

People Re-Identification in Video Surveillance Systems Using Angle Information

Ali Sebti^{1*}, Hamid Hassanpour²

^{1*}- Department of Computer Engineering, Faculty of Engineering, Golestan University, Gorgan, Iran.

²- Faculty of Computer Engineering and Information Technology, Shahrood University of Technology, Iran.

^{1*} a.sebti@gu.ac.ir, ² h.hassanpour@shahroodut.ac.ir

Corresponding author address: Ali Sebti, Gorgan Faculty of Technology and Engineering, Golestan University, Gorgan, Iran.

Abstract- Intelligent video surveillance is one of the main applications in machine vision. People re-identification as part of these systems is of particular importance. Indeed, the accuracy in this part improves the efficiency of many types of monitoring algorithms. The re-identification task in human mind is performed consciously and is based on a prior knowledge of the 3D attributes of the human body. One of these attributes is the orientation of the body relative to the camera. In other words, a human supervisor at the matching stage uses the angle information to estimate the appearance of the person at different angles. In this research, the above process is modeled. Thus, in this research, first the body orientation is automatically extracted in the image, and accordingly, upper part of the clothing is extracted, which might be hidden at different angles. Removing or re-sampling these areas reduces the destructive effects on the matching process. For evaluation and comparison, the proposed method was used in two of the efficient re-identification algorithms. Experiments were performed on the ViPer dataset and the results show % 1.3 percent improvements in the recognition rate for 316 people.

Keywords- People re-identification, Angle information, Video surveillance, 3D attributes.

بازشناسی انسان در سیستم‌های نظارت ویدئویی با بهره‌گیری از اطلاعات زاویه‌ای

علی سبلی^{۱*}، حمید حسن پور^۲

*۱- گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه گلستان، گرگان، ایران.

۲- دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شاهرود، شاهرود، ایران.

^{۱*} a.sebti@gu.ac.ir, ^۲ h.hassanpour@shahroodut.ac.ir

* نشانی نویسنده مسئول: علی سبلی، گرگان، دانشگاه گلستان، دانشکده فنی و مهندسی گرگان.

چکیده- نظارت ویدئویی هوشمند از کاربردهای اصلی در بینایی ماشین می‌باشد. بازشناسی انسان به عنوان بخشی از این سیستم‌ها از اهمیت ویژه‌ای برخوردار است. به طوری که صحت عملکرد در این بخش منجر به کارآمدی انواع الگوریتم‌های نظارتی می‌گردد. فرآیند بازشناسی در منظر ناظر انسانی به صورت آگاهانه و برپایه اطلاعات و دانش پیشین از ویژگی‌های سه بعدی پیکره انسان انجام می‌شود. یکی از این ویژگی‌ها اطلاعات زاویه قرارگیری شخص نسبت به دوربین است. به عبارتی ناظر انسانی در مرحله تطبیق با بهره‌گیری از اطلاعات زاویه، تخمینی از ظاهر شخص در زوایای مختلف خواهد داشت. در این پژوهش نیز فرآیند مذکور مدل سازی می‌گردد. در روش پیشنهادی ابتدا زاویه قرارگیری شخص در تصویر بصورت خودکار استخراج می‌شود و متناسب با آن و در صورت نیاز، بخش‌هایی از تصویر بالاتنه که در زوایای مختلف پنهان می‌شود، استخراج می‌گردد. حذف یا نمونه افزایشی این نواحی، تاثیرات مخرب آن بر فرآیند تطبیق را کاهش می‌دهد. به منظور ارزیابی و مقایسه، روش پیشنهادی بر روی دو الگوریتم کارآمد از الگوریتم‌های بازشناسی اعمال شده است. نتایج آزمایش‌های انجام شده بر روی مجموعه داده‌های ViPer، بهبود ۱.۳٪ درصد را در نرخ تشخیص بازشناسی برای ۳۱۶ شخص در تصویر نشان می‌دهد.

واژه‌های کلیدی: بازشناسی انسان، اطلاعات زاویه‌ای، نظارت ویدئویی، ویژگی‌های سه بعدی.

۱- مقدمه

سیستم نظارتی، با در دست داشتن یک تصویر یا دنباله‌ای از تصاویر فرد از یک دوربین خاص [۱]. به عبارتی یک سیستم نظارتی با ثبت تصویر یک فرد در یک دوربین، بررسی می‌کند که این فرد قبلاً در کدام یک از دوربین‌ها رؤیت شده است. در صورتی که این فرد قبلاً رؤیت شده باشد، سیستم برچسبی معادل برچسب فرد پیشین را در نظر می‌گیرد. در غیر این صورت شخص مذکور فردی تازه وارد محسوب می‌شود، و سیستم یک برچسب جدید برای آن در نظر می‌گیرد.

سیستم‌های نظارت ویدئویی امروزه در موقعیت‌های مختلفی از جمله اماکن عمومی مورد استفاده گسترده می‌باشند. هدف نهایی در این سیستم‌ها کشف و تشخیص رفتارهای غیر نرمال، نوع فعالیت، ردیابی افراد و تشخیص رویداد می‌باشد [۴]. بازشناسی افراد یکی از گام‌های اولیه و بخش‌های پیش‌نیاز در دستیابی به اهداف فوق است. لذا صحت عملکرد در این بخش، منجر به کارآمدی پردازش‌های بعدی می‌شود.

برای ردیابی افراد نیز بازشناسی می‌تواند مورد استفاده قرار گیرد. زمانی که یک فرد متحرک در دنباله‌ای از تصاویر، برای مدتی ناپدید می‌شود (برای مثال عبور فرد از پشت ستون یا فرد دیگر)، تشخیص

بر اساس تحقیقات صورت گرفته، بازشناسی افراد عبارت است از: "فرآیند شناسایی مجدد یک فرد در دوربین‌های مختلف یک

واقعی، کاملاً ممکن است. در رویکرد پیشنهادی، روشی کاملاً آگاهانه در مواجهه با شرایطی که تغییرات زاویه‌ای در تصاویر وجود دارد، معرفی شده است.

به بیانی دیگر زاویه قرارگیری شخص در مقابل دوربین باعث می‌شود که بخش‌هایی از پوشش فرد در تصاویر آشکار یا پنهان گردند و در نهایت ظاهری متفاوت از یک شخص در زوایای مختلف خواهیم داشت. در سیستم‌هایی که ناظر انسانی عمل نظارت را انجام می‌دهد 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 پژوهش ارائه راهکاری برای مدل‌سازی این الگوریتم تطبیق و یکسان‌سازی است. چیزی که پژوهش جاری را از الگوریتم‌های این حوزه متمایز می‌کند تمرکز بر ویژگی‌های سطح بالاتری از تصویر است. یکی از مهمترین این ویژگی‌ها اطلاعات زاویه است که با در نظر گرفتن آن فرآیند بازشناسی آگاهانه‌تر انجام می‌شود. همچنین در این پژوهش روشی جدید برای استخراج نواحی خاصی از پوشش فرد ارائه شده است که در عین سادگی و سرعت اجرا، از کارآمدی خوبی برخوردار است.

