی احتمالاتی برای هر کلاس استفاده شده است. بدین ترتیب بر روی مقادیر احتمالاتی به دست آمده، عمل مرتب‌سازی صورت گرفته که بیشترین مقدار، برچسب کلاس خروجی است. مقدار دوم در مقادیر مرتب‌سازی شده کاندید دوم برای برچسب نمونه تستی می‌باشد. با استفاده از این اطلاعات، نمودار CMC^1 به دست آمده که اطلاعات دقیقی از رفتار کلاسه‌بند نهایی را در اختیار قرار می‌دهد. همچنین با استفاده از ماتریس درهم‌ریختگی^۲ مقایسه‌ای بین روش پایه و روش پیشنهادی انجام شده است.

شکل ۳-۱۴ نمودار CMC برای دو حالت با استفاده از روش پیشنهادی و حالت پایه نشان داده شده است. این نمودار نشان می‌دهد که بیشترین مقدار در خروجی کلاسه‌بند برای حالت پایه دارای دقت ۴۲٪ و برای حالتی که ویژگی پیشنهادی به ویژگی HOG اضافه گردیده دارای دقت ۴۶٪ می‌باشد. همین نمودار نشان می‌دهد احتمال وقوع برچسب درست در دو بزرگ‌ترین مقدار خروجی کلاسه‌بند، برای حالت پایه ۶۲٪ و برای حالت پیشنهادی ۶۴٪ می‌باشد.

^۱ Cumulative Match Curve

^۲ Confusion matrix



شکل ۳-۱۴: نتایج به دست آمده از کلاسه بندی نهایی با استفاده از ویژگی پیشنهادی بر روی نمودار CMC

شکل ۳-۱۵: ماتریس درهم ریختگی را برای نتایج به دست آمده در دو حالت نشان می دهد. در حالت اول تنها از ویژگی HOG در کلاسه بند SVM استفاده شده است (حالت پایه). در حالت دوم ویژگی HOG به همراه ویژگی پیشنهادی که مبتنی بر اطلاعات کانتوری است، در کلاسه بند SVM استفاده شده است.

Confusion Matrix (HOG + Contour)

90	185 14.6%	20 1.6%	16 1.3%	0 0.0%	0 0.0%	0 0.0%	11 0.9%	39 3.1%	68.3% 31.7%
135	52 4.1%	36 2.8%	10 0.8%	0 0.0%	0 0.0%	0 0.0%	5 0.4%	3 0.2%	34.0% 66.0%
180	0 0.0%	9 0.7%	37 2.9%	0 0.0%	0 0.0%	0 0.0%	8 0.6%	4 0.3%	63.8% 36.2%
225	1 0.1%	0 0.0%	7 0.6%	0 0.0%	0 0.0%	0 0.0%	4 0.3%	1 0.1%	0.0% 100%
270	55 4.4%	31 2.5%	40 3.2%	0 0.0%	0 0.0%	1 0.1%	58 4.6%	37 2.9%	0.0% 100%
315	3 0.2%	4 0.3%	7 0.6%	0 0.0%	0 0.0%	0 0.0%	24 1.9%	0 0.0%	0.0% 100%
0	8 0.6%	18 1.4%	76 6.0%	0 0.0%	0 0.0%	0 0.0%	265 21.0%	42 3.3%	64.8% 35.2%
45	39 3.1%	4 0.3%	16 1.3%	0 0.0%	0 0.0%	0 0.0%	27 2.1%	61 4.8%	41.5% 58.5%
	53.9% 46.1%	29.5% 70.5%	17.7% 82.3%	NaN% NaN%	NaN% NaN%	0.0% 100%	65.9% 34.1%	32.6% 67.4%	46.2% 53.8%
	90	135	180	225	270	315	0	45	

Target Class

Confusion Matrix (HOG)

90	164 13.0%	12 0.9%	7 0.6%	0 0.0%	0 0.0%	0 0.0%	11 0.9%	38 3.0%	70.7% 29.3%
135	57 4.5%	32 2.5%	8 0.6%	0 0.0%	0 0.0%	0 0.0%	2 0.2%	2 0.2%	31.7% 68.3%
180	0 0.0%	7 0.6%	37 2.9%	0 0.0%	0 0.0%	0 0.0%	8 0.6%	3 0.2%	67.3% 32.7%
225	2 0.2%	0 0.0%	7 0.6%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	1 0.1%	0.0% 100%
270	79 6.3%	48 3.8%	58 4.6%	0 0.0%	0 0.0%	1 0.1%	81 6.4%	53 4.2%	0.0% 100%
315	4 0.3%	4 0.3%	13 1.0%	0 0.0%	0 0.0%	0 0.0%	30 2.4%	0 0.0%	0.0% 100%
0	6 0.5%	15 1.2%	72 5.7%	0 0.0%	0 0.0%	0 0.0%	247 19.5%	36 2.8%	65.7% 34.3%
45	31 2.5%	4 0.3%	7 0.6%	0 0.0%	0 0.0%	0 0.0%	22 1.7%	54 4.3%	45.8% 54.2%
	47.8% 52.2%	26.2% 73.8%	17.7% 82.3%	NaN% NaN%	NaN% NaN%	0.0% 100%	61.4% 38.6%	28.9% 71.1%	42.2% 57.8%
	90	135	180	225	270	315	0	45	

Target Class

شکل ۳-۱۵: ماتریس درهم‌ریختگی برای نتایج کلاسه‌بندی نهایی

همان‌طور که مشاهده می‌شود در ماتریس درهم‌ریختگی، برای کلیه کلاس‌ها دقت بیشتری نسبت به حالت پایه به‌دست‌آمده است که این نشان‌دهنده وجود اطلاعات مفید در ویژگی پیشنهادی است. همچنین با توجه به نمودار شکل ۳-۱۴، ترکیب ویژگی پیشنهادی با ویژگی HOG، در مقایسه با ویژگی

حالت پایه، از دقت تشخیص بالاتری برخوردار است. میزان این بهبود در رتبه‌های یک و دو (Rank 1, Rank 2) تقریباً ۰.۴٪ می‌باشد. در تشریح بهبود کلاسه‌بندی با استفاده از ویژگی پیشنهادی می‌توان گفت که برخلاف هیستوگرام گرادیان جهتی که برگرفته از اطلاعات گرادیانی در نواحی محلی در تصویر می‌باشد، ویژگی پیشنهادی مبتنی بر اطلاعات کانتور تصویر است که دیدی یکپارچه نسبت به تصویر دارد. اطلاعات بدست آمده از ابرکانتور تطبیق‌یافته با تصویر ورودی، اطلاعات تکمیلی و متفاوتی نسبت به توصیفگر HOG در اختیار ما می‌گذارد. در نتیجه ترکیب ویژگی حاصله در روش پیشنهادی با روش پایه، منجر به بهبود در نتایج شده است. در زیربخش‌های پیش رو نحوه تعیین پارامترهای روش پیشنهادی و همچنین مقایسه روش پیشنهادی با روش پایه از منظر پیچیدگی زمانی، بررسی خواهد شد.

۳-۲-۱-۸- تعیین حد آستانه در ساخت و پیمایش درخت خوشه‌بندی

یکی از پارامترهای روش پیشنهادی، حد آستانه در ساخت و پیمایش درخت خوشه‌بندی می‌باشد. انتخاب درست این پارامتر بر پیچیدگی زمانی و دقت الگوریتم تأثیرگذار خواهد بود. همان‌طور که در بخش ۳-۳ ذکر شد پیمایش درخت خوشه‌بندی به گره‌برگی منتهی می‌شود که حاوی هیستوگرام زاویه‌ای از کانتورهای آن گره می‌باشد. مقدار آستانه بهینه در روش پیشنهادی مقداری است که ابرکانتورهای برگ از منظر زاویه نه خیلی خاص و نه خیلی کلی باشند. به عبارتی، می‌خواهیم حد آستانه‌ای را پیدا کنیم که هیستوگرام زاویه‌ای متناظر با برگ‌ها، یک توزیع تک‌قله‌ای باشد و همچنین واریانس این قله نیز متناسب با اندازه چند زاویه مجاور باشد. بررسی‌های صورت گرفته نشان می‌دهد که مقدار بهینه این حد آستانه با در نظر گرفتن فرض بالا، مقداری در حدود ۲۰۰ کانتور در برگ‌ها می‌باشد. در واقع این انتخاب براساس حداکثر عمقی از درخت است که مورد پیمایش قرار گرفته و در عین حال هیستوگرام زاویه‌ای متناظر با آن تک‌قله‌ای می‌باشد.

انتخاب تابع هسته^۱ مناسب در کلاسه‌بند SVM از اهمیت خاصی برخوردار است. این انتخاب به جنس داده‌های آموزشی بستگی دارد. به عبارت دیگر معیار شباهت بین داده‌ها می‌تواند ما را در انتخاب هسته مناسب راهنمایی کند. در اکثر کاربردها معیار شباهت اقلیدسی معیار مناسبی می‌باشد. در چنین حالتی هسته گاوسی که تابعی از فاصله اقلیدسی است، مناسب می‌باشد. در مواردی که معیار شباهت بین داده‌ها کسینوسی است هسته‌های خطی و چندجمله‌ای مناسب می‌باشند. همان‌طور که در بخش‌های قبل اشاره شد جنس مؤلفه‌های بردار ویژگی نهایی ترکیبی در این مقاله از جنس مؤلفه‌های هیستوگرام هستند. مجموعه این هیستوگرام‌ها به نوعی بافت کلی تصویر را مشخص می‌کنند. همچنین به دلیل وجود شرایط نوری متفاوت در تصاویر، تناسب و همبستگی مؤلفه‌های بردار ویژگی بیش از اختلاف اندازه این مؤلفه‌ها اهمیت دارند. لذا از بین انواع معیارهای ارائه شده برای محاسبه شباهت دو هیستوگرام یا دو توزیع احتمالاتی، معیارهایی نظیر باتچریا، کسینوسی، فیدلیتی^۲ که همگی بر پایه ضرب داخلی می‌باشند، انتخاب‌های مناسب‌تری خواهند بود [۶۹]. بر این اساس هسته‌هایی از قبیل خطی و چندجمله‌ای مناسب می‌باشند. در پیاده‌سازی‌های انجام شده با بررسی داده‌های آموزشی و تست، از هسته چندجمله‌ای با پارامتر درجه ۳ به‌عنوان هسته در ماشین بردار پشتیبان استفاده شده است. کلاسه‌بند SVM مورد استفاده در این مقاله برگرفته از بسته نرم‌افزاری LibSvm می‌باشد [۷۰].

۳-۲-۲- تشخیص زاویه با استفاده از ترکیب کلاسه‌بندهای رگرسیون لجستیک

در این رویکرد توصیفگر HOG، به دلیل سرعت در محاسبات به‌عنوان ویژگی اصلی جهت تخمین زاویه، انتخاب شده است. این توصیفگر بر روی چهار ناحیه در تصویر محاسبه می‌گردد [۲]. دو نوع ناحیه در روش پیشنهادی در نظر گرفته شده: نواحی ایستا و پویا. نواحی ایستا شامل سه ناحیه: کلی، بالاتنه و پایین‌تنه می‌باشند و ناحیه پویا شامل ناحیه‌ای از تصویر است که سر و شانه فرد در آن قرار دارد. بنابراین در ابتدا باید ناحیه پویا موقعیت‌یابی گردد. کلاسه‌بند رگرسیون لجستیک، هم به‌عنوان

^۱ Kernel function

^۲ Fidelity

کلاسه‌بند زاویه‌ای و هم به‌عنوان کلاسه‌بند به‌منظور موقعیت‌یابی، مورد استفاده قرار گرفته است. در این رویکرد نشان می‌دهیم که این کلاسه‌بند به لحاظ دقت کلاسه‌بندی و پیچیدگی زمانی از کارآمدی کافی برخوردار است. برای هر یک از چهار ناحیه ذکر شده در قبل، یک کلاسه‌بند رگرسیون لجستیک آموزش می‌یابد و تخمین نهایی، ترکیبی از خروجی این کلاسه‌بندها خواهد بود. از آنجایی که در روش پیشنهادی از ویژگی HOG و کلاسه‌بند رگرسیون لجستیک در سطح جزئیات استفاده شده، در ادامه به تشریح مختصری از هر یک می‌پردازیم.

۳-۲-۱- توصیفگر HOG

توصیفگر HOG یکی از مرسوم‌ترین توصیفگرها در پردازش تصویر و بینایی است. اولین بار این توصیفگر برای کاربرد تشخیص انسان در تصویر معرفی گردید. در واقع این توصیفگر ویژگی‌های ساختاری را برای نواحی مختلف تصویر، بازنمایی می‌کند. ایده اصلی این توصیفگر در این موضوع نهفته که بیشترین اطلاعات شکل و ساختار محتوای یک تصویر در لبه‌ها و اطلاعات گرادیانی آن منعکس می‌شود. به‌طور کلی در این توصیفگر ابتدا کل تصویر به زیرناحیه‌هایی با نام سلول تقسیم می‌شود. در هر سلول توزیع فراوانی یا هیستوگرام گرادیان‌های تصویر در چندین سطح گسسته محاسبه می‌شود. به‌منظور کاهش تأثیرات تغییرات نوری و سایه‌ها، نرمال‌سازی بر روی مقادیر چندین سلول مجاور که بلاک نام دارد صورت می‌گیرد. این بلاک‌ها می‌توانند در راستای افقی و عمودی همپوشانی داشته باشند. توصیفگر نهایی از کنار هم قرار دادن مقادیر هیستوگرام برای بلاک‌های مختلف حاصل می‌شود. از مزایای این توصیفگر مقاومت در برابر تغییرات نوری، تغییرات هندسی (با شدت پایین)، جابجایی‌های اندک در تصویر و همچنین پیچیدگی محاسباتی نسبتاً پایین می‌باشد. شکل ۳-۱۶ خلاصه‌ای از گام‌های توصیفگر HOG در محاسبه بردار ویژگی را نشان می‌دهد.

تعمیم یافته نسخه دودویی است. در این حالت پارامتر β به صورت یکپارچه برای همه کلاس‌ها با استفاده از بیشینه‌گر احتمال پسین^۱ برآورد می‌گردد. تابع توزیع احتمال پسین برای کلاس j بر اساس رابطه زیر محاسبه می‌شود.

$$\begin{cases} Pr(Y_i = j | X_i = x_i) = \frac{e^{F_j(x_i)}}{\sum_{k=1}^J e^{F_k(x_i)}} \\ F_j(x_i) = \beta_j^T \cdot x_i \end{cases} \quad (۸-۳)$$

در روش پیشنهادی از نسخه خاصی از رگرسیون لجستیک که به صورت افزایشی^۲ آموزش می‌یابد و همچنین عمل انتخاب ویژگی را نیز انجام می‌دهد استفاده شده که تحت عنوان Simple Logistic در ابزار Weka قابل دسترس است [۷۱]. خروجی این الگوریتم مقادیر بردار β برای هر کلاس می‌باشد.

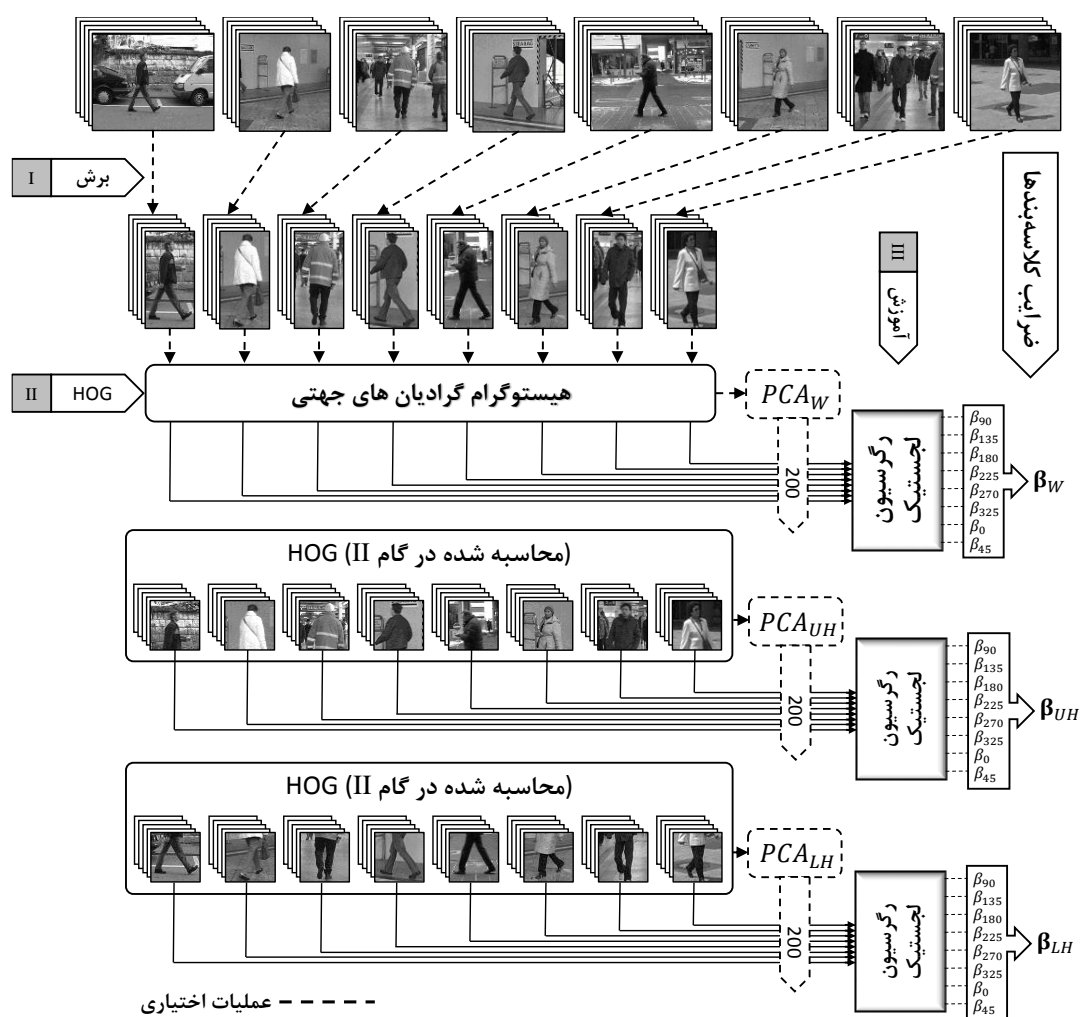
۳-۲-۲-۳- استخراج نواحی ایستا، پویا و آموزش کلاسه‌بندها

مجموعه داده مورد استفاده در این رویکرد، شامل تصاویری از افراد، در هشت زاویه قرارگیری نسبت به دوربین (0° ، 45° ، 90° ، 135° ، 180° ، 225° ، 270° ، 315°) می‌باشد [۶۸]. نه نقطه مشخص از هر فرد در تصاویر این مجموعه داده به صورت دستی علامت گذاری شده است. با استفاده از این نقاط ابتدا ناحیه کلی شخص را با ابعاد 200×200 پیکسل در 100×100 پیکسل جدا می‌کنیم. سپس نواحی ایستا که شامل کل عکس (W)، نیمه بالایی عکس (UH) و نیمه پایینی عکس (LH) از تصاویر استخراج می‌گردد. ناحیه پویا شامل ناحیه سر و شانه (HS) است که با استفاده از اطلاعات نقاط علامت گذاری شده استخراج می‌شود. از آنجایی که تأثیرات ناشی از جابجایی، تغییرات حالت، تداخل با پس‌زمینه، انسدادهای جزئی و پوشش‌های خاص، در کل نواحی تصویر یکسان نیست، لذا ترکیب خروجی چهار کلاسه‌بند متناظر با چهار ناحیه، می‌تواند منجر به بهبود نتایج گردد.