سیستم پیشنهادی در این پژوهش می‌تواند به‌عنوان پیش‌پردازشی در بسیاری از الگوریتم‌های بازشناسی مورد استفاده قرار گیرد. بدین ترتیب زمانی که دو تصویر از زوایای مختلف به سیستم وارد می‌شود ابتدا سیستم زاویه شخص در تصاویر را تخمین می‌زند. سپس با استفاده از اطلاعات زاویه‌ای و ناحیه قرارگیری سر در تصویر، نواحی خاصی که تحت چرخش زاویه‌ای می‌تواند الگوریتم‌های بازشناسی را با اخلال مواجه کند، استخراج می‌گردد. سپس با وزن‌دهی یا نمونه‌افزایی این نواحی، یکسان‌سازی بین دو تصویر انجام می‌شود. خروجی این سیستم می‌تواند با تغییرات اندکی متناسب با الگوریتم‌های دیگر بازشناسی، مورد استفاده قرار گیرد.

بخش‌های پیش رو در این پژوهش بصورت زیر سازماندهی شده است: بخش ۲ به مروری بر تحقیقات پیشین در بازشناسی انسان پرداخته و بخش ۳ به معرفی روش پیشنهادی می‌پردازد. همچنین ارزیابی روش پیشنهادی و مقایسه با الگوریتم‌های مرسوم در این حوزه در بخش ۴ صورت می‌گیرد. در نهایت بخش ۵ به نتیجه‌گیری از پژوهش جاری و پیشنهاداتی برای کارهای آتی در راستای این پژوهش می‌پردازد.

۲- پیشینه پژوهش

روش‌های موجود در بازشناسی افراد به دو دسته تقسیم‌بندی می‌شود. دسته اول روش‌هایی هستند که تنها از اطلاعات تصاویر

مجدد مسیر حرکت آن فرد مستلزم عمل بازشناسی است که می‌توان از روش‌های معرفی‌شده در این حوزه استفاده نمود. از عواملی که دقت فرآیند بازشناسی را می‌کاهد می‌توان به تعدد افراد، کیفیت پایین دوربین‌های نظارتی، شرایط نوری متفاوت صحنه‌ها و زاویه‌های مختلف افراد در دوربین‌های مختلف اشاره نمود [۵].

از فرضیاتی که در اکثر روش‌های ارائه شده برای بازشناسی در نظر گرفته می‌شود عدم تغییر قابل توجه پوشش ظاهری شخص در مدت‌زمانی است که در سیستم نظارتی رؤیت می‌گردد. در حالتی که پوشش افراد تغییر نماید تنها خصوصیات بیومتریک هستند که می‌توانند در بازشناسی مورد استفاده قرار گیرند. عواملی نظیر فاصله زیاد افراد از دوربین‌ها، شرایط نوری متغیر محیط و کیفیت سنسور دوربین، منجر به کاهش کیفیت تصاویر ثبت‌شده می‌شود. در نتیجه بسیاری از خصوصیات بیومتریک از جمله ویژگی‌های چهره و طریقه راه رفتن [۶، ۷] قابل استفاده نمی‌باشد. همچنین تغییرات زاویه‌ای در تصویر شخص منجر به تغییرات بزرگ در خصوصیات بیومتریکی شده و عمل تشخیص را بسیار دشوار می‌سازد. لذا فرض عدم تغییر قابل توجه در پوشش ظاهری، فرضی منطقی می‌باشد.

انجام فرآیند بازشناسی مستلزم پیش‌پردازش‌هاییست که عبارتند از: ردیابی و موقعیت‌یابی فرد در تصویر [۸، ۹]، حذف پس‌زمینه [۱۰]، و حذف سایه [۱۱]. منظور از ردیابی، دنبال کردن تصویر فرد در یک دوربین است. روش‌های بسیار متنوعی در ردیابی ارائه شده‌اند اما از آنجایی که شی هدف در بازشناسی، انسان است می‌توان از الگوریتم‌هایی استفاده نمود که محدوده تصویر انسان را در تصویر مشخص می‌کنند [۹]. در حالتی که دنباله‌ای از تصاویر در اختیار باشد با مدل‌سازی پس‌زمینه می‌توان نواحی مرتبط با پس‌زمینه را از تصویر حذف نمود. از روش‌های مرسوم و کارآمد در مدل‌سازی پس‌زمینه می‌توان به مدل مخلوط گاوسی^۲ اشاره نمود. در بسیاری از روش‌های حذف پس‌زمینه، سایه فرد در تصویر به عنوان بخشی از تصویر فرد در نظر گرفته می‌شود که فرآیند بازشناسی و ردیابی را دشوار می‌سازد. به‌منظور رفع این مشکل از روش‌های حذف سایه استفاده می‌شود.

بررسی‌های صورت گرفته در این پژوهش نشان می‌دهد که تغییرات زاویه قرارگیری شخص در تصویر، منجر به بروز مشکلاتی در الگوریتم‌های بازشناسی می‌شوند که در تحقیقات دیگر کمتر مورد توجه قرار گرفته است. به عبارتی تحقیقات و روش‌های کنونی از اطلاعات زاویه قرارگیری شخص استفاده نکرده و در اغلب آنها تاثیرات منجر از تغییرات زاویه‌ای نادیده گرفته می‌شوند. در حالی که ثبت تصاویر از زوایای مختلف، در دوربین‌های مختلف یک محیط

آن‌ها برای انسان است. در نتیجه در سیستم‌هایی که ویژگی‌هایی از فرد خاصی مدنظر باشد سیستم می‌تواند پیشنهاداتی برای کاربر داشته باشد.

در [۲۴] توصیفگر جدیدی با استفاده از ترکیب شبکه عصبی کانولوشن و ویژگی $ELF16^7$ [۲۵] معرفی شده است. این دو بخش مجدداً در ساختار یک شبکه عصبی ترکیب می‌شود. تعداد خروجی‌های این شبکه به تعداد افراد در سیستم است که برای یک شخص خاص، یکی از این خروجی‌ها مقدار یک به خود می‌گیرد. بعد از یادگیری شبکه، لایه ماقبل آخر از بخش کانولوشن شبکه، به عنوان بخشی از بردار ویژگی نهایی در نظر گرفته می‌شود. در مرحله آزمون، ویژگی $ELF16$ برای تصویر ورودی محاسبه می‌شود. همچنین خروجی شبکه عصبی کانولوشن نیز محاسبه می‌گردد. بردار ویژگی نهایی از کنار هم قرار دادن دو بخش فوق حاصل می‌شود.

در دسته روش‌های فعال می‌توان به روش‌هایی اشاره نمود که سعی در استخراج یک توصیفگر با خاصیت متمایزکنندگی قوی دارند. برای مثال در مقاله [۲۶] تصویر بدن انسان به دو ناحیه بالاتنه و پایین‌تنه قطعه‌بندی می‌گردد. سپس برای هر ناحیه خوشه‌بندی مقادیر ویژگی‌های آن برای مجموعه‌ای از تصاویر انسان‌های مختلف محاسبه می‌گردد. برای هر بخش مراکز خوشه‌ها نگه‌داری شده و در مرحله بازنمایی، برای هر فرد برداری از میزان عدم شباهت با هر یک از این مراکز خوشه‌ها محاسبه می‌شود. در برخی روش‌ها ویژگی‌های مورد استفاده به‌طوری وزن‌دهی می‌شوند که بیشترین تمایز بین افراد حاصل گردد. این وزن‌دهی می‌تواند در یک فرآیند یادگیری تطبیقی صورت گیرد [۲۵]. در [۲۷] هدف یادگیری الگوهای تصویری است که می‌توانند بصورت همزمان خاصیت بازنمایی و تمایز را داشته باشند. بردار ویژگی ساخته شده برای هر ناحیه، ترکیبی از توصیفگر $SIFT^8$ برای سه کانال فضای رنگی Lab^9 و هیستوگرام در سه مقیاس می‌باشد. نواحی مذکور در این مقاله فیلترهای سطح میانی نام‌گذاری شده است. به منظور استخراج فیلترهایی با خاصیت متمایزکنندگی بیشتر، برای هر فیلتر میزان نزدیکترین فاصله تا فیلترهای متناظر در تصاویر مختلف محاسبه می‌گردد. بدین ترتیب فیلتری انتخاب می‌شود که به لحاظ فراوانی در مجموعه داده‌ها، دارای مقدار متوسطی باشد.

گروهی از روش‌ها در یک فرآیند یادگیری، سعی در یادگیری تابع انتقال بین دو تصویر گرفته شده برای یک زوج دوربین دارند. بدین ترتیب با استفاده از مجموعه‌ای تصاویر گرفته شده از افراد برای یک زوج دوربین، تابع انتقال این زوج دوربین مدل می‌شود. زوج تصاویر

استفاده می‌کنند. در دسته دوم از اطلاعات پیکربندی دوربین‌ها به همراه اطلاعات تصاویر استفاده می‌شود. دسته اول شامل روش‌های فعال و غیرفعال می‌باشد. منظور از روش‌های فعال روش‌هایی است که در مراحل مختلف، از فرآیند یادگیری استفاده می‌کنند. روش‌های غیرفعال فاقد فاز یادگیری می‌باشند. در گروه روش‌های غیرفعال می‌توان به روش‌هایی که از توصیفگر کوواریانس^۲ استفاده می‌کنند اشاره نمود. این توصیفگر با استفاده از سنجش همبستگی بین ویژگی‌های مختلف یک فرد، به نوعی بافت را در تصویر بازنمایی می‌کند که در نتیجه نسبت به بسیاری از تغییرات مقاوم است [۱۳-۱۵]. از جمله ویژگی‌های مورد استفاده در محاسبه توصیفگر کوواریانس می‌توان به موقعیت، کانال‌های رنگی، اطلاعات گرادیانی و پاسخ فیلتری تصویر اشاره نمود. در مقاله [۱۶] با انطباق تصویر فرد به یک مدل مثلثی و با بهره‌گیری از یک روش قطعه‌بندی عمل بازشناسی انجام می‌شود. در مقاله [۱۷] توصیفگر ترکیبی ارائه گردیده که از دو بخش هیستوگرام در مدل رنگی HSV و آنالیز جوهره^۳ تشکیل می‌شود. در مقاله [۱۸] توصیفگری مشتمل بر سه بخش معرفی شده است. این سه بخش شامل هیستوگرام در فضای رنگی HSV، توصیفگر $MSCR^5$ و نواحی با خصوصیت ساختاری بالا و مکرر ($RHSP^6$) می‌باشد. همچنین ناحیه بدن به سه بخش سر، بالاتنه و پایین‌تنه تقسیم می‌گردد. برای قسمت بالاتنه و پایین‌تنه محور قرینگی عمودی استخراج و ویژگی ترکیبی بالا برای هر ناحیه بصورت مجزا استخراج می‌شود.

در بسیاری از روش‌ها از هیستوگرام به‌عنوان ویژگی مورد استفاده قرار می‌گیرد [۲]. در گروهی از این روش‌ها فضای رنگی به مجموعه شناخته شده‌ای از رنگ‌ها تقسیم می‌شود. سپس توزیع رنگی تصویر بر روی این مجموعه محاسبه می‌گردد. بدین ترتیب هیستوگرامی از فراوانی بر روی مجموعه‌ای گسسته از مقادیر رنگی خواهیم داشت [۱۹، ۲۰]. برای نمونه در [۲۱] تصویر شخص به ۶ بخش افقی تقسیم شده، و برای هر بخش توزیع رنگی روی مجموعه رنگی شامل ۱۶ رنگ برجسته محاسبه می‌شود. برای محاسبه میزان تعلق یک رنگ به یک عضو در مجموعه مذکور از الگوریتم k نزدیکترین همسایه استفاده شده است.

یکی از مشکلات هیستوگرام، نادیده گرفتن ارتباط بین پیکسل‌ها است. لذا روش‌های مختلفی جهت رفع این نقیصه ارائه شده است [۱۹، ۲۲]. در مقاله [۲۳] خصیصه‌های سطح بالاتری نظیر قدبلندی، کوتاه‌قدی، کوله‌پشتی به همراه داشتن و بلند بودن موی فرد در تصویر تشخیص داده می‌شود. سپس این خصیصه‌ها در فرآیند بازشناسی استفاده می‌شوند. مزیت این خصیصه‌ها قابل فهم بودن

پیکربندی دوربین‌ها سعی در کوچک‌تر کردن فضای جستجو می‌باشند [۳۳]. برای مثال با استفاده از مدل‌سازی مکان‌های ورود و خروج افراد در دوربین‌های مختلف، تخمینی از موقعیت احتمالی فرد در هنگام خروج از یک دوربین و ورود به دوربین دیگر حاصل می‌شود. این تخمین می‌تواند در مورد زمان ورود به دوربین‌های مختلف نیز باشد. بدین ترتیب با کمک این روش‌ها می‌توان فضای جستجو را کاهش داد و نرخ تشخیص بازشناسی را بهبود بخشید.

با بررسی روش‌های موجود به این نتیجه خواهیم رسید که هیچ یک از این روش‌ها راهکاری دقیق برای شرایطی که تصاویر از زوایای گوناگون ثبت شده باشند ندارند. در واقع کلیه روش‌های کنونی با معرفی توصیفگرهای مختلف سعی در کاهش تاثیرات منفی ناشی از تغییرات زاویه‌ای دارند. در حالی که تغییرات زاویه‌ای در برخی موارد منجر به تغییراتی در ظاهر شخص می‌شود که جز با تشخیص زاویه و تحلیل ظاهر شخص، نمی‌توان اثرات منفی آن را بر صحت فرآیند بازشناسی کاست.

۳- روش پیشنهادی

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 روشن‌تر شدن این مسئله شکل ۱ سه جفت تصویر مرتبط با سه شخص در دو دوربین مختلف را نشان می‌دهد. تصاویر مذکور از مجموعه داده‌های ViPer انتخاب شده است [۳۴].