^۱ Maximum a posteriori

^۲ Boosting

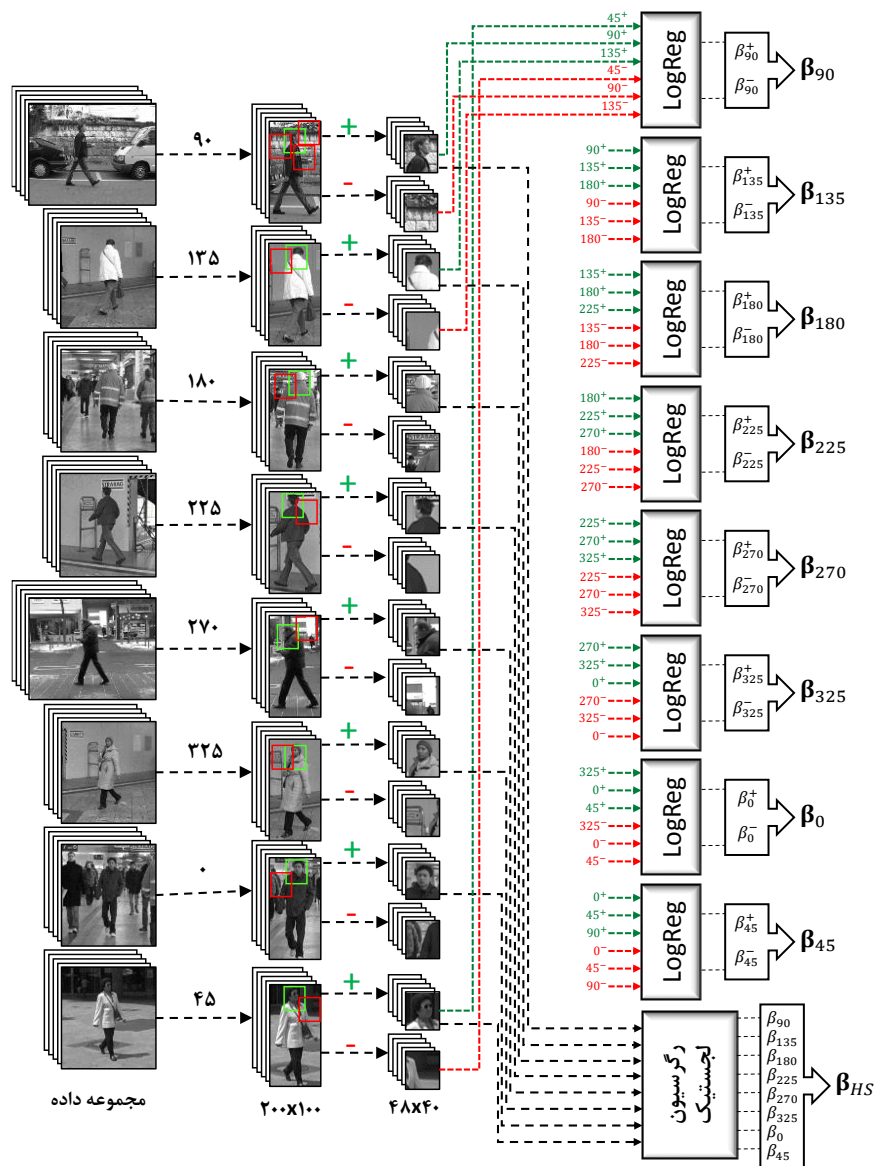
پس از استخراج نواحی مذکور، برای هر ناحیه با توجه به زاویه آن، یک کلاسه‌بند رگرسیون لجستیک آموزش داده می‌شود. از نقاط قوت این کلاسه‌بند می‌توان به سادگی و پیچیدگی محاسباتی پایین در هنگام تست و همچنین قابلیت تعمیم‌پذیری بالا برای داده‌های رؤیت نشده اشاره نمود [۷۲]. برای نواحی (LH, UH, W) بعد از محاسبه توصیفگر HOG، خروجی مذکور با استفاده از نگاشت PCA به ۲۰۰ ویژگی کاهش می‌یابد. این عمل کاهش بعد به صورت اختیاری در مراحل انجام الگوریتم اجرا می‌شود. استفاده از PCA زمان آموزش و تأثیرات نویز را کاهش می‌دهد اما پیچیدگی در هنگام تست را (به دلیل انجام عملیات نگاشت) افزایش می‌دهد. شکل ۳-۱۷ خلاصه گام‌های فاز آموزش را در روش پیشنهادی نشان می‌دهد.



شکل ۳-۱۷: آموزش کلاسه‌بندهای مرتبط با نواحی W, UH و LH

موقعیت‌یابی ناحیه HS با استفاده از کلاسه‌بند باینری بر روی ویژگی HOG انجام می‌شود. از آنجایی که اندازه ناحیه HS مقدار ۴۸ پیکسل در ۴۰ پیکسل است، بنابراین در توصیفگر HOG تعداد ۶ سلول در ۵ سلول را اشغال می‌کند. در نتیجه پنجره‌ای با ابعاد ۶ در ۵ بر روی ویژگی HOG که قبلاً یک‌بار محاسبه شده بود لغزنده می‌شود. در هر موقعیت کلاسه‌بند باینری مذکور مشخص می‌کند که پنجره مورد نظر به کلاس HS یا کلاس غیر HS تعلق می‌گیرد. در نتیجه حاصل این عمل، نقشه‌ای است که هر نقطه از آن مشخص‌کننده درجه وقوع کلاس HS در آن موقعیت از تصویر است.

آزمایشات انجام شده نشان می‌دهد که دقت تخمین زاویه با تنها استفاده از ناحیه W برای تخمین دقیق مقدار ۵۰ درصد و برای تخمین با حد خطای $\pm 45^\circ$ مقدار ۷۰ درصد می‌باشد. در روش پیشنهادی از این موضوع در جهت موقعیت‌یابی دقیق‌تر ناحیه HS استفاده شده است. برای این منظور هشت کلاسه‌بند باینری برای هشت زاویه را، به صورت مجزا آموزش می‌دهیم. در مرحله آموزش نمونه‌های مثبت شامل نواحی است که سر و شانه در آن واقع شده و نمونه‌های منفی شامل نواحی است که تنها بخشی از ناحیه هدف در آن واقع شده است (مساحت همپوشان با ناحیه هدف بین ۱۰ تا ۴۰ درصد). برای هر یک از هشت کلاسه‌بند (برای مثال کلاسه‌بند برای زاویه i°) نمونه‌های مثبت از مجموعه زاویه‌های $\{(i-45)^\circ, i^\circ, (i+45)^\circ\}$ از تصاویر پایگاه و نمونه‌های منفی نیز به همین ترتیب انتخاب می‌شود. شکل ۳-۱۸ خلاصه‌ای از عملیات صورت گرفته برای آموزش کلاسه‌بندهای مرتبط با ناحیه HS به منظور موقعیت‌یابی و تخمین زاویه نشان می‌دهد.



شکل ۳-۱۸: خلاصه‌ای از عملیات انجام شده در آموزش کلاسه‌بندهای مرتبط با ناحیه HS

در مرحله تست زمانی ابتدا یکبار توصیفگر HOG برای تصویر محاسبه می‌شود. سپس با استفاده از کلاسه‌بند مرتبط با ناحیه W تخمین اولیه‌ای از زاویه تصویر صورت می‌گیرد. بر اساس زاویه بدست آمده در مرحله قبل یکی از هشت کلاسه‌بند مرتبط با ناحیه HS به منظور موقعیت‌یابی انتخاب می‌گردد. همان‌طور که در بخش‌های قبل گفتیم خروجی کلاسه‌بند رگرسیون لجستیک از ضرب داخلی بردار $\langle I, x_1, x_2, \dots, x_n \rangle$ در بردار ضرایب β بدست می‌آید. بنابراین لغزاندن پنجره 5×6 سلولی بر روی توصیفگر HOG و محاسبه خروجی کلاسه‌بند همانند عمل کانولوشن می‌باشد. این لغزاندن با گام‌های

۸ پیکسلی (متناسب با عرض سلول‌های توصیفگر) انجام می‌شود. بنابراین خروجی الگوریتم موقعیت‌یابی، نقشه‌ای از احتمال وقوع ناحیه HS در نقاط می‌باشد که دقتی برابر ۸ پیکسل دارد. فرض کنید هر موقعیت در نقشه بدست آمده، یک متغیر تصادفی گسسته چند مؤلفه‌ای باشد. این متغیر تصادفی دو مؤلفه x و y دارد. در نتیجه هر نقطه از نقشه، مشاهده‌ای از یک متغیر تصادفی چند مؤلفه‌ای است. بنابراین امید ریاضی این متغیر تصادفی موقعیت دقیقی از ناحیه HS را به ما می‌دهد. تفسیر دیگری از امید ریاضی در این موضوع، میانگین وزن دار مشاهدات می‌باشد. رابطه ۳-۹ طریقه محاسبه مقدار امید را نشان می‌دهد.

$$E[X] = \frac{(p_1x_1 + p_2x_2 + \dots + p_nx_n)}{p_1 + p_2 + \dots + p_n} \quad (9-3)$$

$$x_i = (a, b)$$

$$a \in \{1, 8, 16, 24, \dots, 80\}, \quad 10 \text{ cell rows}$$

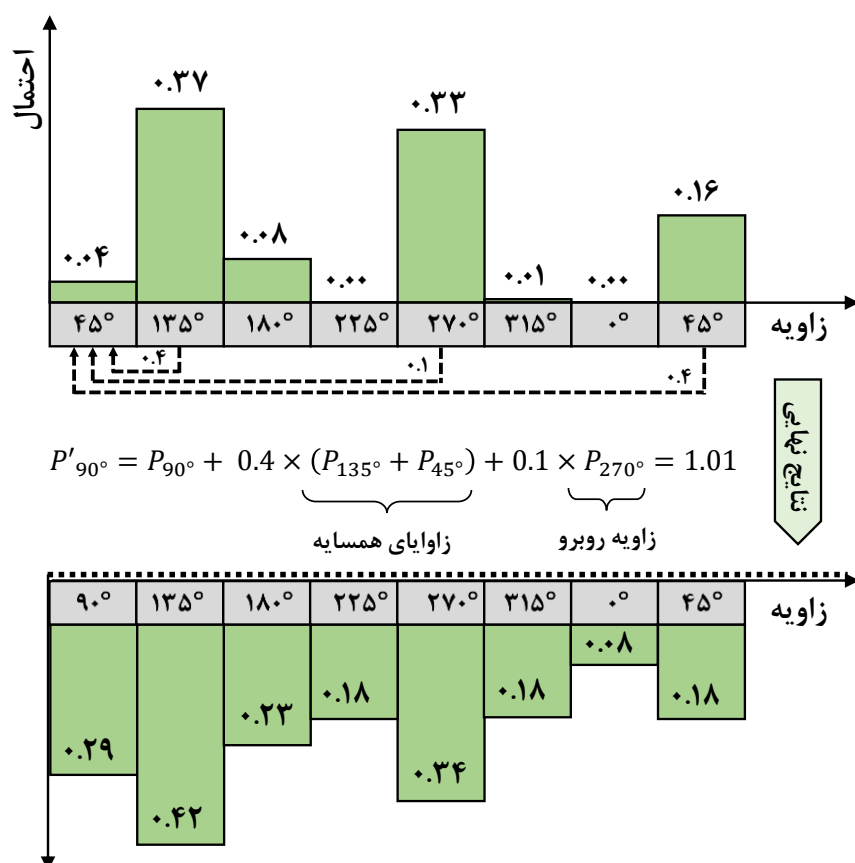
$$b \in \{1, 8, 16, 24, \dots, 88\}, \quad \text{all columns}$$

در رابطه بالا برای موقعیت‌یابی ناحیه سر و شانه تنها ۱۰ سطر سلول اول و تمامی ستون‌ها مورد بررسی قرار می‌گیرند. همچنین مقادیری از خروجی کلاسه‌بند که درجه وقوع ناحیه HS مربوط به آن کمتر از ۰,۵ باشد به مقدار صفر تغییر می‌یابد.

۳-۲-۲-۴- ترکیب نتایج کلاسه‌بندها

در این بخش خروجی چهار کلاسه‌بند که شامل توزیع احتمال بر روی هشت زاویه است، با هم ترکیب شده و خروجی نهایی را تشکیل می‌دهند. در بسیاری از مسائل کلاسه‌بندی نمونه‌های کلاس‌های مختلف با یکدیگر شباهتی ندارند. برای مثال در مسئله تشخیص اشیاء کلاس‌هایی نظیر هواپیما، گربه وجود دارد که به لحاظ شکل کاملاً بی‌ارتباط می‌باشند. اما در مسئله تشخیص زاویه، هشت کلاس زاویه به یکدیگر مرتبط می‌باشند. برای مثال تصاویر انسان در زاویه قرارگیری 90° نسبت به دوربین شباهت‌هایی به تصاویر در زوایای 45° و 135° و همچنین زاویه روبروی (270°) آن دارد. بدین ترتیب قبل از ترکیب چهار کلاسه‌بند، از یک مکانیزم وزن‌دهی، برپایه ایده ذکر شده، استفاده می‌کنیم. در این

مکانیزم وزن‌دهی، دو ثابت یکی برای تأثیرات زوایای همسایه (N) و دیگری برای تأثیرات زاویه روبرو (O) در نظر گرفته شده است. شکل ۱۹-۳ مکانیزم وزن‌دهی را برای مقادیر پارامتر $N=0.4$ و $O=0.1$ نشان می‌دهد.



شکل ۱۹-۳: مثالی از مکانیزم وزن‌دهی روش پیشنهادی برای خروجی یکی از کلاس‌بندها

در گام بعد نتایج وزن‌دهی شده کلاس‌بندها با استفاده از ضرب مؤلفه‌های متناظر کلاس‌بندها با هم ترکیب می‌گردند. بدین ترتیب بزرگ‌ترین مقدار در عبارت نهایی، اندیس کلاس را مشخص می‌کند.

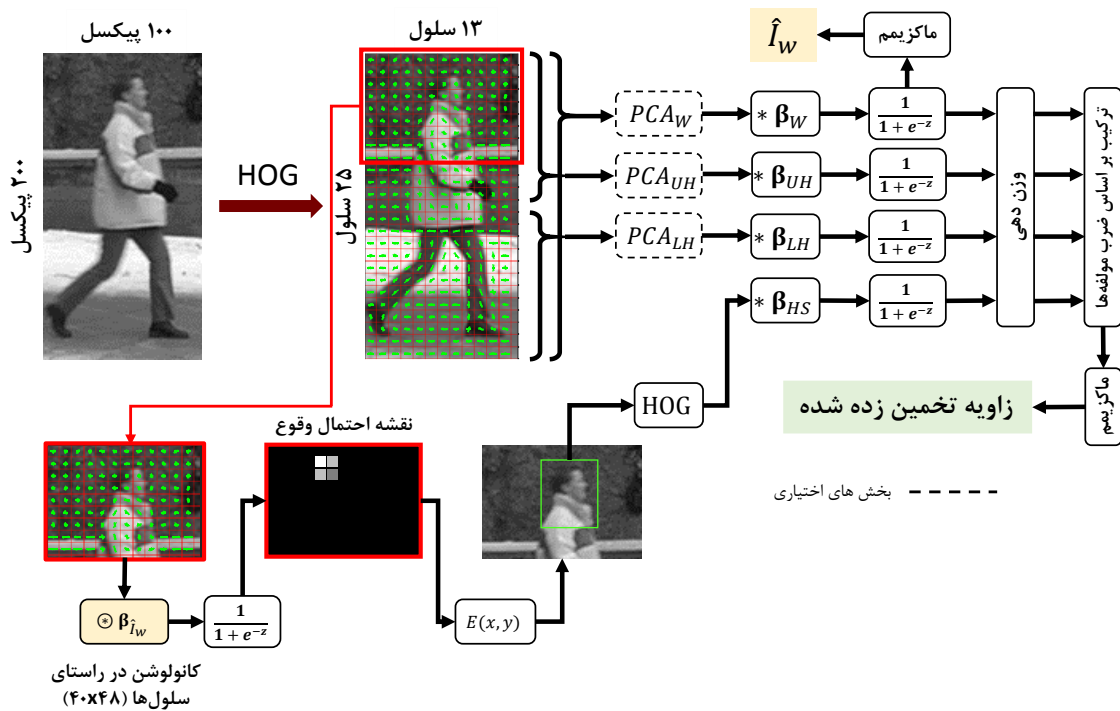
$$h_{final}(x) = \underset{j}{argmax} \mu_j(x) \quad (۱۰-۳)$$

$$\mu_j(x) = \prod_{r \in R} p_{r,j}(x)$$

$$R = \{W, UH, LH, HS\}$$

در این عبارت، مقدار $p_{r,j}$ مقدار تعلق به کلاس j برای خروجی نتایج کلاس‌بند ناحیه r می‌باشد. شکل ۲۰-۳

خلاصه‌ای از کلیه بخش‌های سیستم پیشنهادی را نشان می‌دهد.



شکل ۳-۲۰: خلاصه‌ای از کلیه بخش‌های سیستم پیشنهادی برای تشخیص زاویه

۳-۲-۵- نتایج آزمایش‌ها

مجموعه داده‌های مورد استفاده، به منظور آموزش و تست، مشابه الگوریتم پیشنهادی قبلی است. در این بخش دقت تخمین زاویه روش پیشنهادی با روش‌های [۵۷-۶۱، ۶۳، ۶۸] مقایسه شده است. آزمایش‌های صورت گرفته روش پیشنهادی در دو بخش دقت کلاسه‌بندی و پیچیدگی محاسباتی با روش‌های موجود مقایسه می‌شود. در بخش اول میانگین دقت کلاسه‌بندی برای تخمین دقیق زاویه و تخمین با حد خطای $\pm 45^\circ$ محاسبه و نتایج مقایسات با روش‌های موجود در جدول ۳-۱ مقایسه شده است.