الف ب پ ت ث ج

شکل ۱: شش تصویر مربوط به سه شخص از مجموعه داده 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 کمتر رؤیت می‌گردد. همین مشکل برای تصاویر دیگر به دلیل تغییرات زاویه‌ای مشاهده می‌شود. اکثر روش‌های موجود در

می‌تواند مربوط به دو دوربین در شرایط کاملاً متفاوت باشد. در برخی روش‌ها تابع معیار فاصله مورد یادگیری قرار می‌گیرد [۲۸] و در برخی دیگر تابع انتقال رنگی برای هر دو دوربین تخمین زده می‌شود [۲۹].

در [۳۰] ویژگی جدیدی بر پایه هیستوگرام رنگ و بافت از تصاویر استخراج می‌گردد. سپس از ویژگی مذکور در یادگیری معیار فاصله استفاده شده است. به عبارتی نگاشتی از بردار فاصله‌ی دو نمونه حاصل می‌شود که تحت نگاشت مذکور مقدار فاصله برای نمونه‌های مرتبط عددی کوچکتر و برای نمونه‌های غیر مرتبط مقداری بزرگتر می‌باشد. در قسمت دیگری از رابطه‌ی بهینه‌سازی این مقاله، عمل کاهش بعد با در نظر گرفتن قید پیشین بصورت همزمان صورت می‌گیرد. این کاهش بعد منجر به کاهش تاثیرات نویز و پس‌زمینه می‌شود.

در [۳۱] 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. استفاده از شبکه عصبی کانولوشن این مزیت را دربر دارد که مرحله استخراج ویژگی درون شبکه بصورت خودکار صورت می‌گیرد. در صورتی که داده‌های آموزشی کافی باشد، ویژگی‌های آموزش‌یافته در لایه‌های کانولوشن بهینه می‌باشند.

در [۳۲] 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 دو دوربین به زیرنواحی کوچکتری بخش‌بندی شود مشاهده خواهد شد که برای یک جفت دوربین، تناظری بین این نواحی به لحاظ شباهت وجود دارد که تقریباً برای تمامی تصاویر ثبت‌شده یکسان خواهد بود. در این مقاله ماتریسی تحت عنوان ماتریس تناظر ساختاری^۱ برای استخراج احتمال تناظر و تطبیق بین هر دو زیرناحیه برای یک جفت دوربین طی یک فرآیند یادگیری ساخته می‌شود. در مرحله آزمون، برای یک جفت تصویر از یک جفت دوربین، تطبیق زیرنواحی دو تصویر بر اساس میزان تناظر بدست آمده از ماتریس تناظر صورت می‌گیرد.

همان‌طور که گفته شد گروهی دیگر از روش‌ها با استفاده از اطلاعات

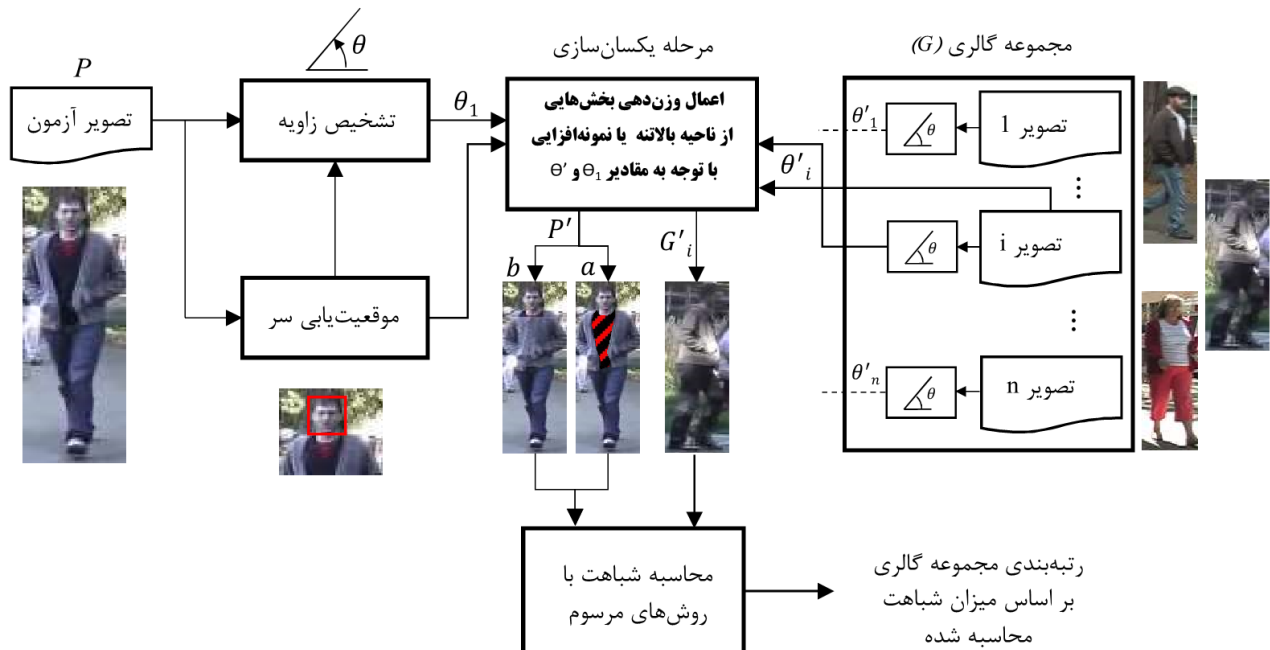
یکسان‌سازی دارند را تشخیص خواهد داد و در موارد دیگر از الگوریتم پایه که می‌تواند هر الگوریتم بازشناسی باشد استفاده می‌کند. شکل ۲ نمای کلی سیستم پیشنهادی را نشان می‌دهد.

ورودی یک سیستم بازشناسی، تصویری است که قصد داریم در مجموعه تصاویر موجود مورد جستجو قرار دهیم و تصویر متناظر ورودی را بازشناسی نماییم. در منابع این حوزه پژوهشی به تصویر مورد جستجو تصویر آزمون^{۱۱} و به مجموعه تصاویر موجود در سیستم مجموعه گالری^{۱۲} اطلاق می‌شود. بدین منظور بردار ویژگی برای کلیه تصاویر استخراج شده و در فضای ویژگی نزدیک‌ترین نمونه یا تصویر از مجموعه گالری انتخاب می‌شود. بر این اساس گام‌های سیستم پیشنهادی در قالب یک الگوریتم به صورت زیر می‌باشد:

- ۱- با استفاده از دسته‌بند CNN موقعیت سر در تصویر مشخص می‌گردد. مجموعه داده آموزشی و معماری شبکه برگرفته از مقاله [۳۵] می‌باشد.
- ۲- زاویه قرارگیری شخص در تصویر آزمون محاسبه می‌شود. تشخیص خودکار زاویه قرارگیری شخص در تصویر با استفاده از روش‌های مختلفی امکان پذیر است [۳۶، ۳]. در این پژوهش از رویکرد ارائه شده در [۳۶] استفاده شده است.
- ۳- زاویه قرارگیری شخص در تصویر آزمون محاسبه می‌شود. تشخیص خودکار زاویه قرارگیری شخص در تصویر با استفاده از روش‌های مختلفی امکان پذیر است [۳۶، ۳].

چنین حالتی قادر به تطبیق دو نمونه نیستند و یا با عدم دقت همراه می‌باشند. 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 ت، پ و الف باشند. در چنین حالتی ناظر انسانی با توجه به دانش پیشینی که از شکل سه‌بعدی بدن انسان در زوایای مختلف دارد به‌خوبی 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. سیستم پیشنهادی در این پژوهش نیز سعی در مدل‌سازی این دانش پیشین دارد. بطور خلاصه در روش پیشنهادی زمانی که در مرحله تطبیق فاصله دو تصویر محاسبه می‌شود تاثیرات چرخش زاویه‌ای بر ظاهر شخص خنثی می‌گردد. در این پژوهش به فرآیند مذکور، یکسان‌سازی تصاویر گفته می‌شود. این فرآیند برای هر مقایسه بصورت مجزا اجرا خواهد شد. به عبارتی زمانی که دو تصویر از نظر زاویه قرارگیری شخص زاویه مشابهی داشته باشند نیازی به فرآیند یکسان‌سازی نیست. بدین ترتیب در مرحله تطبیق پس از استخراج زاویه قرارگیری شخص در دو تصویر، در صورتیکه تفاوت این دو زاویه از حدی بیشتر باشد فرآیند یکسان‌سازی در الگوریتم پیشنهادی انجام می‌شود. علاوه بر این الگوریتم پیشنهادی با استخراج نواحی از بدن فرد، دقیقاً مواردی را که احتیاج به



شکل ۲: نمای کلی از روش پیشنهادی

بسیار دشوار است. درحالی‌که در روش دوم، تصویر نمونه‌افزایی شده به‌سادگی می‌تواند به‌عنوان ورودی شبکه مورد استفاده قرار گیرد. 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 پژوهش از دو الگوریتم معرفی شده در [۳۰] و [۳۷] 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 مثالی از گام یکسان‌سازی تصویر آزمون و دو تصویر نمونه از مجموعه گالری را نشان می‌دهد. فرض کنید تصویر آزمون سمت چپ و تصاویر مجموعه گالری، دو تصویر سمت راست از شکل ۳ باشند. هدف نهایی الگوریتم بازشناسی رتبه‌بندی دو تصویر مجموعه گالری براساس میزان شباهت با تصویر آزمون است. فرآیند یکسان‌سازی قبل از سنجش میزان شباهت با هر یک از دو تصویر از مجموعه گالری، بصورت مجزا بررسی می‌کند که آیا نیاز به تغییراتی (ماسک کردن، نمونه‌افزایی) در هیچ یک از دو تصویر آزمون یا گالری هست یا نه. در مثال شکل ۳ از آنجایی که زاویه تصویر آزمون با تصویر اول از گالری تفاوت زیادی دارد لذا فرآیند یکسان‌سازی نیاز می‌باشد. از آنجایی که تصویر آزمون از روبرو است الگوریتم یکسان‌سازی، این تصویر را به منظور تغییر (ماسک کردن، نمونه‌افزایی) انتخاب می‌کند. در مقایسه تصویر آزمون و تصویر دوم از گالری، از آنجایی که هر دو تصویر از روبرو هستند نیازی به انجام فرآیند یکسان‌سازی نیست و در نتیجه دو تصویر می‌توانند به روش عادی مقایسه گردند. بدین ترتیب خروجی الگوریتم یکسان‌سازی برای مقایسه تصویر آزمون و تصویر اول از گالری شامل ماسک و تصویر نمونه‌افزایی شده‌ای است که می‌تواند در گام استخراج ویژگی از هر الگوریتم بازشناسی مورد استفاده قرار گیرد. در آزمایش‌های انجام شده نشان خواهیم داد که استفاده از سیستم پیشنهادی منجر به بهبود در نرخ تشخیص فرآیند بازشناسی می‌شود.

۳-۱- استخراج بخش میانی از تصویر بالاتنه

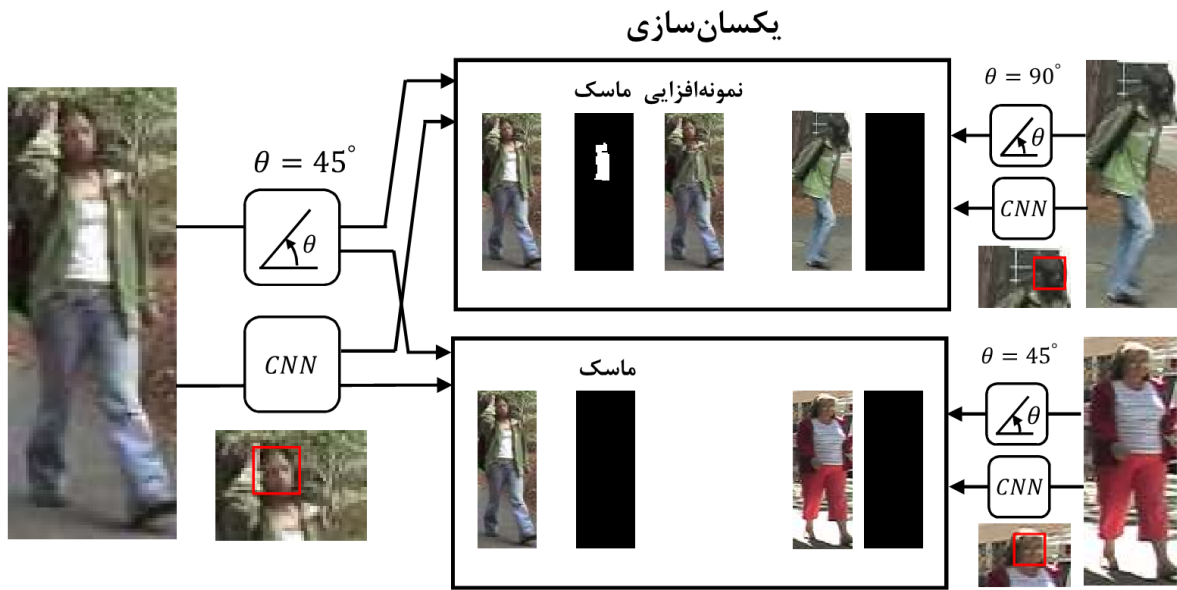
همانطور که در زیربخش قبل گفته شد متناسب با نوع پوشش و زاویه قرارگیری شخص در تصویر، بخش‌هایی از بالاتنه می‌تواند در تصویر پنهان شود. الگوریتم زیر با در دست داشتن زاویه تخمینی شخص و همچنین موقعیت سر، نواحی مذکور را در تصویر استخراج

۴- در مرحله یکسان‌سازی در صورتی که زاویه قرارگیری شخص در تصویر آزمون، تصویری از روبرو باشد (زاویه بین 45° تا $+45^{\circ}$)، به کمک الگوریتم خوشه‌بندی در فضای رنگی Lab، ناحیه‌ای از پوشش فرد که در زیر ناحیه سر واقع شده، در صورتی که شرایط رنگی متفاوتی از نواحی اطراف داشته باشد، استخراج می‌گردد. دلیل انتخاب زوایای 45° تا $+45^{\circ}$ در این است که تحت این زوایا پوشش شخص در زیر ناحیه سر، قابل رؤیت بوده و تحت چرخش زاویه‌ای ممکن است غیر قابل مشاهده باشد. در رویکرد پیشنهادی خروجی این مرحله به دو حالت در نظر گرفته می‌شود. 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^{۱۳} 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 ۲ به ترتیب با برچسب‌های a و b مشخص شده‌اند. برای تصاویر موجود در مجموعه گالری نیز فرآیند مشابهی انجام می‌گیرد.