جدول ۱-۳: مقایسه نتایج کلاسه‌بندی کلی روش پیشنهادی و روش‌های موجود

دقت تشخیص		روش
(دقیق) °	(نادقیق) ±۴۵°	
۵۷,۹%	۸۳,۷%	روش پیشنهادی
۵۳,۰%	۸۱,۵%	[۶۳] ERT+MoAWG
۵۳,۲%	۷۸,۸%	[۵۹, ۵۷] HOG+SVM+PCA
۵۱,۶%	۷۹,۰%	[۶۸] HOG+SVM(1-vs-all)+PCA
۳۸,۵%	۷۹,۶%	[۶۸] HOG+SVM-adj(1-vs-all)+PCA
۳۷,۰%	۶۴,۳%	[۵۹, ۵۷] HOG+SVM
۲۶,۰%	۵۱,۶%	[۵۸] Wavelet+SVM+PCA
۲۵,۸%	۴۵,۰%	[۶۱] CovDescriptor

همچنین جدول ۲-۳ و جدول ۳-۳ ماتریس درهم‌ریختگی را برای روش پیشنهادی و الگوریتم [۶۳] که بهترین نتایج ارائه شده تاکنون است را نشان می‌دهد. این نتایج برتری روش پیشنهادی نسبت به روش‌های کنونی از منظر دقت کلاسه‌بندی را نشان می‌دهد.

جدول ۲-۳: ماتریس درهم‌ریختگی برای الگوریتم ارائه شده در [۶۳]

زاویه تخمین زده شده	90	135	180	225	270	315	0	45	
90	49 9.9%	11 2.2%	1 0.2%	1 0.2%	3 0.6%	2 0.4%	0 0.0%	18 3.6%	57.6% 42.4%
135	4 0.8%	11 2.8%	3 0.6%	2 0.4%	0 0.0%	0 0.0%	3 0.6%	3 0.6%	46.4% 53.6%
180	0 0.0%	4 0.8%	46 9.3%	10 2.0%	0 0.0%	2 0.4%	13 2.6%	0 0.0%	61.3% 38.7%
225	0 0.0%	0 0.0%	1 0.2%	28 6.8%	6 1.2%	0 0.0%	0 0.0%	1 0.2%	77.8% 22.2%
270	1 0.2%	3 0.6%	1 0.2%	23 4.6%	33 6.7%	17 3.4%	3 0.6%	3 0.6%	39.3% 60.7%
315	0 0.0%	5 1.0%	14 2.8%	9 1.8%	1 0.2%	52 10.5%	24 4.8%	2 0.4%	48.6% 51.4%
0	0 0.0%	0 0.0%	14 2.8%	0 0.0%	0 0.0%	2 0.4%	21 5.4%	8 1.6%	52.9% 47.1%
45	2 0.4%	2 0.4%	2 0.4%	2 0.4%	0 0.0%	0 0.0%	7 1.4%	15 3.0%	50.0% 50.0%
	87.5% 12.5%	34.2% 65.8%	56.1% 43.9%	37.3% 62.7%	76.7% 23.3%	69.3% 30.7%	35.1% 64.9%	30.0% 70.0%	53.0% 47.0%
	90	135	180	225	270	315	0	45	

جدول ۳-۳: ماتریس درهم‌ریختگی برای روش پیشنهادی

90	40 8.1%	4 0.8%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	7 1.4%	76.9% 23.1%
135	9 1.8%	15 3.0%	4 0.8%	2 0.4%	0 0.0%	0 0.0%	2 0.4%	0 0.0%	46.9% 53.1%
180	0 0.0%	8 1.6%	56 11.3%	14 2.8%	0 0.0%	1 0.2%	11 2.2%	2 0.4%	60.9% 39.1%
225	1 0.2%	0 0.0%	3 0.6%	37 7.5%	8 1.6%	5 1.0%	3 0.6%	2 0.4%	62.7% 37.3%
270	1 0.2%	0 0.0%	3 0.6%	13 2.6%	30 6.0%	10 2.0%	1 0.2%	2 0.4%	50.0% 50.0%
315	1 0.2%	1 0.2%	2 0.4%	4 0.8%	2 0.4%	43 8.7%	7 1.4%	4 0.8%	67.2% 32.8%
0	0 0.0%	6 1.2%	11 2.2%	2 0.4%	2 0.4%	14 2.8%	44 8.9%	11 2.2%	48.9% 51.1%
45	4 0.8%	4 0.8%	3 0.6%	2 0.4%	1 0.2%	2 0.4%	9 1.8%	22 4.4%	46.8% 53.2%
	71.4% 28.6%	39.5% 60.5%	68.3% 31.7%	49.3% 50.7%	69.8% 30.2%	57.3% 42.7%	57.1% 42.9%	44.0% 56.0%	57.9% 42.1%
	90	135	180	225	270	315	0	45	

کلاس هدف

به‌منظور انجام مقایسات از منظر پیچیدگی محاسباتی، تعداد عملیات مورد نیاز شمارش می‌شود. این عملیات شامل: ضرب، جمع، مقایسه و ارجاع می‌باشند.

• رگرسیون لجستیک

بر طبق رابطه ۳-۸ ضرب داخلی بردار β و بردار ویژگی x نیازمند $\|\beta\|^{01}$ ضرب و $\|\beta\|^{0}$ عمل جمع است. همچنین برای هر یک از این دو عمل یک عمل ارجاع به بردارهای β و x می‌باشد. برای سادگی، یک عمل جمع، یک عمل ضرب و دو عمل ارجاع را یک واحد محاسباتی MA2I می‌نامیم.

• درخت تصمیم

^۱ $\|\cdot\|^{0}$: نرم صفر به معنای تعداد عناصر غیر صفر بردار

در یک درخت تصمیم باینری، برچسب کلاس‌ها در برگ‌ها قرار دارد. برای پیمایش درخت سه عمل انجام می‌شود: دو عمل مقایسه (مقایسه با کلید و بررسی برگ بودن یا نبودن) و یک عمل ارجاع (انتقال به زیردرخت راست یا چپ). این مجموعه عملیات را یک واحد محاسباتی I2C می‌نامیم.

• SVM

پیچیدگی محاسباتی ماشین بردار پشتیبان در مرحله تست وابسته به تعداد بردارهای پشتیبانی است که در مرحله آموزش بدست آمده است. عبارت ۱۱-۳ رابطه مرز تصمیم‌گیری ماشین بردار پشتیبان را در حالتی که از تابع هسته استفاده شود برای حالت دو کلاسه نشان می‌دهد.

$$y = \text{sgn} \left(\sum_{i=1}^n w_i y_i k(x_i, x') \right) \quad (11-3)$$

که در آن $x' \in R^n$ یک نمونه تستی و $x_i \in R^n$ i امین نمونه آموزشی می‌باشد. $y_i \in \{-1, +1\}$ خروجی i امین نمونه آموزشی و w_i وزن i امین نمونه آموزشی در این رابطه را نشان می‌دهد. در مقایسات انجام شده تابع هسته k تابع تانژانت سیگموئیدی انتخاب شده است. در این رابطه نمونه‌های آموزشی متناظر با بردارهای پشتیبان، مقدار w_i غیر صفری خواهند داشت. براساس این رابطه در ماشین بردار پشتیبان برای تعیین کلاس خروجی، ضرب داخلی بردار ورودی با تمامی بردارهای پشتیبان را خواهیم داشت. در صورتی که n بردار پشتیبان داشته باشیم به تعداد n واحد محاسباتی MA2I نیاز می‌باشد.

به‌منظور مقایسه دو واحد محاسباتی MA2I و I2C پیاده‌سازی در زبان C# صورت گرفت و این دو واحد به‌صورت تکراری در میلیون‌ها بار اجرا شدند. نتایج اجرا نشان می‌دهد که $O(I2C) = 1.23 \times O(MA2I)$. بنابراین پیچیدگی زمانی واحد MA2I کمتر از I2C می‌باشد. جدول ۳-۴ مقایسه‌ای از

روش پیشنهادی و روش‌های دیگر موجود را از منظر پیچیدگی محاسباتی نشان می‌دهد. بر طبق این نتایج، روش پیشنهادی از روش درختان تصادفی دارای پیچیدگی کمتری بوده و همچنین الگوریتم پیشنهادی قابلیت پیاده‌سازی موازی را دارا می‌باشد. از منظر پیچیدگی فضای حافظه، روش پیشنهادی نیازمند ذخیره‌سازی تنها مؤلفه‌های غیر صفر بردارهای β می‌باشد که در آزمایش‌های انجام شده در مجموع ۱۲۰۰ عدد اعشاری ذخیره می‌شود که به صورت قابل توجهی از الگوریتم‌های موجود عملکرد بهتری دارد

جدول ۳-۴: مقایسه پیچیدگی محاسباتی روش پیشنهادی و روش‌های دیگر موجود

روش‌ها	پیچیدگی محاسباتی
روش پیشنهادی	$O(HOG) + MA2I$ $\times (\ \beta_W\ _0 + \ \beta_{UH}\ _0 + \ \beta_{UL}\ _0 + \ \beta_{HS}\ _0) \times 8 + \ \beta_{Loc}\ _0 \times k$ $\cong O(HOG) + MA2I$ $\times ((157 + 178 + 193 + 108) \times 8 + 674 \times 32)$ $= O(HOG) + MA2I \times 26656$
Extremely Randomized Tree+MoAWG [۶۳]	$O(HOG) \times (1 + 0.25 + 0.125) + I2C \times (d \times n \times 8)$ $\cong O(HOG) \times 1.375 + I2C \times (10 \times 300 \times 8)$ $= O(HOG) \times 1.375 + MA2I \times 1.23 \times 24000$ $= O(HOG) \times 1.375 + MA2I \times 29520$
HOG with SVM+ PCA [۵۷, ۵۹]	$O(HOG) + MA2I \times$ $(D_{PCA} \times Dim(x) + nSV \times D_{PCA} \times D_{PCA})$ $\cong O(HOG) + MA2I \times (200 \times 9500 + 2700 \times 200 \times 200)$ $= O(HOG) + MA2I \times 109900000$

$O(\cdot)$: پیچیدگی محاسباتی الگوریتم (O بزرگ)

k : تعداد نواحی مورد بررسی در موقعیت یابی ناحیه HS

d : میانگین عمق پیمایش درخت از ریشه به برگ‌ها

n : تعداد درخت‌های تصادفی برای کلاسه بندی یک زاویه

D_{PCA} : تعداد بردارهای ویژه انتخاب شده در الگوریتم PCA برای کاهش بعد

$Dim(x)$: تعداد مؤلفه‌های بردار ویژگی

nSV : تعداد بردارهای پشتیبان در SVM آموزش یافته

۳-۳- استخراج ناحیه سر در تصویر

یکی از بخش‌های مهم از رویکرد پیشنهادی، استخراج ناحیه سر در تصویر می‌باشد. این ناحیه کلیدی هم اطلاعات مهمی از زاویه قرارگیری شخص را درون خود دارد و هم مشخص‌کننده موقعیت نسبی بالاتنه شخص می‌باشد. بنابراین استخراج این ناحیه باید با دقت بالا صورت گیرد. بدین منظور می‌توان از توصیفگرهای مختلف و کلاسه‌بندهای مختلفی جهت استخراج این ناحیه استفاده کرد از جمله توصیفگر HOG، LBP و کلاسه‌بندهایی نظیر SVM و رگرسیون لجستیک. الگوریتم کلی استخراج ناحیه، مشابه بخش استخراج ناحیه HS در تشخیص زاویه است. دقت در استخراج این ناحیه تأثیر مستقیمی بر عملکرد الگوریتم پیشنهادی نهایی خواهد گذاشت.

بررسی‌های صورت گرفته نشان داد که دقت الگوریتم استخراج ناحیه HS برای هدف نهایی ما، کافی نمی‌باشد. بدین منظور از شبکه عصبی کانولوشن (CNN) برای استخراج این ناحیه از تصویر استفاده گردید. بنابراین در زیر بخش آتی مروری اجمالی بر شبکه عصبی کانولوشن خواهیم داشت. سپس به معرفی مجموعه داده‌های آموزشی و معماری شبکه مورد استفاده خواهیم پرداخت. همچنین روش جدیدی برای دستیابی به دقت بیشتر در موقعیت‌یابی ناحیه سر ارائه می‌گردد. در نهایت مقایساتی بین روش پیشنهادی و الگوریتم استخراج ناحیه HS انجام خواهیم داد که نتایج، دربردارنده دقت بسیار بالاتر رویکرد پیشنهادی می‌باشند.

۳-۳-۱- شبکه عصبی کانولوشن

شبکه عصبی کانولوشن یکی از الگوریتم‌های یادگیری ماشین است که به لحاظ ساختاری جزء شبکه‌های عصبی مصنوعی پیش‌خور^۱ عمیق قرار می‌گیرد. به دلیل موفقیت‌های اخیر سخت‌افزاری و استفاده راحت و استاندارد از پردازنده‌های گرافیکی در انجام محاسبات موازی، استفاده از این نوع شبکه

^۱ Feed-forward artificial neural network

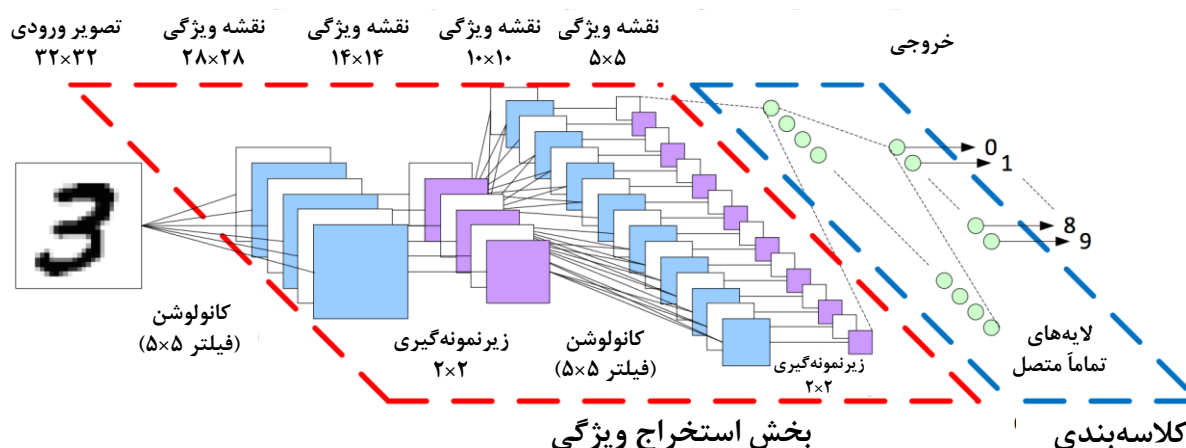
مورد توجه ویژه محققین قرار گرفته است. به طوری که در بسیاری از کاربردهای پردازش تصویر و بینایی ماشین بهترین نتایج را شبکه‌های عصبی کانولوشن عمیق ارائه نموده‌اند. معماری و ساختار این نوع شبکه نیز روز به روز در حال ارتقاء می‌باشد که از مهم‌ترین این ساختارها می‌توان به AlexNet، VGG، GoogleNet و Microsoft ResNet اشاره نمود [۷۳-۷۶].

این نوع شبکه، نسخه‌ای از شبکه پرسپترون چندلایه^۱ است که به دلیل ساختار ویژه آن، نیازمند کمترین عملیات پیش‌پردازشی توسط شخص استفاده‌کننده را دارد. از ویژگی‌های بسیار مهم این نوع شبکه، عدم وابستگی شبکه به جابجایی مکانی عناصر تصویر می‌باشد. این ویژگی محصول ساختار وزن‌دهی اشتراکی^۲ است. براساس این ساختار، یکی از لایه‌های اصلی و کلیدی شبکه ایجاد می‌گردد که لایه کانولوشن می‌باشد. همان‌طور که می‌دانیم عملگر کانولوشن، شامل فیلتری است که بر روی یک تصویر اعمال گردیده و درواقع مقادیر این فیلتر برای کل تصویر مقدار مشترکی است. مکانیزم وزن‌دهی اشتراکی نیز ایده مشابهی را در خود دارد [۷۷].

در بسیاری از روش‌های یادگیری ماشین، استخراج ویژگی جزء اولین گام‌های طراحی سیستم می‌باشد. استخراج ویژگی مناسب، از عملیات زمان‌بری است که شخص خبره با بررسی‌ها و آزمایش‌های گوناگون در جهت یافتن آن می‌باشد. اما شبکه عصبی کانولوشن در لایه‌های نخستین شبکه، مرحله استخراج ویژگی را درون خود دارد. به عبارتی فیلترهای کانولوشن در این نوع شبکه تحت مکانیزم یادگیری، تعیین می‌گردند. در نتیجه بهترین نوع ویژگی با استفاده از این روش قابل دستیابی است که نتایج، در حوزه‌های مختلف، مؤید این موضوع است. شکل ۳-۲۱ یکی از معماری‌های مرسوم این نوع شبکه را برای مسئله تشخیص ارقام دست‌نویس نشان می‌دهد.

^۱ Multilayer perceptron

^۲ Shared-weights architecture



شکل ۳-۲۱: ساختار کلی یک شبکه عصبی کانولوشن برای مسئله تشخیص ارقام دستنویس لاتین

لایه‌های مرسوم در این نوع شبکه که در تحقیق جاری نیز استفاده شده به شرح زیر است:

- **لایه کانولوشن:** در این لایه عملگر کانولوشن بر روی ورودی اعمال می‌شود. هر یک از نورون‌های این لایه، ویژگی خاصی را از تصویر استخراج می‌کنند. در نتیجه هرچه در لایه‌های کانولوشن شبکه پیش می‌رویم، ویژگی‌های سطح بالاتری استخراج می‌گردد.
- **لایه تجمیع^۱:** در این لایه نورون‌های یک لایه با یکدیگر ترکیب می‌شوند. برای مثال در حالت ماکزیمم، بیشترین مقدار نورون‌های لایه جاری به لایه بعد انتقال می‌یابد.
- **لایه ReLU^۲:** با افزودن این تابع فعالسازی خاص به لایه‌های شبکه، خاصیت غیرخطی بودن شبکه بیشتر می‌گردد و در نتیجه عملکرد نهایی شبکه بهبود می‌یابد.
- **لایه تماماً متصل^۳:** در این لایه هر یک از نورون‌ها به تمامی نورون‌های لایه بعد متصل می‌گردند.

^۱ Pooling layer

^۲ Fully connected layer

^۳ Rectified Linear Units

- لایه تابع اتلاف^۱: این تابع نحوه انتشار خطا به آخرین لایه از شبکه را مشخص می‌کند. برای مثال در حالت کلاسه‌بندی از تابع اتلاف آنروپی متقاطع سیگموئیدی^۲ استفاده می‌گردد.

۳-۲- آماده‌سازی داده‌ها و انتخاب معماری شبکه

یکی از محدودیت‌های استفاده شبکه عصبی کانولوشن، نیازمندی آن به تعداد زیاد داده‌های آموزشی می‌باشد. برای این منظور مجموعه داده‌های HollywoodHeads به‌منظور آموزش استفاده شده است [۷۸]. این مجموعه داده شامل ۳۶۹۸۴۶ تصویر سر مشخص شده در ۲۲۴۷۴۰ فریم ویدئویی از ۲۱ فیلم هالیوود می‌باشد. تصاویر سر در این مجموعه داده از زوایای مختلفی نظیر روبرو، پهلو و پشت است. لذا شبکه عصبی که در نهایت آموزش می‌یابد قادر به شناسایی ناحیه سر در هر زاویه دلخواهی می‌باشد. ابتدا تصاویر نواحی سر از این مجموعه داده استخراج می‌شود. سپس این نواحی به اندازه 16×16 پیکسل و به‌صورت سطح خاکستری تغییر می‌یابد. برای داده‌های منفی به ازای هر تصویر سر، دو تصویر: پس‌زمینه و تصویری با همپوشانی جزئی با ناحیه سر انتخاب شده است. بنابراین فرض بر این است که ناحیه سر در تصویر، در پنجره‌ای به اندازه 16×16 پیکسل قرار دارد. در هنگام تست، پنجره‌ای با این اندازه، بر روی بخش بالایی تصویر لغزانده می‌شود. برای هر موقعیت، تصویر متناظر، به‌عنوان ورودی به شبکه عصبی کانولوشن داده می‌شود و خروجی، درجه عضویت به کلاس سر را نشان خواهد داد. در نتیجه مقدار ماکزیمم در نقشه خروجی، مکان ناحیه سر را مشخص می‌کند. بر این اساس معماری شبکه به‌صورت جدول ۳-۵ در نظر گرفته شد.

^۱ Loss layer

^۲ Sigmoid cross-entropy

جدول ۳-۵: معماری شبکه عصبی کانولوشن برای تشخیص ناحیه سر از ناحیه غیر سر

شماره لایه	نام لایه	ابعاد خروجی لایه
۱	لایه ورودی	۱۶×۱۶
۲	لایه کانولوشن (۵ فیلتر ۶×۶×۱ با اندازه گام ۲×۲)	۶×۶
۳	لایه ReLU	۶×۶
۴	لایه کانولوشن (۱۰ فیلتر ۳×۳×۵ با اندازه گام ۲×۲)	۴×۴
۵	لایه ReLU	۴×۴
۶	لایه تجمیع (ماکزیمم‌گیری ۲×۲ با اندازه گام ۲×۲)	۲×۲
۷	لایه تماما متصل ۲ نورونی	۲×۱
۸	لایه اتلاف از نوع Softmax	۲×۱
۹	لایه خروجی	۲×۱

شکل ۳-۲۲ پنج فیلتر آموزش یافته با اندازه ۶×۶ در لایه دوم شبکه عصبی کانولوشن مذکور نشان می‌دهد.



شکل ۳-۲۲: پنج فیلتر آموزش یافته در لایه دوم شبکه عصبی

با دقت در این مقادیر شاهد آن هستیم که در لایه‌های شبکه عصبی کانولوشن به‌مرور بافت‌ها و الگوهای خاصی استخراج می‌شود که در واقع عمل استخراج ویژگی با کارآمدی بالایی صورت می‌گیرد. نحوه انتخاب معماری شبکه پیشنهادی بدین صورت است که ابتدا از یک شبکه با لایه‌های کم و تعداد فیلترهای کم در لایه‌های کانولوشن شروع می‌کنیم. سپس با اضافه کردن به تعداد فیلترها در هر لایه و همچنین تعداد لایه‌ها تا زمانی که به دقت قابل قبول در مجموعه آموزشی دست یابیم شبکه را بزرگ‌تر می‌کنیم. بدین ترتیب به شبکه‌ای دست خواهیم یافت که علاوه بر دارا بودن دقت کافی در طبقه‌بندی، پیچیدگی محاسباتی حداقلی نیز خواهد داشت.

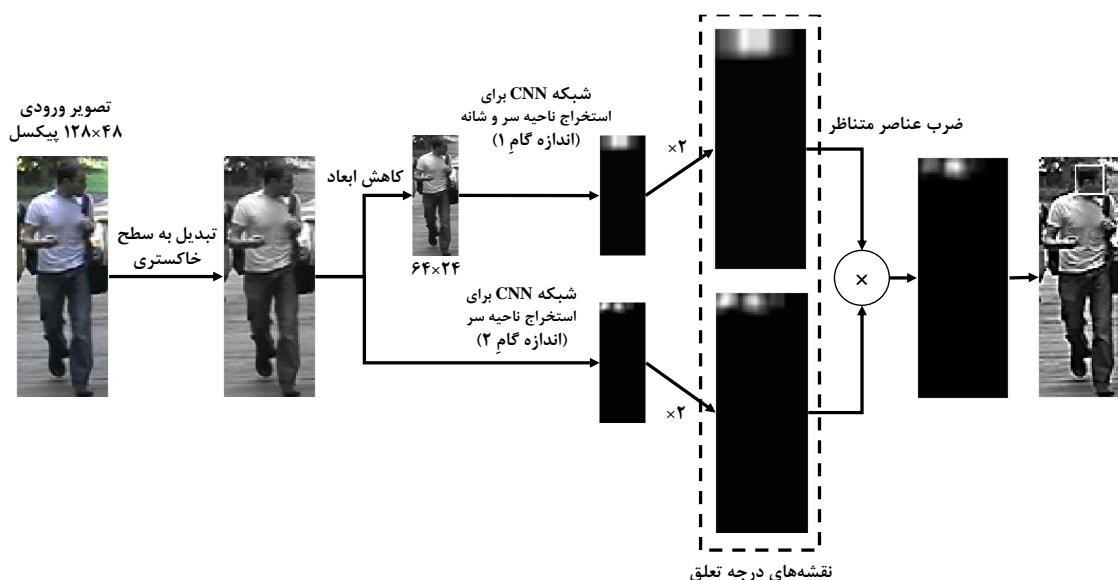
در استفاده از شبکه آموزش یافته بالا درمی‌یابیم که در بسیاری حالات و نواحی در تصویر به‌اشتباه به‌عنوان ناحیه سر، درجه تعلق بالایی به خود می‌گیرد. دلیل این موضوع شباهت بخشی از پس‌زمینه به سر، انسداد و تنوع ظاهری ناحیه سر می‌باشد. در چنین مواردی در نظر گرفتن ناحیه بزرگتری نظیر سر و شانه می‌تواند بر مشکلات ذکرشده در بالا غلبه نماید و در تصویر قابل تشخیص باشد. بنابراین استفاده توأمان از خروجی دو کلاسه‌بند یکی برای ناحیه سر و دیگری برای ناحیه سر و شانه می‌تواند دقت نهایی را افزایش دهد. برای این منظور ابتدا ناحیه سر و شانه را در تصاویر بریده و سپس ناحیه مذکور به اندازه 16×16 پیکسل تغییر می‌یابد. طریقه انتخاب نواحی منفی نیز مشابه کلاسه‌بند مربوط به ناحیه سر می‌باشد. جدول ۳-۶ معماری شبکه عصبی کانولوشن در استخراج ناحیه سر و شانه را نشان می‌دهد.

جدول ۳-۶: معماری شبکه عصبی کانولوشن برای تشخیص ناحیه سر و شانه

شماره لایه	نام لایه	ابعاد خروجی لایه
۱	لایه ورودی	16×16
۲	لایه کانولوشن (۵ فیلتر $5 \times 5 \times 1$ با اندازه گام 1×1)	12×12
۳	لایه ReLU	12×12
۴	لایه تجمیع (ماکزیمم‌گیری 2×2 با اندازه گام 2×2)	6×6
۵	لایه کانولوشن (۱۰ فیلتر $3 \times 3 \times 5$ با اندازه گام 1×1)	4×4
۶	لایه ReLU	4×4
۷	لایه تجمیع (ماکزیمم‌گیری 2×2 با اندازه گام 2×2)	2×2
۸	لایه تماماً متصل ۲ نورونی	2×1
۹	لایه اتلاف از نوع Softmax	2×1
۱۰	لایه خروجی	2×1

همان‌طور که قبلاً اشاره شد خروجی کلاسه‌بندها بر روی تصویر، نقشه‌ای از درجه تعلق به کلاس هدف را نشان می‌دهد. در روش پیشنهادی، الگوریتمی استفاده شده که در عین کاهش پیچیدگی محاسباتی، از دقت بیشتری در استخراج ناحیه سر نیز برخوردار است. برای این منظور با استفاده از شبکه آموزش یافته برای ناحیه سر، تصویر ورودی را با اندازه گام دو مرور کرده و بدین ترتیب نقشه‌ای

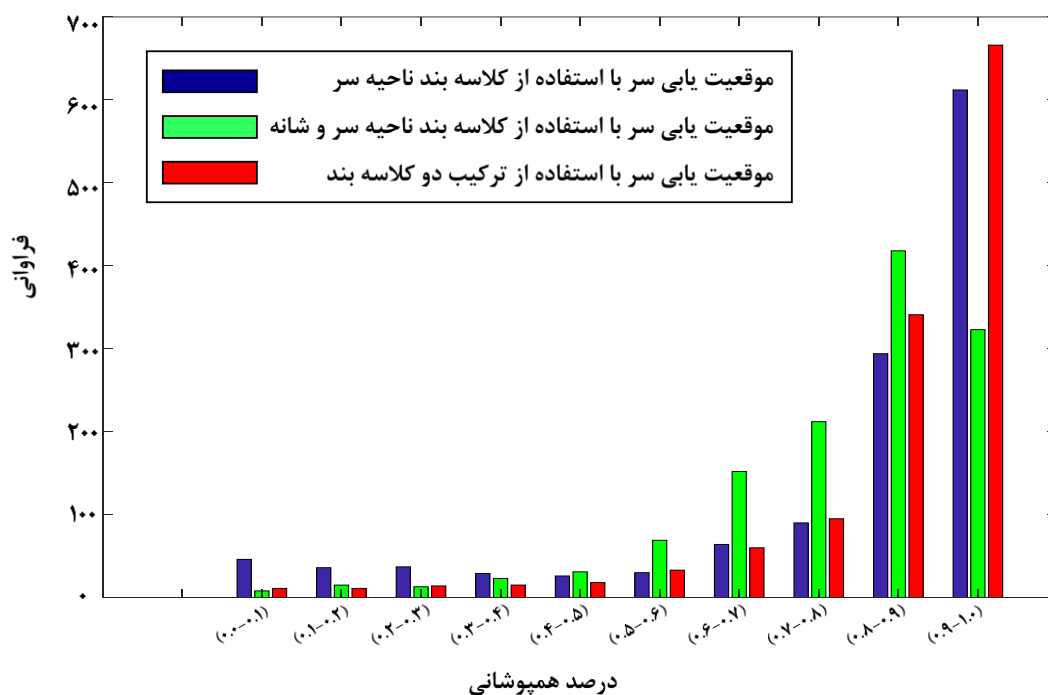
با اندازه یک‌چهارم سایز تصویر حاصل می‌گردد. در نتیجه میزان محاسبات یک‌چهارم حالتی است که اندازه گام ۱ باشد. در گام بعد اندازه تصویر ورودی به مقدار نصف کاهش می‌یابد. سپس به کمک کلاسه‌بند ناحیه سر و شانه، تصویر ورودی را با اندازه گام یک مرور کرده و نقشه متناظر با آن حاصل می‌شود. بدین ترتیب دو نقشه‌ی درجه تعلق، خواهیم داشت که به لحاظ اندازه برابر می‌باشند. بدین ترتیب با ضرب مؤلفه‌های متناظر این دو نقشه به نقشه‌ای خواهیم رسید که دقت بالاتری در استخراج ناحیه سر خواهد داشت. شکل ۳-۲۳ بخش‌های الگوریتم مذکور را به همراه یک مثال نشان می‌دهد.



شکل ۳-۲۳: استخراج موقعیت ناحیه سر با استفاده از ترکیب خروجی دو شبکه عصبی کانولوشن

الگوریتم پیشنهادی در مقایسه با حالتی که کل تصویر را با استفاده از شبکه عصبی کانولوشن سر و با گام یک مرور کنیم دارای پیچیدگی زمانی کمتر و به میزان نصف حالت اول خواهد بود. به منظور نشان دادن کارآمدی روش پیشنهادی، موقعیت ناحیه سر در مجموعه داده ViPER به صورت دستی علامت‌گذاری، و خروجی شبکه بر روی این مجموعه داده تست گردید. نمودارهای شکل ۳-۲۴ نتایج را در چهار حالت با هم مقایسه می‌کند. در این نمودارها میزان همپوشانی ناحیه استخراج شده، با ناحیه مطلوب محاسبه می‌گردد. سپس برای کل مجموعه داده، هیستوگرام فراوانی با فواصل یک درصد، محاسبه می‌شود. برای مثال نمودار میله‌ای قرمز رنگ در شکل ۳-۲۴، نشان می‌دهد که در ۶۶۶ تصویر

از ۱۲۶۴ تصویر مجموعه داده، همپوشانی ناحیه تخمین زده شده توسط الگوریتم پیشنهادی با ناحیه مطلوب، بین ۹۰ تا ۱۰۰ درصد می‌باشد. این نتایج نشان می‌دهد که رویکرد ترکیبی از دقت بالاتری نسبت به هریک از روش‌ها به تنهایی برخوردار است.

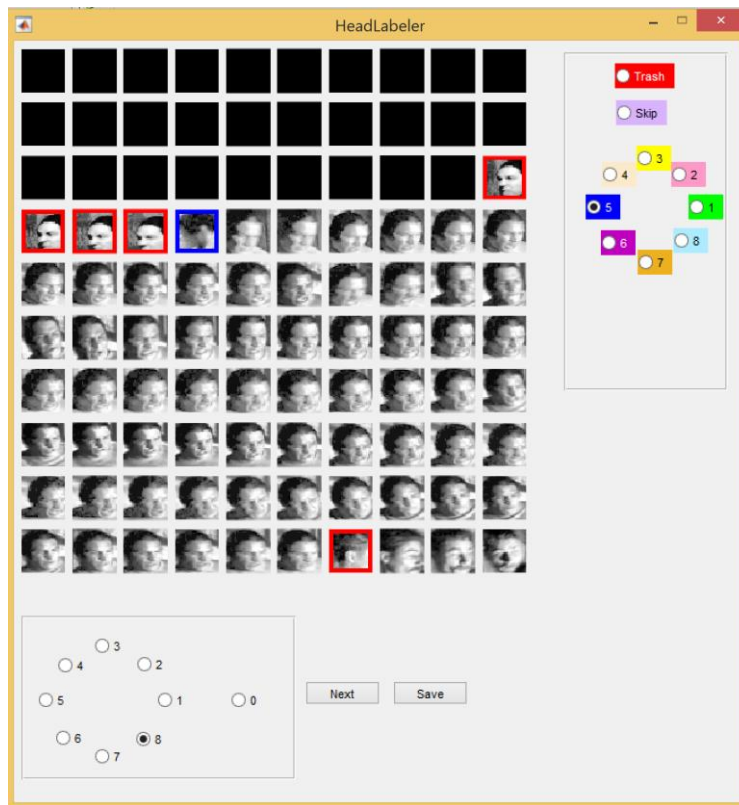


شکل ۳-۲۴: مقایسه هیستوگرام فراوانی درصد همپوشانی با ناحیه مطلوب، برای سه روش موقعیت یابی ناحیه سر، بر روی داده‌های ViPER

۳-۳-۳- استفاده از شبکه عصبی کانولوشن در تشخیص زاویه

به منظور افزایش دقت در تخمین زاویه، زیرسیستم استخراج موقعیت سر و شانه را با استخراج ناحیه سر با استفاده از شبکه عصبی کانولوشن که در زیربخش پیشین تشریح شد، جایگزین کردیم. بدین ترتیب ناحیه سر به دو منظور استخراج می‌گردد: برای عملیات یکسان‌سازی و تخمین زاویه. بعد از اینکه ناحیه مربوط به سر مشخص گردید زاویه قرارگیری سر در تصویر نیز باید تعیین گردد. بدین منظور مجموعه داده‌های HollywoodHeads به صورت دستی برچسب‌گذاری زاویه‌ای شده‌اند. از آنجایی که تعداد تصاویر این مجموعه بسیار زیاد است نرم‌افزاری جهت سرعت بخشیدن به عملیات برچسب‌گذاری تهیه گردید.

با توجه به اینکه تصاویر این مجموعه داده، از فریم‌های ویدئویی پشت سر هم و مربوط به ۲۵ فیلم است، بنابراین ارتباطی بین تصاویر وجود دارد. همچنین زاویه سر مربوط به یک شخص در طول فریم‌های متوالی تغییر چندانی نمی‌کند. لذا به منظور برچسب‌گذاری، ابتدا نواحی سر در تصاویر استخراج می‌گردد. سپس این نواحی به سطح خاکستری تبدیل شده و اندازه آن به ابعاد 32×32 پیکسل تغییر می‌یابد. الگوریتم نرم‌افزار تهیه‌شده بدین صورت است که ۱۰۰۰ تصویر ناحیه سر از مجموعه تصاویر استخراج می‌شود. سپس بردار ویژگی HOG برای این ۱۰۰۰ تصویر محاسبه، و با استفاده از الگوریتم K-Means خوشه‌بندی می‌گردند. در هنگام نمایش، این خوشه‌ها به صورت مجزا به کاربر نمایش داده می‌شوند. در صورتی که تعداد عناصر یک خوشه بیش از ۱۰۰ نمونه بود، در مجموعه‌های ۱۰۰ تایی به کاربر نمایش داده می‌شود. بنابراین با این مکانیزم این امکان فراهم می‌شود که تصاویر مشابه و در تعداد قابل توجه به کاربر نمایش داده شده و کاربر با انتخاب زاویه مربوطه به صورت یکجا، تعداد زیادی تصویر را برچسب‌گذاری کند. شکل ۳-۲۵ نمایی از نرم‌افزار تهیه‌شده را نشان می‌دهد.



شکل ۳-۲۵: نرم‌افزار نوشته‌شده برای برچسب‌گذاری زاویه برای مجموعه داده HollywoodHeads

در این نرم‌افزار هر یک از زوایا، با رنگی مشخص می‌شود. رنگ قرمز نشان‌دهنده تصاویری است که مناسب آموزش شبکه نیست. ناحیه پایین در سمت چپ، زاویه نمونه‌هایی را تعیین می‌کند که با رنگی مشخص نشده‌اند. برای مثال در شکل ۳-۲۵ موارد قرمز تصاویر نامناسب و مورد آبی تصویری با زاویه 270° را مشخص می‌کند. بقیه تصاویر از آنجایی که در کادر پایین مقدار ۸ انتخاب شده، در بردارنده تصاویری با برچسب زاویه 45° می‌باشند.