۵- در مرحله تطبیق و محاسبه شباهت، در دو حالت، از ماسک‌های بدست آمده استفاده می‌شود: حالت اول زمانی که تصویر آزمون (P) از روبرو و تصویر مورد بررسی در مجموعه گالری (Gi)، در زاویه‌ای غیر از روبرو باشد. این حالت بدین معنی است ناحیه استخراج شده از پوشش فرد در تصویر P، در تصویر Gi قابل رؤیت نیست و ناحیه مذکور ماسک می‌گردد. منظور از ماسک کردن در نظر نگرفتن ناحیه مذکور در فرآیند استخراج ویژگی می‌باشد. حالت دوم زمانی است که تصویر Gi از روبرو و تصویر P در زاویه روبرو نباشد. در این حالت، تصویر Gi ماسک می‌گردد.

۶- 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 الگوریتم‌ها قابل استفاده می‌باشد. اعمال روش ماسک‌گذاری مستلزم انجام تغییرات بیشتری در گام استخراج ویژگی است. برای مثال در الگوریتم‌هایی که استخراج ویژگی آنها توسط شبکه عصبی کانولوشنی عمیق^{۱۴} صورت می‌گیرد، استفاده از روش ماسک‌گذاری

می‌کند. در صورتی که زاویه تخمین زده شده از روبرو نباشد الگوریتم پیش رو اجرا نمی‌شود و به عبارت دیگر ناحیه‌ای برای ماسک شدن انتخاب نمی‌گردد. جزئیات این الگوریتم در قالب گام‌های زیر بیان شده است:



شکل ۳: مرحله یکسان‌سازی در روش پیشنهادی

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.

$$Select(I_j^k) = \begin{cases} 1 & \frac{sum(I_j^k \otimes M)}{sum(I_j^k)} > \tau_1 \ \& \ \frac{sum(C_k \otimes M)}{sum(C_k)} > \tau_2 \\ 0 & else \end{cases} \quad (4)$$

در این رابطه C_k و I_j^k مقدار دودویی متناظر با نواحی C_k و I_j^k هستند. عملگر \otimes ضرب درایه‌های متناظر دو ماتریس می‌باشد و همچنین تابع $sum(\cdot)$ مجموع درایه‌های ماتریس آرگومان را محاسبه می‌کند. با توجه به رابطه ۴، بخش (I_j^k) در صورتی انتخاب می‌شود که بیشتر از τ_1 درصد از مساحت آن زیر ناحیه سر قرارگیرد. همچنین τ_2 درصد از مساحت خوشه متناظر با بخش مذکور نیز باید در زیر ناحیه سر قرار گیرد.

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

$$\mathbb{O} = \bigcup_{j=1}^n I_j^k \times Select(I_j^k) \quad (5)$$

۷- اتمام.

بطور کلی در الگوریتم بالا بخش‌های بدست آمده از الگوریتم قطعه‌بندی، مجدداً خوشه‌بندی می‌شوند. در واقع عمل خوشه‌بندی،

۱- اولین پیش‌پردازش مورد نیاز قطعه‌بندی^{۱۵} تصویر است. در این پژوهش روش‌های قطعه‌بندی گوناگونی مورد بررسی قرار گرفته است [۳۸]. به منظور پیاده‌سازی از بسته نرم‌افزاری VLFeat به منظور استفاده شده است.

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

در رابطه ۱ تصویر ورودی I به n بخش قطعه‌بندی می‌شود. ۲- برای هر یک از n بخش، بردار ویژگی شامل مؤلفه‌های فضای رنگی^{۱۶} Lab و موقعیت مرکز بخش (x و y) مورد نظر، تشکیل می‌شود.

$$f_i = \langle L(I_i), a(I_i), b(I_i), x(I_i^{center}), y(I_i^{center}) \rangle \quad (2)$$

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

$$\begin{cases} Cluster(F) = \bigcup_{k=1}^4 C_k \\ C_k = \bigcup_{j \in \{1, \dots, n\}} I_j^k \end{cases} \quad (3)$$

به عبارتی در این گام، بخش‌های بدست آمده از قطعه‌بندی گام نخست مجدداً خوشه‌بندی شده و در نتیجه بخش‌های بزرگ‌تری حاصل می‌شود. با توجه به ویژگی‌های مورد استفاده، بخش‌هایی که در یک خوشه قرار می‌گیرند دارای شباهت رنگی و مکانی بصورت همزمان می‌باشند.

به صورت همزمان با یادگیری معیار فاصله صورت می گیرد. روش پیشنهادی در این پژوهش تغییری در قسمت یادگیری معیار فاصله ایجاد نمی کند.



شکل ۴: مثالی از نتایج الگوریتم استخراج بخش میانی تصویر بالاتنه

در [۳۷] بازنمایی جدیدی تحت عنوان توصیفگر گاوسی سلسله مراتبی^{۱۹} بر پایه توصیفگر کوواریانس معرفی شده است. در این بازنمایی برای نواحی مختلف تصویر توزیع آماری توصیفگرهای کوواریانس برای زیر نواحی کوچکتر محاسبه شده و در نتیجه خصوصیات رنگی و بافتی برای هر ناحیه بازنمایی می شود. در زیربخش های بعدی به نحوه تلفیق روش پیشنهادی در دو الگوریتم [۳۷] و [۳۰] پرداخته می شود.

۳-۲-۱- تلفیق ویژگی ماکزیمم وقوع محلی با روش پیشنهادی

در ویژگی معرفی شده در [۳۰] ابتدا تبدیلی بر اساس تبدیل Retinex به صورت چندمقیاسی^{۲۰} بر روی تصاویر ورودی به عنوان پیش پردازش انجام می شود. در این تبدیل شرایط نوری و رنگی تصویر با الگوبرداری از سیستم ادراک بینایی انسان اصلاح می گردد. سپس تصویر حاصله به فضای رنگی HSV نگاشت می گردد. در این ویژگی علاوه بر توصیف رنگ، از توصیفگر SILTP^{۲۱} که نسخه بهبود یافته ای از توصیفگر LBP^{۲۱} است به منظور توصیف بافت ها در تصویر استفاده می شود [۳۹]. از ویژگی های این توصیفگر مقاومت در برابر تغییرات شرایط نوری و وجود نویز در تصویر است. توصیفگر LOMO از هیستوگرام های رنگی HSV و SILTP^{۲۲} در سه مقیاس تشکیل می گردد.