با استفاده از این نرم‌افزار مجموعه داده HollywoodHeads برچسب‌گذاری گردید و با استفاده از آن کلاسه‌بندی با ۸ خروجی و با استفاده از شبکه عصبی کانولوشن آموزش یافت و این کلاسه‌بند جایگزین بخش استخراج ناحیه سر و شانه در سیستم تخمین زاویه گردید.

۳-۴- یکسان سازی تصاویر

همان طور که در ابتدای این فصل بیان شد مقصود از یکسان سازی، به دو صورت استخراج ماسک‌هایی برای وزن‌دهی پوشش شخص در تصویر و همچنین نمونه‌افزایی نواحی استخراج شده بر اساس نواحی مجاور می‌باشد. استفاده از گام یکسان سازی برای محاسبه شباهت تصویر آزمون با هر تصویر به صورت مجزا صورت می‌گیرد. بررسی‌های صورت گرفته نشان می‌دهد که تنها بخش بالاتنه پوشش فرد تحت چرخش زاویه‌ای دستخوش تغییر می‌شود. لذا در رویکرد پیشنهادی تنها برای بخش بالاتنه ماسک متناظر با آن استخراج و همچنین نمونه‌افزایی برای نواحی مذکور صورت می‌گردد.

پس از محاسبه زاویه قرارگیری شخص در تصویر و استخراج ناحیه سر، با استفاده از الگوریتم زیر نواحی خاصی از بالاتنه که تحت چرخش زاویه‌ای آشکار و نهان می‌شود، در صورت وجود استخراج می‌گردد. الگوریتم زیر زمانی استفاده می‌شود که زاویه تخمینی از روبرو باشد، در غیر این صورت تمامی نواحی تصویر مورد استفاده قرار می‌گیرد و عمل یکسان سازی انجام نمی‌شود. الگوریتم استخراج این نواحی به صورت گام‌های زیر است:

۱- تصویر ورودی بخش‌بندی^۱ می‌گردد (با استفاده از ابزار VLFeat).

$$Seg(I) = \bigcup_{i=1}^n I_i \quad (12-3)$$

۲- برای هر یک از n بخش، بردار ویژگی شامل رنگ در فضای رنگی Lab و موقعیت مرکز بخش مورد نظر، تشکیل می‌شود.

$$f_i = \left(L(I_i), a(I_i), b(I_i), x(I_i^{center}), y(I_i^{center}) \right) \quad (13-3)$$

در این رابطه I_i^{center} پیکسل مرکزی در بخش i ام می‌باشد.

^۱ Segmentation

۳- بردارهای ویژگی با استفاده از الگوریتم K-Means خوشه‌بندی می‌گردند (۴ خوشه)

$$\begin{cases} Cluster^{kmean(k=4)}(F) = \bigcup_{k=1}^4 C_k \\ C_k = \bigcup_{j \in \{1, \dots, n\}} I_j^k \end{cases} \quad (۱۴-۳)$$

۴- ماسکی براساس موقعیت قرارگیری ناحیه سر ساخته می‌شود (M).

۵- در صورتی که نواحی I_j^k دارای شرایط زیر باشند به‌عنوان نواحی خاص در تصویر علامت‌گذاری می‌گردند.

$$Select(\hat{I}_j^k) = \begin{cases} 1 & \frac{\hat{I}_j^k \circ M}{|\hat{I}_j^k|_{\oplus}} > \tau_1 \quad \& \quad \frac{\hat{C}_k \circ M}{|\hat{C}_k|_{\oplus}} > \tau_2 \\ 0 & else \end{cases} \quad (۱۵-۳)$$

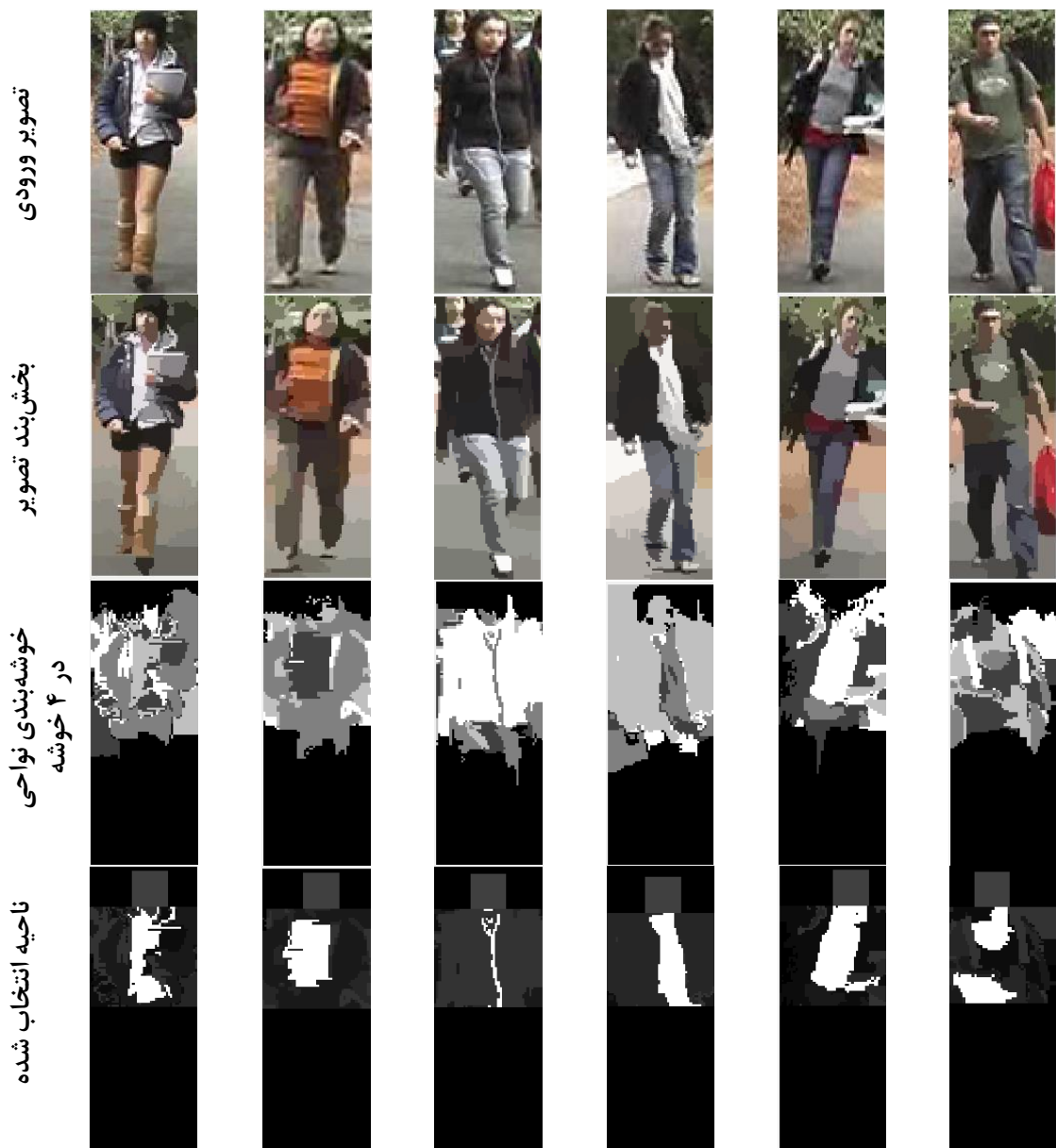
در این رابطه \hat{I}_j^k و \hat{C}_k مقدار باینری متناظر با نواحی I_j^k و C_k هستند. عملگر \circ مجموع ضرب درایه‌های متناظر دو ماتریس می‌باشد و همچنین تابع $|\cdot|_{\oplus}$ مجموع درایه‌های ماتریس آرگومان خود را محاسبه می‌کند. رابطه ۱۵-۳ برای یک ناحیه خاص، در صورتی که بیش از τ_1 درصد و خوشه متناظر با آن ناحیه نیز بیش از τ_2 درصد آن در زیر ناحیه سر واقع شده باشد انتخاب می‌گردد.

۶- برای تمامی نواحی I_j^k تابع $Select(\cdot)$ شرایط انتخاب شدن را بررسی می‌کند. بدین ترتیب ناحیه خروجی (O) از الگوریتم پیشنهادی به‌صورت زیر انتخاب می‌شود.

$$O = \bigcup_{j=1}^n \hat{I}_j^k \times Select(\hat{I}_j^k) \quad (۱۶-۳)$$

۷- اتمام

شکل ۳-۲۶ خروجی الگوریتم بالا را برای چند نمونه تصویر برای مقادیر $\tau_1 = 0.7$ و $\tau_2 = 0.4$ نشان می‌دهد.



شکل ۳-۲۶: مثالی از اجرای الگوریتم پیشنهادی در استخراج نواحی خاص به منظور یکسان سازی

۳-۵- تلفیق مرحله یکسان سازی با الگوریتم های بازشناسی

به منظور ارزیابی رویکرد پیشنهادی، دو الگوریتم بازشناسی [۲۵] و [۲۸] انتخاب و مرحله یکسان سازی در گام استخراج ویژگی آنها استفاده شده است. در [۲۸] بازنمایی جدیدی با نام ماکزیمم وقوع محلی به عنوان بردار ویژگی، برپایه هیستوگرام معرفی شده است. همچنین معیار فاصله و زیرفضای

جدیدی تحت عنوان^۱ XQDA معرفی شده است. تفاوت روش یادگیری معیار فاصله در این مقاله با بقیه روش‌های هم‌گروه، در کاهش بعدی است که توأمان با یادگیری معیار فاصله صورت می‌گیرد. رویکرد پیشنهادی در این رساله تغییری در قسمت یادگیری معیار فاصله ایجاد نمی‌کند. در [۲۵] بازنمایی جدیدی تحت عنوان توصیفگر گاوسی سلسله مراتبی^۲ برپایه توصیفگر کوواریانس معرفی شده است. در این بازنمایی برای نواحی مختلف تصویر توزیع آماری توصیفگرهای کوواریانس برای زیر نواحی کوچکتر محاسبه می‌شود و در نتیجه خصوصیات رنگی و بافتی برای هر ناحیه بازنمایی می‌گردد. در زیربخش‌های بعدی به نحوه تلفیق رویکرد پیشنهادی در بازنمایی‌های ذکر شده در بالا خواهیم پرداخت.

۳-۵-۱- ویژگی ماکزیمم وقوع محلی و تلفیق با روش پیشنهادی

در ویژگی معرفی شده در [۲۸] ابتدا تبدیلی براساس الگوریتم Retinex به صورت چندمقیاسی^۳ بر روی تصاویر ورودی به‌عنوان پیش‌پردازش انجام می‌شود. در این تبدیل شرایط نوری و رنگی تصویر با الگوبرداری از سیستم ادراک بینایی انسان اصلاح می‌گردد. سپس تصویر حاصله به فضای رنگی HSV نگاشت می‌گردد. در این ویژگی علاوه بر توصیف رنگ، از توصیفگر SILTP که نسخه بهبودیافته‌ای از توصیفگر LBP است به‌منظور توصیف بافت‌ها در تصویر استفاده می‌شود [۷۹]. از ویژگی‌های این توصیفگر مقاومت در برابر تغییرات شرایط نوری و وجود نویز در تصویر است. همانطور که در فصل دوم بدان اشاره شد بازنمایی معرفی شده در این مقاله (LOMO) از هیستوگرام‌های رنگی HSV و SILTP در سه مقیاس تشکیل می‌گردد.

به‌منظور ماسک‌گذاری و وزن‌دهی نواحی استخراج شده در تصویر و ترکیب با توصیفگر LOMO، باید از هیستوگرام وزن‌دار بجای هیستوگرام مرسوم استفاده کرد. بدین ترتیب که برای پنجره‌های لغزنده شده بر روی تصویر، نگاشت HSV و SILTP، مشابه حالت قبل محاسبه می‌گردد. با توجه به بازه‌های

^۱ Crossview Quadratic Discriminant Analysis (XQDA)

^۲ Hierarchical Gaussian Descriptor
^۳ Multiscale

فراوانی مورد نظر در محاسبه هیستوگرام HSV، برای هر پیکسل از تصویر مقداری عددی خواهیم داشت که حاصل چندی‌سازی در هر کانال رنگی است. خروجی توصیفگر SILTP نیز تصویری شامل مقادیر عددی با توجه به پارامترهای توصیفگر می‌باشد. در این مرحله با در دست داشتن ماسک وزنی، هیستوگرام وزن‌دار بر طبق رابطه زیر قابل محاسبه می‌باشد.

$$wHist^i(I, M) = \sum_{\{x,y|I(x,y)=i\}} M(x, y) \quad (17-3)$$

در این رابطه اندیس i ، بازه فراوانی خاصی در هیستوگرام وزن‌دار نهایی را نشان می‌دهد. به عبارتی در محاسبه هیستوگرام وزن‌دار، به ازای وقوع یک مقدار خاص در تصویر، به اندازه وزن متناظر با موقعیت آن پیکسل در ماتریس ماسک، شمارش می‌گردد. بدین ترتیب در مرحله یکسان‌سازی زمانی که تصمیم بر ماسک کردن نواحی از تصویر آزمون یا تصویری از مجموعه گالری باشد، با استفاده از روش بالا محاسبه هیستوگرام‌های HSV و SILTP را برای سه مقیاس تغییر می‌دهیم.

در روش نمونه‌افزایی ابتدا ناحیه ماسک در تصویر مشابه قبل استخراج می‌گردد. سپس برای سطرهایی از تصویر که شامل پیکسل‌های ماسک هستند روال زیر را اجرا می‌کنیم:

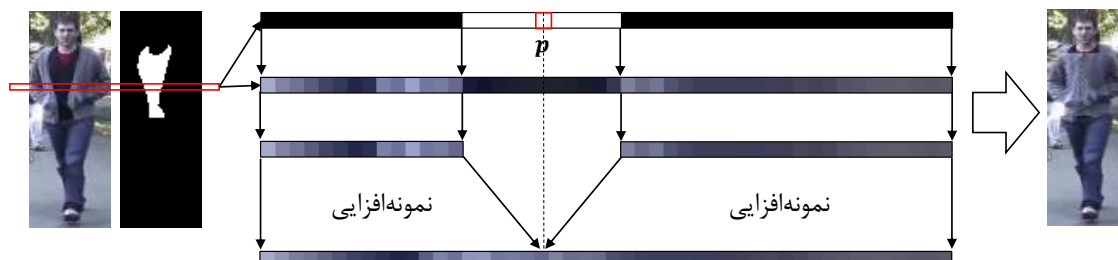
۱- موقعیت پیکسل میانی در پیکسل‌های ماسک مشخص می‌گردد (p).

۲- پیکسل‌های غیرماسک در سمت راست پیکسل p به اندازه فاصله پیکسل p تا ضلع راست تصویر، نمونه‌افزایی می‌شود.

۳- پیکسل‌های غیرماسک در سمت چپ پیکسل p به اندازه فاصله پیکسل p تا ضلع چپ تصویر، نمونه‌افزایی می‌شود.

۴- از کنار هم قرار دادن بخش‌های نمونه‌افزایی شده سطر جدید حاصل می‌گردد.

در شکل ۳-۲۷ خلاصه روال فوق برای یک سطر از تصویر و ماسک متناظر، نشان داده شده است. در نهایت برای تصویر نمونه‌افزایی شده ویژگی LOMO به صورت استاندارد قابل محاسبه می‌باشد.



شکل ۳-۲۷: خلاصه الگوریتم نمونه‌افزایی برای یک تصویر و ماسک استخراج شده متناظر آن

۳-۵-۲- توصیفگر گاوسی سلسله‌مراتبی و تلفیق با روش پیشنهادی

همانطور که در فصل دوم اشاره شد در توصیفگر گاوسی سلسله‌مراتبی، برای نواحی و زیرنواحی تصویر، بصورت سلسله‌مراتبی و پیاپی، مدلی گاوسی به منظور توصیف رنگ و بافت ایجاد می‌شود. این مدل گاوسی شامل بردار میانگین و ماتریس کوواریانس می‌باشد. بدین ترتیب که ابتدا تصویر به نواحی سطری بخش‌بندی می‌گردد. سپس برای هر ناحیه زیرنواحی استخراج می‌شوند. برای هر زیرناحیه مدل گاوسی تک‌مده شامل بردار میانگین و ماتریس کوواریانس محاسبه شده و سپس تبدیل به برداری در فضای مماسی اقلیدسی می‌گردد. در نتیجه برای هر ناحیه مجموعه‌ای از بردارها خواهیم داشت که مجدداً برای آنها مدل گاوسی دیگری تخمین زده می‌شود. بردار ویژگی نهایی، حاصل تبدیل مدل گاوسی نهایی به فضای مماسی اقلیدسی می‌باشد. با توجه به رابطه ۲-۲۶، تخمین مدل گاوسی نهایی می‌تواند با استفاده از وزن‌دهی بردارهای فضای مماسی مربوط به زیرنواحی محاسبه گردد. به منظور اعمال ماسک وزنی به این توصیفگر، وزن‌های مربوط به زیرنواحی متناظر با ماسک، در رابطه ۲-۲۱ تغییر داده می‌شود. همچنین برای اعمال روش نمونه‌افزایی به این توصیفگر مشابه آنچه در روش قبل گفته شد عمل خواهیم نمود.