به منظور ماسک گذاری و وزن دهی نواحی استخراج شده در تصویر و ترکیب با توصیفگر LOMO، باید از هیستوگرام وزن دار بجای

وابستگی رنگی و مکانی بخش های مختلف تصویر را مشخص می کند. رابطه (۴) بیان می دارد که یک بخش، زمانی به عنوان ناحیه خاص استخراج می شود که بیش از τ_1 درصد از مساحت آن در زیر ناحیه سر واقع گردد. همچنین حداقل τ_2 درصد از مساحت خوشه متناظر با بخش مذکور نیز باید در زیر ناحیه سر واقع شده باشد. برقراری دو شرط مذکور، خاص بودن ناحیه زیر سر را به لحاظ رنگی تضمین می کند. برای مثال در صورتی که بخشی از تصویر که زیر ناحیه سر قرار دارد به لحاظ رنگی مشابه با دیگر نواحی باشد، شرط τ_2 نقض شده و بخش مذکور انتخاب نمی شود.

شکل ۴ خروجی الگوریتم بالا را بر روی چند نمونه تصویر، برای مقادیر $\tau_1 = 0.7$ و $\tau_2 = 0.4$ نشان می دهد. از آنجایی که الگوریتم بالا تنها بر روی تصاویر از روبرو اجرا می شود، در اندک مواردی منجر به تشخیص غلط می گردد. انتخاب مقدار برای این دو پارامتر با توجه به معنای این دو پارامتر انجام شده است. با توجه به هدف مسئله، باید سعی شود نواحی انتخاب گردد که درصد اشتراک بیشتری با ناحیه زیر سر داشته و این به معنای انتخاب مقدار بزرگ برای پارامتر τ_1 می باشد. برای خوشه های متناظر با نواحی انتخابی نیز تحلیل مشابهی قابل انجام است. مقدار خیلی کوچک برای پارامتر τ_2 نشان دهنده خوشه هایی از تصویر است که متعلق به پس زمینه یا پوشش کلی شخص است که مطلوب الگوریتم پیشنهادی نیست. بدین ترتیب برای پارامتر τ_2 مقدار کمتری نسبت به τ_1 مناسب می باشد. بدین ترتیب مقادیر این دو پارامتر با توجه به پیش فرض های بیان شده و آزمایش های صورت گرفته بر روی مجموعه اعتبارسنجی برای این بخش از الگوریتم پیشنهادی، تعیین گردیده است. اشاره به این نکته ضروری است که هرچه کارآمدی این الگوریتم بالاتر باشد، نرخ تشخیص نهایی بازشناسی بیشتر می شود و بهبود بیشتری را شاهد خواهیم بود. در واقع نتایج گزارش شده برآیندی از اعمال پی در پی بخش های رویکرد پیشنهادی است. بطور مشابه دقت در تشخیص ناحیه سر نیز بر کارآمدی نهایی الگوریتم تاثیر خواهد گذاشت.

۳-۲-۲- تلفیق مرحله یکسان سازی با الگوریتم های بازشناسی

به منظور ارزیابی روش پیشنهادی، دو الگوریتم بازشناسی [۳۷] و [۳۰] انتخاب و مرحله یکسان سازی در گام استخراج ویژگی آنها استفاده شده است. در [۳۰] بازنمایی جدیدی با نام ماکزیمم وقوع محلی (LOMO)^{۱۷} به عنوان بردار ویژگی، بر پایه هیستوگرام معرفی شده است. همچنین معیار فاصله و زیرفضای جدیدی تحت عنوان XQDA^{۱۸} معرفی شده است. تفاوت روش یادگیری معیار فاصله در این مقاله با بقیه روش های هم گروه، در کاهش بعدی است که

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

در توصیفگر گاوسی سلسله‌مراتبی، برای نواحی و زیرنواحی تصویر، به‌صورت سلسله‌مراتبی و پیاپی، مدلی گاوسی به منظور توصیف رنگ و بافت ایجاد می‌شود. این مدل گاوسی شامل بردار میانگین و ماتریس کوواریانس است. بدین منظور ابتدا تصویر به نواحی سطری تقسیم‌بندی می‌شود. این تقسیم‌بندی به گونه‌ای است که سطرها دارای همپوشانی باشند. سپس هر سطر به زیرنواحی به شکل پنجره‌های مربعی تقسیم می‌شود. برای هر زیرناحیه مدل گاوسی تک‌مده^{۲۴} شامل بردار میانگین و ماتریس کوواریانس محاسبه می‌شود. از آنجایی که ماتریس کوواریانس در فضای ریمانی واقع شده، تخمینی از این فضا به فضای اقلیدسی انجام می‌گردد. در این توصیفگر توزیع گاوسی مذکور تبدیل به برداری در فضای مماسی^{۲۵} اقلیدسی می‌شود. در نتیجه برای هر ناحیه مجموعه‌ای از بردارها خواهیم داشت. برای یک ناحیه مجدداً مدل گاوسی دیگری بر روی بردارهای مذکور تخمین زده می‌شود. بردار ویژگی نهایی، از تبدیل مدل گاوسی نهایی به فضای مماسی اقلیدسی حاصل می‌شود. در مدل‌سازی گاوسی نهایی در توصیفگر گاوسی سلسله‌مراتبی، بردارهای فضای مماسی متناظر با زیر نواحی وزن‌دهی می‌گردند.

انگیزه این وزن‌دهی اهمیت‌دهی به زیرنواحی مرتبط با مرکز تصویر است. در واقع با دور شدن از محور عمودی مرکز تصویر، احتمال تعلق زیرنواحی به بدن شخص کمتر می‌شود. روابط زیر مکانیزم وزن‌دهی مذکور را نشان می‌دهند.

$$\mu^G = \frac{1}{\sum_{s \in G} w_s} \sum_{s \in G} w_s g_s \quad (7)$$

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

که در آن μ^G و Σ^G به ترتیب بردار میانگین و ماتریس کوواریانس متناظر با ناحیه G ، g_s بردار فضای مماسی و w_s وزن متناظر با زیرناحیه s از ناحیه G می‌باشند. به منظور اعمال ماسک وزنی به این توصیفگر، وزن‌های مربوط به زیرنواحی متناظر با ماسک در رابطه (۷) تغییر داده می‌شود. همچنین اعمال روش نمونه‌افزایی به این توصیفگر، مشابه روش قبل می‌باشد. در مقاله مذکور از توصیفگر گاوسی سلسله‌مراتبی با اختصار GoG^{۲۶} نام برده می‌شود. نتایج گزارش شده در این مقاله با استفاده از یادگیری معیار فاصله (الگوریتم XQDA) می‌باشد.