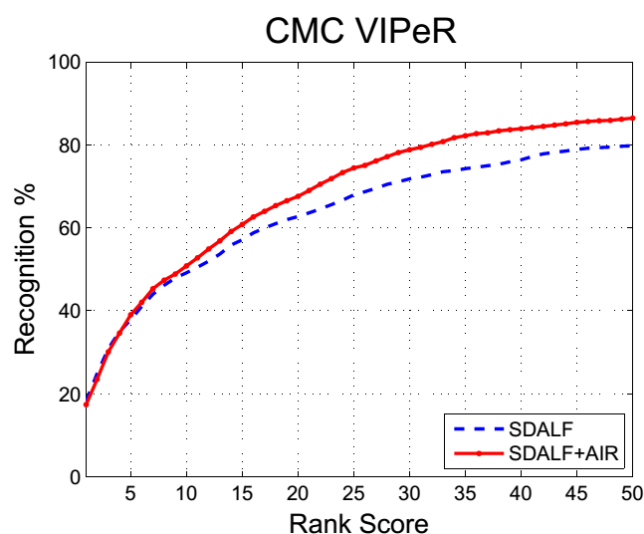
۳-۶- جمع‌بندی

در این فصل به تشریح رویکرد پیشنهادی این رساله پرداختیم. نشان دادیم که اطلاعات زاویه‌ای به عنوان یک ویژگی سه‌بعدی مهم و تاثیرگذار بر فرآیند بازشناسی، باید مورد توجه قرار گیرد. در رویکرد پیشنهادی برای مقایسه و تطبیق دو تصویر، ابتدا زاویه قرارگیری شخص در تصویر مشخص می‌شود. بدین منظور دو راهکار برای تخمین زاویه ارائه نمودیم. در مرحله بعد با استفاده از اطلاعات زاویه، فرآیند یکسان‌سازی بین دو تصویر مذکور انجام شد. منظور از فرآیند یکسان‌سازی، اعمال تغییراتی در تصاویر یا بردار ویژگی مربوط به آنهاست به طوری که تاثیرات ناشی از چرخش زاویه‌ای به حداقل کاهش یابد. دو روش برای یکسان‌سازی ارائه شد که شامل هیستوگرام وزن‌دار و نمونه‌افزایی تصویر است. در نهایت رویکرد پیشنهادی در دو الگوریتم مهم در کاربرد بازشناسی، مورد استفاده قرار گرفت. نحوه استفاده از رویکرد پیشنهادی در این دو الگوریتم نیز تشریح گردید. بدین ترتیب مشاهده می‌کنیم که رویکرد پیشنهادی در بسیاری از الگوریتم‌های بازشناسی قابل استفاده می‌باشد

فصل ۴. ارزیابی و نتایج آزمایش‌ها

در فصل قبل به تشریح روش پیشنهادی پرداختیم. ارزیابی نتایج و مقایسات برای بخش تشخیص زاویه و موقعیت یابی ناحیه سر از الگوریتم پیشنهادی، در زیربخش‌های مربوطه ارائه شد. در این فصل به ارائه نتایج و مقایسات انجام شده و ارزیابی آن برای کل رویکرد پیشنهادی و استفاده از آن در فرآیند بازشناسی، خواهیم پرداخت. به منظور ارزیابی رویکرد پیشنهادی مجموعه داده ViPER مورد استفاده قرار گرفته است. این مجموعه داده‌ای شامل تصویر ۶۳۲ فرد می‌باشد. برای هر فرد دو تصویر مربوط به دو دوربین مختلف تهیه شده است. دو تصویر مربوط به یک فرد از زوایای مختلف می‌باشند. شرایط نوری در دو دوربین متنوع می‌باشد. تصاویر هر فرد به مستطیلی محدود گشته و در اندازه ۱۲۸ در ۴۸ پیکسل می‌باشد. از آنجایی که رویکرد پیشنهادی در این رساله تمرکز ویژه‌ای بر روی تغییرات زاویه‌ای شدید دارد از بین مجموعه داده‌های استاندارد موجود، مجموعه ViPER انتخاب شده که دارای چنین ویژگی است. به طور کلی این مجموعه داده سخت‌ترین مجموعه داده‌ای است که در بازشناسی مورد استفاده قرار گرفته است.

معیارهای ارزیابی بازشناسی را می‌توان در دو حالت مجموعه بسته و مجموعه باز بررسی نمود. در حالت مجموعه بسته که مجموعه موارد آزمون در مجموعه گالری وجود دارد، مسئله بازشناسی در قالب یک مسئله رتبه‌بندی در نظر گرفته می‌شود. در این حالت به منظور نشان دادن دقت تشخیص از نمودار CMC استفاده می‌شود. در این نمودار روی محور افقی رتبه‌های مختلف را رسم نموده است. برای مثال نمودار شکل ۴-۱ نشان می‌دهد زمانی که الگوریتم SDALF نمونه‌های گالری را برای تمام حالات مورد آزمون مرتب می‌کند به طور میانگین در ۲۰ درصد موارد نمونه درست در رتبه ۱ قرار می‌گیرد. یا برای مثال نشان می‌دهد الگوریتم SDALF در ۸۰ درصد موارد نمونه درست در رتبه‌ای زیر ۵۰ قرار می‌گیرد.



شکل ۴-۱: یک نمونه از نمودار CMC

زمانی که مسئله به صورت مجموعه باز است نمودار CMC نمی‌تواند به درستی تمام ویژگی‌های الگوریتم را نشان دهد. در حالت مجموعه باز مواردی خواهیم داشت که نمونه مورد آزمون در مجموعه گالری نیست. در این حالت ارزیابی باید بر اساس دو معیار دقت (رتبه ۱) و نرخ پذیرش خطا محاسبه (FAR^۱) گردد. منظور از دقت همان موارد تشخیص TP^۲ می‌باشد که نشان‌دهنده درصد نمونه‌هایی از مجموعه آزمون است که به درستی تشخیص داده شده است. رابطه نرخ پذیرش خطا به صورت زیر می‌باشد:

$$FAR = \frac{MMs + FPs}{N_p} \quad (۱-۴)$$

در این رابطه MMs تعداد مواردی را نشان می‌دهد که مورد آزمون در مجموعه گالری بوده ولی نادرست تشخیص داده می‌شود، و FPs تعداد مواردی را نشان می‌دهد که مورد آزمون در مجموعه گالری نیست اما الگوریتم آن را برچسب‌گذاری کرده است. به منظور مشخص کردن رفتار سیستم نمودار Accuracy-FAR رسم می‌گردد. برای مثال می‌توان مقدار آستانه‌ای را که سیستم تصمیم به رد کردن

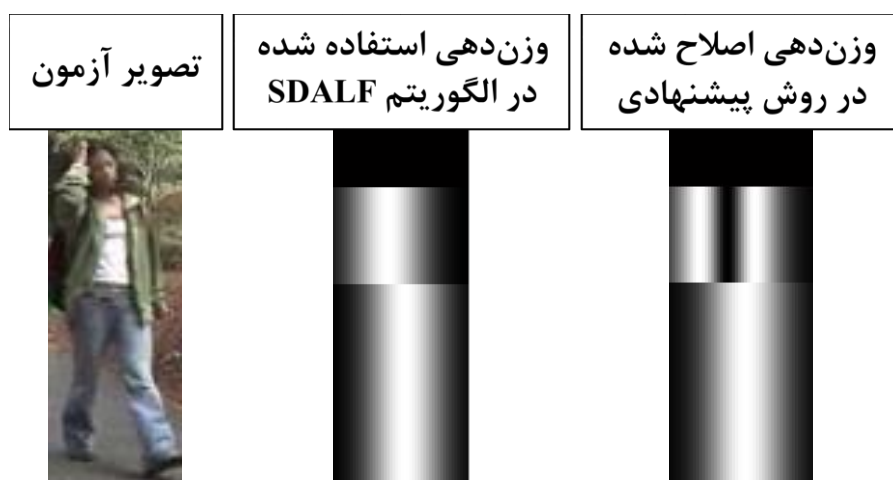
^۱ False Acceptance Rate

^۲ True Positive

یک نمونه مورد آزمون، می‌کند را از ۰ تا ۱ با گام‌های ۰,۰۵ مقدار دقت و نرخ پذیرش خطا را محاسبه نمود و کلیه این اطلاعات را بر روی نمودار Accuracy-FAR مشاهده کرد.

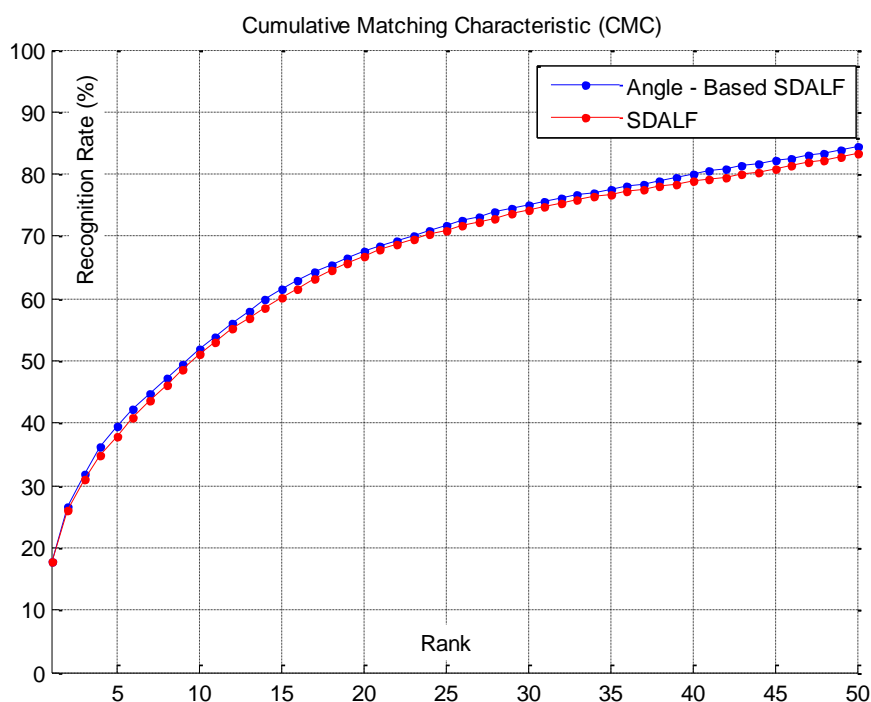
۴-۱- ارزیابی روش ناآگاهانه با دانستن اطلاعات زاویه‌ای

اولین ایده از رویکرد پیشنهادی با استفاده از یکسان‌سازی ناآگاهانه انجام شد. منظور از یکسان‌سازی ناآگاهانه، ماسک کردن ناحیه جلویی از بالاتنه فرد بدون در نظر گرفتن نوع پوشش آن است. به عبارت دیگر در موارد بسیاری نوع پوشش فرد به گونه‌ای است که نیازی به اعمال ماسک‌گذاری در تصویر وجود ندارد. در واقع این روش نسخه اولیه از رویکرد پیشنهادی در این رساله می‌باشد. در این روش، دقت بازشناسی الگوریتم ارائه شده در [۱۵] با استفاده از اطلاعات زاویه و با استفاده از وزن‌دهی ناآگاهانه بهبود یافته است. بدین منظور فرض شده که اطلاعات زاویه‌ای بصورت دقیق در دسترس است. این فرض خصوصاً زمانی که تصاویر از دنباله ویدئویی استخراج شود، فرض کاملاً منطقی می‌باشد. در واقع زمانی که دنباله‌ای از تصاویر ویدئویی در دسترس است، ردیابی شخص در تصویر امکان‌پذیر می‌باشد. اطلاعات حاصل از الگوریتم ردیابی، دقت تشخیص زاویه را به مقدار قابل توجهی افزایش خواهد داد. آزمایش‌ها بر روی داده‌های ViPER صورت گرفته است. اعمال وزن‌دهی بر روی بخش‌های مختلف الگوریتم SDALF نیز با استفاده از رابطه ۳-۱۷ انجام شده است. شکل ۴-۲ نمونه‌ای از وزن‌دهی برای یک تصویر را نشان می‌دهد.



شکل ۲-۴: نمونه‌ای از مرحله یکسان‌سازی غیرآگاهانه برای تصویری از روبرو، بر اساس وزن دهی پیشنهادی

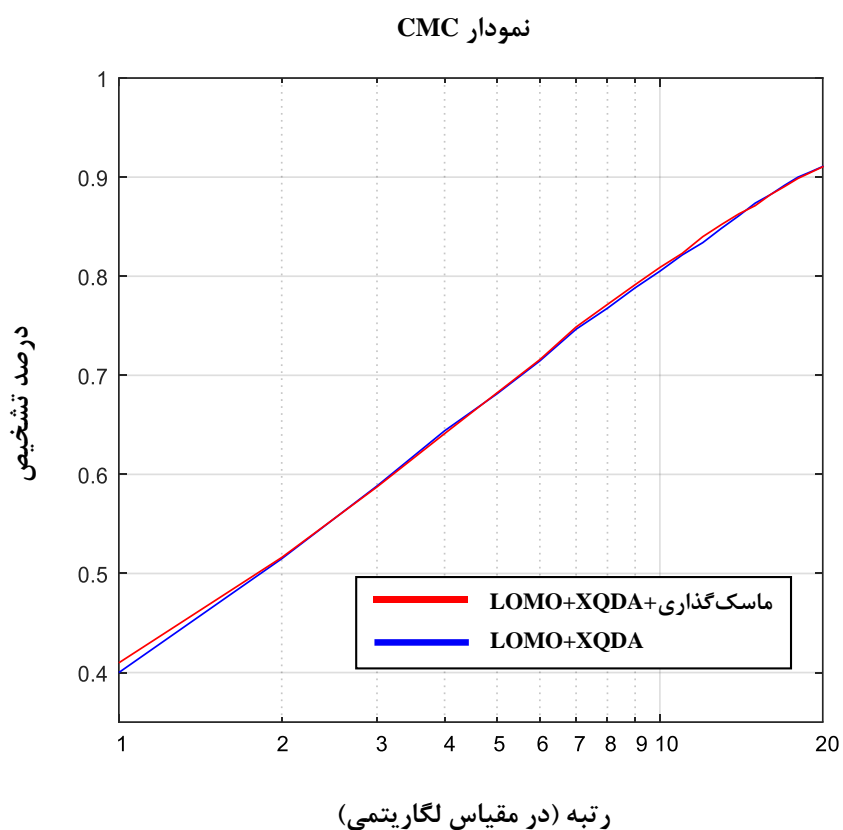
آزمایش‌های صورت گرفته در حالت مجموعه بسته می‌باشد. نتایج مقایسات انجام شده برای نسخه استاندارد الگوریتم SDALF و نسخه اصلاح شده در شکل ۳-۴ بر روی نمودار CMC نشان داده شده است. همانطور که شاهد هستیم نرخ بازشناسی بطور میانگین برای رتبه‌های مختلف به میزان ۱,۵ درصد بهبود یافته است.



شکل ۳-۴: مقایسه نسخه استاندارد SDALF و نسخه اصلاح شده به روش پیشنهادی

۴-۲- ارزیابی روش آگاهانه به همراه تخمین خودکار زاویه

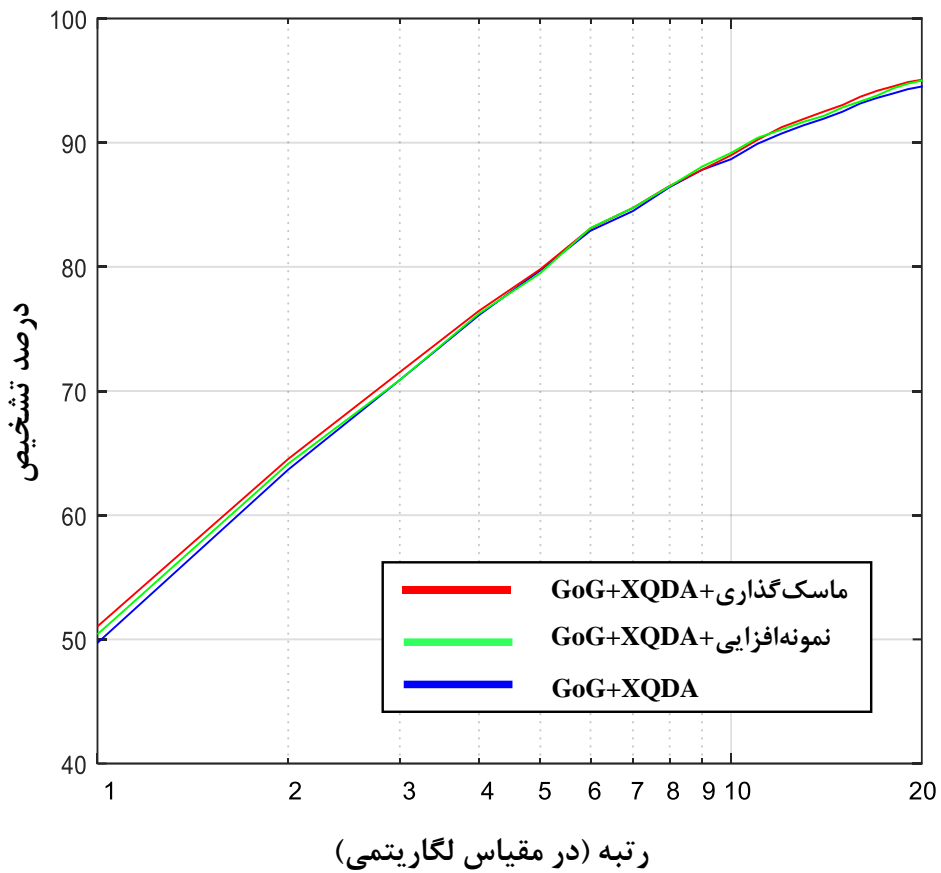
آزمایش‌های این بخش بر روی رویکرد نهایی این رساله صورت گرفته است. در این آزمایش‌ها مرحله یکسان‌سازی بصورت آگاهانه انجام می‌شود. منظور از آگاهانه، ماسک‌گذاری متناسب با نوع پوشش و ظاهر فرد می‌باشد. برای مثال در مواردی که پوشش فرد به لحاظ رنگی یکنواخت باشد، در حالت آگاهانه ماسکی استخراج نمی‌گردد. همچنین تخمین زاویه نیز به صورت خودکار با راهکارهای پیشنهادی در این رساله، تعیین گردیده است. همانطور که در فصل سوم بدان اشاره شد رویکرد پیشنهادی قابل ترکیب با هر الگوریتم بازشناسی می‌باشد. به عبارتی در الگوریتم پیشنهادی با در دست داشتن اطلاعات زاویه‌ای، برحسب شرایط زاویه، از مرحله یکسان‌سازی استفاده می‌شود. در این بخش به منظور تلفیق رویکرد پیشنهادی، دو الگوریتم بازشناسی مورد بررسی و پیاده‌سازی قرار گرفته است [۲۵]، [۲۸]. جزئیات و نحوه این تلفیق در زیربخش ۳-۵- تشریح گردید. مقایسات در قالب نمودار CMC و جدولی از مقایسه روش‌های مختلف براساس رتبه‌های مختلف می‌باشد. در [۲۸] ویژگی جدیدی با نام LOMO معرفی شده و همچنین روشی برای یادگیری معیار فاصله با نام XQDA ارائه شده است. در [۲۵] ویژگی جدیدی با نام GoG معرفی شده و برای یادگیری معیار فاصله نیز از روش مقاله قبلی (XQDA) استفاده می‌کند. به منظور انجام فرآیند آموزش معیار فاصله، مجموعه داده‌های آموزشی بصورت تصادفی به دو قسمت تقسیم می‌شود. بدین ترتیب ۳۱۶ جفت تصویر برای آموزش و ۳۱۶ جفت تصویر برای مرحله آزمون استفاده می‌شود. این تقسیم بندی به دفعات تکرار می‌گردد و نتیجه نهایی، شامل میانگین گیری از نتایج بدست آمده برای آزمایش‌های انجام شده می‌باشد. آزمایش‌های صورت گرفته در این بخش شامل دو حالت ماسک‌گذاری وزنی و نمونه‌افزایی می‌باشد. شکل ۴-۴ نمودار CMC تا رتبه ۲۰ را برای الگوریتم LOMO+XQDA و ترکیب آن با رویکرد پیشنهادی برای دو حالت یکسان‌سازی (ماسک‌گذاری وزنی و نمونه‌افزایی) نشان می‌دهد.



شکل ۴-۴: مقایسه الگوریتم LOMO+XQDA در دو حالت (تلفیق رویکرد پیشنهادی-ماسک گذاری و بدون تلفیق) در قالب نمودار CMC

شکل ۴-۵ نمودار CMC تا رتبه ۲۰ را برای الگوریتم GoG+XQDA و ترکیب آن با رویکرد پیشنهادی برای دو حالت یکسان سازی (ماسک گذاری وزنی و نمونه‌افزایی) نشان می‌دهد.

نمودار CMC



شکل ۴-۵: مقایسه الگوریتم GoG+XQDA در سه حالت (تلفیق با رویکرد پیشنهادی-ماسک گذاری و نمونه افزایی و بدون تلفیق) در قالب نمودار CMC

جدول ۴-۱ نتایج مقایسه‌های انجام شده برای کلیه حالات را به صورت عددی برای رتبه‌های ۱، ۱۰ و ۲۰ نشان می‌دهد. در این جدول روش پیشنهادی با جدیدترین الگوریتم‌های بازشناسی که از جمله بهترین دقت‌ها را ارائه کرده‌اند مقایسه شده است.

جدول ۱-۴: نتایج بازشناسی برای تلفیق رویکرد پیشنهادی با الگوریتم‌های XQDA+LOMO و GoG+XQDA و مقایسه با روش‌های دیگر برای سه رتبه ۱، ۱۰ و ۲۰

رتبه ۲۰	رتبه ۱۰	رتبه ۱	روش
۹۵,۰۶	۸۸,۹۹	۵۱,۰۴	GoG+XQDA + ماسک‌گذاری
۹۵,۰۰	۹۰,۳۸	۵۰,۴۱	GoG+XQDA + نمونه‌افزایی
۹۴,۵۲	۸۸,۶۷	۴۹,۷۱	GoG+XQDA [۲۵]
۹۱,۰۴	۸۰,۸۹	۴۰,۹۸	LOMO+ XQDA + ماسک‌گذاری
۹۱,۰۴	۸۱,۳۰	۴۰,۰۹	LOMO+ XQDA + نمونه‌افزایی
۹۱,۰۸	۸۰,۵۱	۴۰,۰۰	LOMO+ XQDA [۲۸]
۹۰,۴۰	۸۱,۲۰	۳۷,۸۰	SCNCD [۳۱]
۸۲,۴۵	۷۰,۷۱	۳۱,۱۱	kBiCov [۱۸]
۷۹,۸۷	۶۵,۹۵	۲۹,۱۱	Mid-level Filter [۴۵]

این نتایج نشان می‌دهد که با استفاده از اطلاعات زاویه‌ای و گام یکسان‌سازی در مرحله تطبیق از الگوریتم‌های بازشناسی، بهبود در دقت تشخیص را خواهیم داشت. استفاده از رویکرد پیشنهادی در الگوریتم LOMO+XQDA منجر به بهبود ۱ درصد نرخ تشخیص برای ۳۱۶ شخص یا کلاس در رتبه ۱ شده است. همچنین برای الگوریتم GoG+XQDA نیز استفاده از رویکرد پیشنهادی منجر به بهبود ۱,۳ درصد در نرخ تشخیص در رتبه ۱ شده است. نتایج GoG+XQDA روی داده‌های ViPER، جزو بهترین نتایجی است که تاکنون گزارش شده است.

۴-۳- بازشناسی چندتصویره (دنباله ویدئویی)

در حالت چندتصویره و یا به طور کلی زمانی که دنباله‌ای از تصاویر برای هر شخص در اختیار باشد (دنباله ویدئویی)، اثربخشی رویکرد پیشنهادی بیشتر می‌شود. به عبارت دیگر تشخیص زاویه فرد در یک دنباله ویدئویی می‌تواند با دقت بیشتری صورت گیرد. تخمین دقیق‌تر زاویه نیز منجر به بهبود عملکرد رویکرد پیشنهادی در این رساله می‌شود. از آنجایی که مسیر حرکت شخص در یک دنباله ویدئویی

محدود (برای مثال ۱۰ فریم متوالی)، بصورت ناگهانی تغییر نمی‌کند، بنابراین زاویه قرارگیری شخص نیز برای تصاویر دنباله مذکور، مقدار مشابهی می‌باشد. بدین ترتیب می‌توانیم الگوریتم تشخیص زاویه را برای تمامی تصاویر دنباله، اعمال و نتایج بدست‌آمده را با یکدیگر ترکیب نماییم و در نتیجه تخمین دقیق‌تری از زاویه شخص حاصل گردد. از آنجایی که دقت و اثربخشی رویکرد پیشنهادی در این رساله وابسته به کارآمدی تخمین زاویه است، بنابراین استفاده از چندین تصویر یا دنباله‌ای از تصاویر می‌تواند منجر به بهبود الگوریتم پیشنهادی شود.

به منظور نشان دادن این موضوع، مجموعه تصاویر CUHK01 استفاده شده است [۸۰]. این مجموعه تصاویر مربوط به ۹۷۱ شخص در دو دوربین مجزا می‌باشد. برای هر دوربین، دو تصویر تقریباً متوالی از هر شخص وجود دارد. این مجموعه تصاویر از دنباله ویدئویی مربوط به یک فضای باز بصورت دستی جدا شده است. دو آزمایش به منظور تخمین زاویه انجام شده است. در آزمایش اول دقت تخمین زاویه بصورت دقیق و حد خطای $\pm 45^\circ$ برای تک تک تصاویر مجموعه داده با استفاده از روش پیشنهادی محاسبه شده است. در آزمایش دوم بصورت مشابه، دقت تخمین زاویه با استفاده از ترکیب نتایج تخمین از تصاویر متوالی محاسبه گردید. نحوه ترکیب نیز با استفاده از رابطه ۳-۱۰ می‌باشد. جدول ۴-۲ مقایسه دو حالت را نشان می‌دهد. همانطور که مشاهده می‌شود دقت تخمین زاویه برای زمانی که از دو تصویر استفاده گردد تقریباً شش درصد بیش از حالت تک تصویر است. بدیهی است که با افزودن بر تعداد تصاویر، دقت تخمین بیشتر می‌گردد.

جدول ۴-۲: مقایسه دقت تخمین زاویه برای حالت تک تصویر و دو تصویر

دقت تشخیص		روش
($\pm 45^\circ$ نادقیق)	(0° دقیق)	
۸۰,۷%	۵۳,۱%	تخمین زاویه (تک تصویر)
۸۶,۰%	۵۹,۱%	تخمین زاویه (دو تصویر)

۴-۴- جمع‌بندی

در این فصل رویکرد پیشنهادی در دو حالت ناآگاهانه و آگاهانه مورد ارزیابی قرار گرفت. در حالت ناآگاهانه در بخش یکسان‌سازی از رویکرد پیشنهادی، قسمت میانی پوشش فرد بدون توجه به نوع پوشش، ماسک‌گذاری می‌گردد. قسمت میانی پوشش فرد نیز با استفاده از محور قرینگی فراهم شده در الگوریتم SDALF قابل دستیابی است. اعمال این یکسان‌سازی منجر به بهبود در دقت بازشناسی برای رتبه‌های مختلف شد. در حالت آگاهانه، با استفاده از الگوریتم پیشنهادی در بخش ۳-۴ قسمت‌هایی از پوشش فرد که تحت چرخش‌های زاویه‌ای ممکن است قابل رؤیت نباشند استخراج می‌شوند. در فرآیند یکسان‌سازی دو رویکرد ارائه شد. در رویکرد اول ناحیه بدست آمده ماسک‌گذاری شده و در مرحله محاسبه بردار ویژگی، نواحی مذکور نادیده گرفته می‌شود. برای مثال در صورتی که ویژگی محاسبه شده برپایه هیستوگرام باشد، استفاده از هیستوگرام وزن دار می‌تواند فرآیند ماسک‌گذاری را فراهم نماید. در رویکرد دوم نواحی استخراج شده بر اساس پیکسل‌های مجاور در هر سطر نمونه‌افزایی می‌گردد. بدین ترتیب تصویر کاملی که فاقد نواحی مذکور است ایجاد می‌شود. این دو رویکرد در الگوریتم‌های استخراج ویژگی LOMO و GoG اعمال گردید و نتایج آزمایش‌ها بهبود در عملکرد بازشناسی را برای هر دو رویکرد نشان می‌دهند.

با دقت در نتایج درمی‌یابیم که روش ماسک‌گذاری نسبت به نمونه‌افزایی عملکرد بهتری دارد. دلیل این امر در تولید نواحی رنگی جدید در تصویر برای رویکرد نمونه‌افزایی است. در واقع با نمونه‌افزایی، وزن بیشتری به نواحی مجاور ماسک خواهیم داد که این امر می‌تواند در برخی موارد تاثیرات نامطلوبی در فرآیند استخراج ویژگی ایجاد نماید. همچنین نمونه‌افزایی می‌تواند در برخی موارد اعوجاج‌های جزئی در تصویر ایجاد کند که از معایب این روش محسوب می‌گردد. از مزایای رویکرد نمونه‌افزایی سهولت استفاده در الگوریتم‌های مختلف بازشناسی می‌باشد. در حالی که در روش ماسک‌گذاری، فرآیند استخراج ویژگی دست‌خوش تغییرات زیادی شده که استفاده از آن را دشوار می‌سازد. برای مثال در الگوریتم‌هایی

که از شبکه عصبی کانولوشن به منظور استخراج ویژگی یا یادگیری معیار فاصله استفاده می‌کنند، اعمال رویکرد ماسک‌گذاری بسیار دشوار است. در حالیکه به کمک نمونه‌افزایی، تصویر کاملی خواهیم داشت که به سادگی قابل استفاده است.

فصل ۵. جمع‌بندی و کارهای آینده

در این فصل به جمع‌بندی کارهای انجام شده خواهیم پرداخت. همچنین پیشنهاداتی برای انجام پژوهش‌های آتی ارائه خواهیم کرد.

۵-۱- جمع‌بندی رویکرد پیشنهادی در رساله

هدف از این رساله ارائه رویکردی مبتنی بر نحوه قضاوت ناظر انسانی در مسئله بازشناسی است. بر این اساس نشان دادیم که مقایسه دو تصویر در مرحله تطبیق از فرآیند بازشناسی، نیازمند مرحله یکسان‌سازی است. در واقع فرآیند یکسان‌سازی سعی دارد برخی تغییرات در تصویر را که به دلیل خصوصیات سه‌بعدی انسان در شرایط طبیعی مسئله ایجاد می‌گردند از تصاویر حذف و سپس عمل مقایسه را انجام دهد. خصوصیتی که در این رساله مورد هدف قرار گرفت، اطلاعات زاویه قرارگیری شخص در تصویر است. به بیان ساده، ایده اصلی رویکرد پیشنهادی بر این قاعده استوار است که نحوه مقایسه و سنجش شباهت تصویر دو شخص، وابسته به نحوه قرارگیری و به‌طور خاص، زاویه قرارگیری آن‌ها نسبت به دوربین می‌باشد.

یکی از بخش‌های اصلی در فرآیند یکسان‌سازی، تخمین زاویه بصورت خودکار است. بدین منظور دو راهکار پیشنهاد گردید. در راهکار اول با استفاده از کانتورهای مصنوعی تولیدشده از یک مدل سه‌بعدی انسان، و تطبیق اطلاعات لبه تصویر ورودی، بر کانتورهای مذکور، حالت شخص و از جمله زاویه قرارگیری آن استخراج شد. این عمل تطبیق در یک ساختار درختی و از طریق پیمایش درخت صورت گرفت. نشان دادیم که گره‌های میانی در این درخت، نوع تعمیم‌یافته‌ای از کانتور می‌باشد. در راهکار دوم با استفاده از استخراج ویژگی هیستوگرام گرادیان‌های جهتی در چهار ناحیه از تصویر و آموزش کلاسه‌بند رگرسیون لجستیک برای هر یک، و در نهایت ترکیب خروجی این چهار کلاسه‌بند، تخمین خوبی از زاویه حاصل گردید. همچنین الگوریتم سریعی برای موقعیت‌یابی یکی از نواحی مذکور ارائه شد. در این الگوریتم نیز از ویژگی هیستوگرام گرادیان‌های جهتی استفاده شده است. نتایج

آزمایش‌ها برتری روش پیشنهادی را از منظر دقت تشخیص و پیچیدگی محاسباتی نسبت روش‌های موجود نشان می‌دهند. در بخشی از آزمایش‌ها، نشان دادیم که با در اختیار داشتن چند تصویر مرتبط با یک شخص در یک دوربین (مشابه دنباله ویدئویی)، می‌توان دقت تخمین زاویه را بصورت قابل توجهی افزایش داد و در نتیجه اثرپذیری الگوریتم پیشنهادی نهایی در این رساله را بهبود بخشید.

بعد از اینکه زاویه قرارگیری شخص در تصاویر مجموعه گالری و آزمون تعیین شد، تحت شرایط خاصی فرآیند یکسان‌سازی برای یکی از تصاویر اعمال می‌گردد. این یکسان‌سازی شامل استخراج نواحی خاصی از تصویر است که ممکن است تحت یک زاویه خاص قابل رؤیت نباشد. به منظور استخراج این ناحیه، ابتدا موقعیت سر در تصویر، توسط شبکه عصبی کانولوشن تعیین می‌شود. بدین منظور، الگوریتمی پیشنهاد گردید که در آن با ترکیب نتایج حاصل از دو شبکه عصبی کانولوشن برای ناحیه سر و ناحیه سر و شانه، موقعیت‌یابی دقیقی از ناحیه سر صورت می‌گیرد. همچنین پیچیدگی محاسباتی الگوریتم مذکور نیز نسبت به استفاده مرسوم کمتر می‌باشد.

در نهایت با استفاده از اطلاعات زاویه‌ای و موقعیت ناحیه سر در تصویر، نواحی خاصی از پوشش بالاتنه شخص استخراج گردید. بخش نهایی فرآیند یکسان‌سازی شامل حذف تاثیر نواحی مذکور بر بردار ویژگی تولید شده از تصویر مربوطه می‌باشد. بدین منظور دو راهکار پیشنهاد شد. در راهکار اول ماسکی بر روی تصویر قرار داده می‌شود و متناسب با نوع ویژگی از ماسک مذکور استفاده می‌گردد. برای مثال در روش‌هایی که از هیستوگرام در ساخت بردار ویژگی خود استفاده می‌کنند، هیستوگرام وزن‌دار می‌تواند به عنوان عمل ماسک‌گذاری در نظر گرفته شود. در راهکار دوم با استفاده از نمونه‌افزایی پیکسل‌های مجاور ناحیه ماسک، تصویر جدیدی هم اندازه با تصویر اولیه ساخته می‌شود که فاقد نواحی ماسک می‌باشد. مزیت راهکار نمونه‌افزایی، سهولت استفاده در انواع روش‌های بازشناسی است. آزمایش‌های صورت گرفته برای دو راهکار ماسک‌گذاری و نمونه‌افزایی بر روی دو الگوریتم جدید و کارآمد

در کاربرد بازشناسی می‌باشد. نتایج بدست‌آمده بهبود دقت در بازشناسی را برای هر دو الگوریتم و هر دو راهکار نشان می‌دهند.

بدین ترتیب دیدیم که تطبیق در فرآیند بازشناسی باید بصورت آگاهانه با در نظر گرفتن ویژگی‌ها و خصیصه‌هایی نظیر خصوصیات سه‌بعدی بدن شخص در تصویر صورت گیرد. استخراج اندام بدن شخص در تصویر و تطبیق بر اساس این اطلاعات می‌تواند راهکاری برای مسئله فوق باشد. اما مشکل این راهکار، نیازمندی به تصاویر با کیفیت بالا و همچنین زمان بر بودن فرآیند تشخیص اندام‌ها در تصویر می‌باشد. بدین ترتیب ما به دنبال ویژگی‌هایی هستیم که علاوه بر تاثیرگذاری بر فرآیند بازشناسی، با کارآمدی مطلوبی قابل استخراج باشند. اطلاعات زاویه‌ای یکی از این ویژگی‌ها است که نشان دادیم با کارآمدی مطلوبی قابل تخمین می‌باشد. نتایج بدست‌آمده اثربخشی و بهبود را در کاربرد بازشناسی نشان می‌دهند. شایان ذکر است که کلیه اطلاعات استخراج شده از تصویر شامل اطلاعات زاویه‌ای و ناحیه سر می‌تواند در بخش‌های دیگری از یک سیستم نظارت ویدئویی قابل استفاده باشد.

۵-۲- کارهای آینده

در این پژوهش با استفاده از اطلاعات زاویه‌ای و استخراج نواحی خاصی در بخش بالاتنه، عمل یکسان‌سازی انجام گردید. در راستای اهداف این پژوهش می‌توان توسعه‌های بیشتری ارائه نمود که در زیر به برخی از آن‌ها اشاره خواهیم کرد.

۱- از آنجایی که پوشش اغلب افراد از قواعدی کلی و تقریباً مشابهی پیروی می‌کند می‌توان این فرض‌ها و قواعد را در تصاویر مشخص و مدل‌سازی نمود. برای مثال می‌توان وجود و یا عدم وجود کوله‌پشتی را در تصویر مشخص نمود. بعد از استخراج این خصیصه و تعیین خصوصیات سه‌بعدی مربوط به آن، می‌توان پیش‌بینی از ظاهر این شخص در زوایای مختلف به عمل آورد. برای مثال فردی که کوله‌پشتی حمل می‌کند، احتمالاً در تصویر زاویه روبرو کوله‌پشتی قابل

رؤیت نخواهد بود. این خصیصه‌ها در کنار ویژگی‌های سه‌بعدی مرتبط با آن‌ها می‌تواند تکمیلی بر پژوهش صورت گرفته در این رساله باشد.

۲- یکی از مشکلات در زمینه یادگیری انواع خصیصه‌ها از جمله آنچه در مورد قبل اشاره شد، کمبود داده‌های آموزشی است. در مسئله بازشناسی که تمرکز بر روی ظاهر فرد است می‌توان مدل‌های سه‌بعدی از انسان با پوشش‌های مختلف در یک نرم‌افزار سه‌بعدی طراحی نمود و سپس این مدل‌های سه‌بعدی را در حالات گوناگون راه رفتن، در محیطی مجازی و سه‌بعدی قرار داد. بدین ترتیب مجموعه عظیمی از داده‌های مصنوعی قابل تولید می‌باشد که می‌تواند در مرحله یادگیری مورد استفاده قرار گیرد. البته لازمه این امر طراحی تعداد زیادی پیکره سه‌بعدی با ویژگی‌های اندامی متنوع می‌باشد. به‌علاوه محیط‌های مجازی که این مدل‌ها در آن‌ها قرار می‌گیرند باید تا حد امکان با محیط‌های طبیعی شبیه باشند. وجود داده‌های آموزشی زیاد استفاده از مدل‌های یادگیرنده قدرتمند نظیر شبکه‌های عصبی کانولوشن عمیق را ممکن می‌سازد.

۳- فرآیند یکسان‌سازی معرفی شده در این رساله می‌تواند بصورت احتمالاتی و فازی صورت گیرد. بدین ترتیب که بجای استخراج دقیق نواحی، نگاشتی احتمالاتی از یک تصویر در زاویه مورد نظر ایجاد نماییم. منظور از نگاشت احتمالاتی، تصویری است که در هر پیکسل از آن بجای وجود یک مقدار قطعی رنگ، توزیع احتمالی از رنگ و بافتی است که می‌تواند برای آن پیکسل تشکیل گردد. این نوع نگرش به مسئله، با شرایط سیستم‌های واقعی که توأم با انواع خطاها می‌باشند سازگارتر است. بدین ترتیب در هنگام تطبیق بین دو فرد، ابتدا نگاشتی احتمالی از دو شخص به یک حالت مشترک و از پیش تعیین شده ایجاد می‌شود. سپس مدل‌های احتمالی مرتبط با دو شخص مذکور با هم مقایسه می‌گردد.

۴- با فرض در دست داشتن تصاویر بسیار زیادی از افراد در زوایای مختلف می‌توان به کمک مدل‌های یادگیری عمیق، نگاشت بین تصاویر در زوایای مختلف را بدست آورد. به‌طوری‌که

به صورت هم‌زمان، هم عمل موقعیت‌یابی و هم نگاهت مذکور حاصل گردد. در نتیجه با در دست داشتن تصویر یک شخص می‌توانیم تخمینی از تصویر شخص در زاویه‌ای دیگر داشته باشیم. به نظر می‌رسد بهترین راه در فراهم کردن داده‌های آموزشی، تولید مصنوعی آن‌هاست که در بند قبل تشریح گردید.

۵- یکی از جنبه‌های مهم در سیستم‌های بازشناسی قابلیت استفاده در شرایط محیطی با مقیاس بزرگ می‌باشد. هرچه محیط و افراد رؤیت شده در سیستم بیشتر باشند فرآیندهایی نظیر بازشناسی با دشواری‌های بیشتری همراه هستند. یکی از این مشکلات مسئله بازشناسی بلادرنگ است. زمانی که افراد کاندید در سیستم زیاد باشند زمان اجرای الگوریتم بیشتر شده و در نتیجه زمان پاسخ سیستم نیز بیشتر می‌گردد. در چنین مواقعی استفاده از راهکارهای موازی‌سازی می‌تواند بر این مشکلات غلبه نماید. رویکرد موازی‌سازی در کنار ساختارهای هرمی می‌تواند بیشترین تاثیر را بر سرعت اجرا داشته باشد. منظور از ساختارهای هرمی استخراج ویژگی از یک سطح کلی به سطح جزئی یا بالعکس می‌باشد که در این ساختار سعی می‌شود از نتیجه محاسبات در گذشته بیشترین استفاده به عمل آید. استفاده از چنین ساختاری مستلزم طراحی ویژگی‌های خاصی است که دارای چنین مشخصاتی باشند. در این راستا استفاده پردازنده‌های گرافیکی می‌تواند راهکار مناسبی باشد.

واژه نامه

A

Active Color Model..... مدل رنگی فعال
Annotation..... حاشیه نویسی
Average linkage..... پیوند میانگین

B

Bhattacharyya..... باتچریا
Boosting بوستینگ
Brightness Transfer Function ... تابع انتقال
رنگی

C

Calibration..... هم‌ترازی
Close set مجموعه بسته
Confidence اطمینان
Confusion matrix..... ماتریس درهم‌ریختگی
Contour..... کانتور
Contrast کنتراست
Convex programming.. برنامه نویسی محدب
Convolutional Neural Network شبکه
عصبی کانولوشن

D

Dense correspondence..... انطباق متراکم
Discriminative model مدل افتراقی
Distance Transform..... تبدیل فاصله
Dominant Color Descriptor . توصیفگر رنگ
غالب

E

Epitome جوهره

F

False Acceptance Rate نرخ پذیرش خطا
Feature map نقشه ویژگی
Feed-forward artificial neural
network شبکه عصبی مصنوعی پیش‌خور
Fidelity وفاداری
Fully connected layer..... لایه تماماً متصل

G

Gabor filter..... فیلتر گابور
Gallery set..... مجموعه گالری
Gaussian mixture model مدل مخلوط
گاوسی

H

Haar-like features..... ویژگی‌های شبه‌هار
Hessian..... هسین
Hierarchical Clustering..... خوشه‌بندی
سلسله‌مراتبی
Histogram of Oriented
Gradients..... هیستوگرام گرادیان‌های جهتی
Homography هموگرافی
Hyper contour ابرکانتور

K

Kernel function تابع هسته

L

Local Maximal Occurrence ماکزیمم وقوع
محلی
Loss layer..... لایه تابع اتلاف

M

Mahalanobis distance..... فاصله ماهالانوبیس

Manifold..... خمینه
 Max pooling layer لایه تجمیع بیشینه
 Maximum a posteriori.... بیشینه‌گر احتمال
 پسین
 Maximum Likelihood Estimation.... برآورد
 درست‌نمایی بیشینه
 Momentum..... مُمنتم
 Multilayer perceptron ... پرسپترون چندلایه
 Multiscale..... چندمقیاسی
 Multi-Shot چندتصویره

N
 Normalized cuts algorithm..... الگوریتم
 برش‌های نرمال

O
 Open set مجموعه باز
 Overfitting..... بیش‌برازش

P
 Panoramic model مدل تمام‌نما
 Parzen window..... پنجره پارزن
 Pooling layer لایه تجمیع
 Principle Component Analysis آنالیز
 مؤلفه‌های اصلی
 Probabilistic احتمالاتی
 Probe آزمون

S
 Schmid filter..... فیلتر اشمیت
 Segmentation..... قطعه‌بندی
 Shared-weights architecture معماری
 وزن‌دهی اشتراکی
 Sigmoid cross-entropy..... آن‌تروپی متقاطع
 سیگموئیدی
 Sigmoid function..... تابع سیگموئیدی
 Single-Shot تک‌تصویره
 Spectral clustering..... خوشه‌بندی طیفی
 Structure element..... مؤلفه ساختاری
 Support Vector Machine ماشین بردار
 پشتیبان

T
 Tangent space..... فضای مماسی
 Tied convolution layer لایه کانولوشن
 گره‌خورده
 True Positive مثبت صحیح

U
 Unimodal distribution توزیع تک‌مده
 Upsampling نمونه‌افزایی

W
 Watershed segmentation قطعه‌بندی حوزه
 آبگیر

- [1] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image and Vision Computing*, vol. 32, pp. 270-286, 2014.
- [2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 886-893.
- [3] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, pp. 1491-1498.
- [4] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*, 1999, pp. 246-252.
- [5] N. Jojic, A. Perina, M. Cristani, V. Murino, and B. Frey, "Stel component analysis: Modeling spatial correlations in image class structure," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 2044-2051.
- [6] N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, pp. 1528-1535.
- [7] J. Lu and E. Zhang, "Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion," *Pattern Recognition Letters*, vol. 28, pp. 2401-2411, 2007.
- [8] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 2007, pp. 1-7.
- [9] A. Ess, B. Leibe, and L. Van Gool, "Depth and appearance for mobile scene analysis," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1-8.
- [10] U.H. Office, i-Lids Multiple Camera Tracking Scenario Definition, 2007.
- [11] ([link].URL)<http://homepages.inf.ed.ac.uk/rbf/CAVIAR/2004>.
- [12] S. Wang, M. Lewandowski, J. Annesley, and J. Orwell, "Re-identification of pedestrians with variable occlusion and scale," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, 2011, pp. 1876-1882.
- [13] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Computing Surveys (CSUR)*, vol. 46, p. 29, 2013.
- [14] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," *Pattern Recognition Letters*, vol. 33, pp. 898-903, 2012.

- [15] L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Computer Vision and Image Understanding*, vol. 117, pp. 130-144, 2013.
- [16] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," *Computer Vision–ECCV 2006*, pp. 589-600, 2006.
- [17] J. Metzler, "Appearance-based re-identification of humans in low-resolution videos using means of covariance descriptors," in *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, 2012, pp. 191-196.
- [18] B. Ma, Y. Su, and F. Jurie, "Bicov: a novel image representation for person re-identification and face verification," in *British Machine Vision Conference*, 2012, p. 11 pages.
- [19] B. Ma, Y. Su, and F. Jurie, "Covariance descriptor based on bio-inspired features for person re-identification and face verification," *Image and Vision Computing*, vol. 32, pp. 379-390, 2014.
- [20] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2288-2295.
- [21] S. Bık, G. Charpiat, E. Corvee, F. Bremond, and M. Thonnat, "Learning to match appearances by correlations in a covariance metric space," in *European Conference on Computer Vision*, 2012, pp. 806-820.
- [22] Y. Zhang and S. Li, "Gabor-LBP based region covariance descriptor for person re-identification," in *Image and Graphics (ICIG), 2011 Sixth International Conference on*, 2011, pp. 368-371.
- [23] E. Corvee, F. Bremond, and M. Thonnat, "Person re-identification using spatial covariance regions of human body parts," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, 2010, pp. 435-440.
- [24] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Multiple-shot human re-identification by mean riemannian covariance grid," in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, 2011, pp. 179-184.
- [25] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical gaussian descriptor for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1363-1372.
- [26] A. Bedagkar-Gala and S. K. Shah, "Part-based spatio-temporal model for multi-person re-identification," *Pattern Recognition Letters*, vol. 33, pp. 1908-1915, 2012.
- [27] R. Satta, G. Fumera, and F. Roli, "Fast person re-identification based on dissimilarity representations," *Pattern Recognition Letters*, vol. 33, pp. 1838-1848, 2012.
- [28] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2197-2206.

- [29] A. D'Angelo and J.-L. Dugelay, "People re-identification in camera networks based on probabilistic color histograms," in *Visual Information Processing and Communication*, 2011, p. 78820K.
- [30] Z. J. Xiang, Q. Chen, and Y. Liu, "Person re-identification by fuzzy space color histogram," *Multimedia tools and applications*, vol. 73, pp. 91-107, 2014.
- [31] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li, "Salient color names for person re-identification," in *European conference on computer vision*, 2014, pp. 536-551.
- [32] S. T. Birchfield and S. Rangarajan, "Spatiograms versus histograms for region-based tracking," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 1158-1163.
- [33] A. Khan, J. Zhang, and Y. Wang, "Appearance-based re-identification of people in video," in *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*, 2010, pp. 357-362.
- [34] O. Javed, K. Shafique, and M. Shah, "Appearance modeling for tracking in multiple non-overlapping cameras," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 26-33.
- [35] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," *Computer Vision—ECCV 2008*, pp. 262-275, 2008.
- [36] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Scandinavian conference on Image analysis*, 2011, pp. 91-102.
- [37] S. Bık, E. Corvee, F. Bremond, and M. Thonnat, "Boosted human re-identification using riemannian manifolds," *Image and Vision Computing*, vol. 30, pp. 443-452, 2012.
- [38] E. Corvee, F. Bremond, and M. Thonnat, "Person re-identification using haar-based and dcd-based signature," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, 2010, pp. 1-8.
- [39] S. Wu, Y.-C. Chen, X. Li, A.-C. Wu, J.-J. You, and W.-S. Zheng, "An enhanced deep feature representation for person re-identification," in *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, 2016, pp. 1-8.
- [40] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch, "Learning implicit transfer for person re-identification," in *Computer Vision—ECCV 2012. Workshops and Demonstrations*, 2012, pp. 381-390.
- [41] C. Liu, S. Gong, C. Loy, and X. Lin, "Person re-identification: What features are important?," in *Computer Vision—ECCV 2012. Workshops and Demonstrations*, 2012, pp. 391-401.
- [42] R. Layne, T. M. Hospedales, and S. Gong, "Towards person identification and re-identification with attributes," in *European Conference on Computer Vision*, 2012, pp. 402-412.

- [43] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3586-3593.
- [44] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2528-2535.
- [45] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 144-151.
- [46] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, pp. 653-668, 2013.
- [47] C. C. Loy, C. Liu, and S. Gong, "Person re-identification by manifold ranking," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, 2013, pp. 3567-3571.
- [48] B. J. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking," in *BMVC*, 2010, p. 6.
- [49] W. Lin, Y. Shen, J. Yan, M. Xu, J. Wu, J. Wang, *et al.*, "Learning Correspondence Structures for Person Re-Identification," *IEEE Transactions on Image Processing*, vol. 26, pp. 2438-2453, 2017.
- [50] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3908-3916.
- [51] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views," *Computer Vision and Image Understanding*, vol. 109, pp. 146-162, 2008.
- [52] T. Gandhi and M. M. Trivedi, "Person tracking and reidentification: Introducing Panoramic Appearance Map (PAM) for feature representation," *Machine Vision and Applications*, vol. 18, pp. 207-220, 2007.
- [53] D. Baltieri, R. Vezzani, and R. Cucchiara, "Sarc3d: a new 3d body model for people tracking and re-identification," *Image Analysis and Processing-ICIAP 2011*, pp. 197-206, 2011.
- [54] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, pp. 1627-1645, 2010.
- [55] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, "Discriminative local binary patterns for human detection in personal album," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8.
- [56] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on riemannian manifolds," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, pp. 1713-1727, 2008.
- [57] T. Gandhi and M. M. Trivedi, "Image based estimation of pedestrian orientation for improving path prediction," in *Intelligent Vehicles Symposium, 2008 IEEE*, 2008, pp. 506-511.

- [58] H. Shimizu and T. Poggio, "Direction estimation of pedestrian from multiple still images," in *Intelligent Vehicles Symposium, 2004 IEEE*, 2004, pp. 596-600.
- [59] M. Enzweiler and D. M. Gavrila, "Integrated pedestrian classification and orientation estimation," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 982-989.
- [60] J. Tao and R. Klette, "Integrated pedestrian and direction classification using a random decision forest," in *Proceedings of the IEEE international conference on computer vision workshops*, 2013, pp. 230-237.
- [61] D. Tosato, M. Spera, M. Cristani, and V. Murino, "Characterizing humans on riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1972-1984, 2013.
- [62] D. M. Gavrila and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," *International journal of computer vision*, vol. 73, pp. 41-59, 2007.
- [63] D. Baltieri, R. Vezzani, and R. Cucchiara, "People orientation recognition by mixtures of wrapped distributions on random trees," *Computer Vision–ECCV 2012*, pp. 270-283, 2012.
- [64] BMLwalker. Analysis and Synthesis of Biological Motion Patterns, Retrieved March 15, 2015 from the world wide web:
<http://www.biomotionlab.ca/Demos/BMLwalker.html>.
- [65] R. Kimmel, N. Kiryati, and A. M. Bruckstein, "Sub-pixel distance maps and weighted distance transforms," *Journal of Mathematical Imaging and Vision*, vol. 6, pp. 223-233, 1996.
- [66] M.-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, 1994, pp. 566-568.
- [67] O. Maimon and L. Rokach, *Data mining and knowledge discovery handbook* vol. 2: Springer, 2005.
- [68] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3d pose estimation and tracking by detection," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 623-630.
- [69] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, p. 27, 2011.
- [70] P. Felzenszwalb and D. Huttenlocher, "Distance transforms of sampled functions," Cornell University 2004.
- [71] M. Sumner, E. Frank, and M. Hall, "Speeding up logistic model tree induction," in *PKDD*, 2005, pp. 675-683.
- [72] P. Komarek and A. W. Moore, "Making logistic regression a core data mining tool with tr-irls," in *Data Mining, Fifth IEEE International Conference on*, 2005, p. 4 pp.

- [73] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
- [74] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [75] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
- [76] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [77] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks*, vol. 8, pp. 98-113, 1997.
- [78] T.-H. Vu, A. Osokin, and I. Laptev, "Context-aware CNNs for person head detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2893-2901.
- [79] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikäinen, and S. Z. Li, "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 1301-1306.
- [80] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Asian Conference on Computer Vision*, 2012, pp. 31-44.

Abstract:

One of the main applications of intelligent video surveillance is in machine vision. The general configuration of these systems is a set of cameras with a non-overlapping field of view. The ultimate goal of these systems is to detect abnormal events and analyze individual behaviors. Human re-identification is one of the main steps in the system, which is to label people in front of a network of cameras, so that a single label is assigned to a particular person. In such systems, a person can be placed in front of a set of cameras at different times and positions.

For various reasons, recognition is one of the complex problems of computer vision. The reason for this complexity is the change in a person's appearance from one camera to another. These changes can be caused by hardware limitations such as low-end cameras, different camera responses to colors, changing optical conditions in different camera positions or angular variations in how people are exposed to cameras. One of the main weaknesses in the existing methods is the lack of attention towards angular information. In fact, when the quality of recorded images is low or people are far from the camera, existing algorithms inevitably ignore important information, such as angular information.

The purpose of this study is to model the judgment of the human supervisor regarding visual changes in the coverage of the target subject, resulting in a change in the angle of view. For this purpose, the orientation of the body towards the camera and the resulting 3D changes are used. In the proposed approach, the orientation of the person in the image is first extracted by two proposed methods, one based on contour information, and the other on the histogram of the oriented gradients and the logistic regression classifier. Then, certain areas that can be seen or hidden under angular rotations are identified. To this end, the region of the head in the image is used as a key point that is extracted using a convolutional neural network. In the matching stage of the re-identification process, the extracted areas are weighed, which we call it unification process in the proposed approach. In designing the proposed system, we have considered the computational complexity in all parts. The significant feature of this system is the ability to use and integrate in most re-identification methods. In this research, the proposed approach was integrated into two pioneering algorithms in the field of re-identification. The results of applying this method to the ViPER dataset, which is considered as one of the most difficult re-identification datasets, indicate an improvement of 1.3% in the recognition rate with respect to the two aforementioned algorithms and the effectiveness of the proposed unification process.

Keywords: people reidentification, video surveillance system, 3D model, body orientation estimation, convolutional neural network



Shahrood University of Technology
Faculty of Computer and IT Engineering

Ph.D. Dissertation in Artificial Intelligence

**People Re-identification using human 3D model attributes in
video surveillance systems**

By: Ali Sebti

Supervisor:
Prof. Hamid Hassanpour

Advisor:
Dr. Ali Soleimani

August 2017