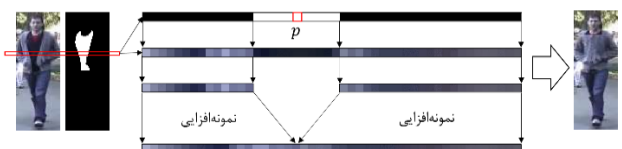
هیستوگرام مرسوم استفاده کرد. بدین ترتیب که برای پنجره‌های لغزنده شده بر روی تصویر، نگاشت HSV و SILTP، مشابه حالت قبل محاسبه می‌گردد. با توجه به بازه‌های فراوانی مورد نظر در محاسبه هیستوگرام HSV، برای هر پیکسل از تصویر مقداری عددی خواهیم داشت که حاصل چندی‌سازی^{۲۳} در هر کانال رنگی است. خروجی توصیفگر 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.

$$wHist^i(I, M) = \sum_{\{x,y|I(x,y)=i\}} M(x,y) \quad (8)$$

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 از الگوریتم پیشنهادی هیستوگرام‌های HSV و SILTP بصورت وزن‌دار محاسبه می‌شوند. در روش نمونه‌افزایی ابتدا ناحیه ماسک در تصویر استخراج می‌شود. سپس برای سطری از تصویر که شامل پیکسل‌های ماسک هستند روال زیر اجرا می‌شود:

- ۱- تعیین موقعیت پیکسل میانی در پیکسل‌های ماسک (p).
- ۲- نمونه‌افزایی پیکسل‌های غیرماسک در سمت راست پیکسل p به اندازه فاصله پیکسل p تا ضلع راست تصویر.
- ۳- نمونه‌افزایی پیکسل‌های غیرماسک در سمت چپ پیکسل p به اندازه فاصله پیکسل p تا ضلع چپ تصویر.

بنابراین پیکسل‌های ناحیه ماسک مقادیر رنگی مشابهی با نواحی مجاور به خود می‌گیرند. شکل ۵ خلاصه روال فوق را برای یک سطر از تصویر و ماسک متناظر نشان می‌دهد. در نهایت برای تصویر نمونه‌افزایی شده ویژگی LOMO به‌صورت استاندارد قابل محاسبه می‌باشد.



شکل ۵: مثالی از نتیجه الگوریتم نمونه‌افزایی برای یک تصویر

۳-۲-۳- پیش فرض‌های روش پیشنهادی

در این بخش پیش فرض‌هایی که در مسئله بازشناسی در نظر گرفته شده بیان می‌شود. برخی از این پیش فرض‌ها مربوط به اکثریت سیستم‌های بازشناسی است و برخی مختص به روش پیشنهادی این پژوهش می‌باشند.

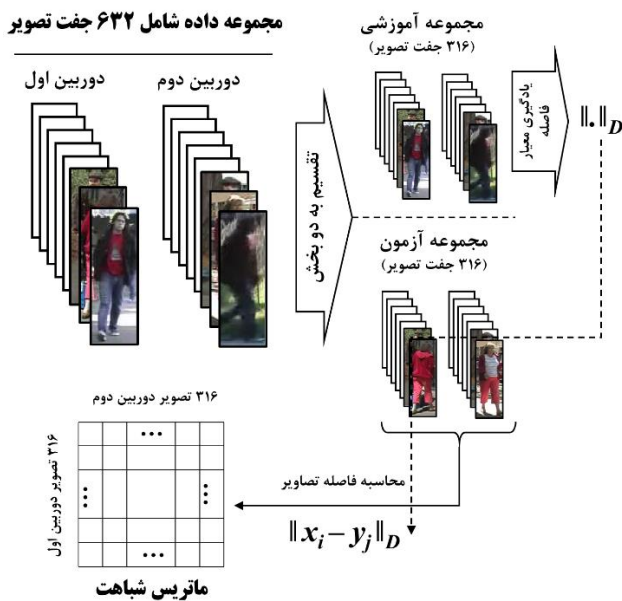
- مسئله بازشناسی بصورت مجموعه بسته^{۲۷} صورت می‌گیرد. منظور از مجموعه بسته، وجود تصویر آزمون در مجموعه گالری است. در مجموعه باز تصویر آزمون می‌تواند در مجموعه گالری نباشد که در این صورت سیستم بازشناسی باید عدم وجود تصویر در مجموعه گالری را تشخیص دهد.
- پوشش فرد از یک دوربین به دوربین دیگر تغییر نمی‌کند. به عبارتی فاصله زمانی قابل توجهی بین ورود و خروج افراد در بین دوربین‌ها وجود ندارد.
- از پیش فرض‌هایی که تحت آن مزیت و توانمندی روش پیشنهادی مشخص می‌شود تغییرات زاویه‌ای زیاد بین تصاویر مختلف سیستم بازشناسی است. در راستای این پیش فرض یکی از مجموعه داده‌های موجود به منظور ارزیابی انتخاب گردید.

۴- ارزیابی و مقایسه روش پیشنهادی

به منظور ارزیابی الگوریتم‌های بازشناسی از مجموعه داده‌های مختلفی استفاده می‌شود. در پژوهش جاری از مجموعه داده ViPer استفاده شده است [۳۴]. دلیل این انتخاب سازگاری تصاویر این مجموعه داده با فرضیات و ایده روش پیشنهادی است. همانطور که پیشتر بیان گردید، روش پیشنهادی در این پژوهش زمانی منجر به بهبود در نرخ بازشناسی می‌شود که تغییرات زاویه‌ای قابل توجه باشد. در چنین شرایطی بخش‌های مختلف پوشش شخص می‌تواند در زوایای مختلف آشکار یا نهان شود. از این منظر مجموعه داده ViPer دارای چنین خاصیتی است. در مجموعه داده‌هایی که دارای چنین شرایطی نباشند اعمال روش پیشنهادی بر روی الگوریتم‌های بازشناسی بهبودی در نرخ بازشناسی حاصل نمی‌کند. بنابراین می‌توان گفت که مجموعه داده ViPer به خوبی می‌تواند مزیت‌ها و کاربرد ایده مطرح شده در این پژوهش را برجسته کرده و تاثیرگذاری آن را نشان دهد.

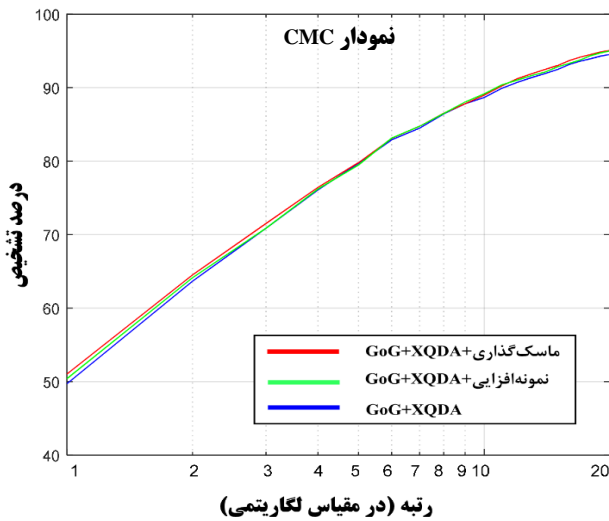
مجموعه داده شامل تصویر ۶۳۲ فرد می‌باشد. برای هر فرد دو تصویر مربوط به دو دوربین مختلف تهیه شده است. دو تصویر مربوط به یک فرد از زوایای مختلف می‌باشند. شرایط نوری در دو دوربین متنوع می‌باشد. تصاویر هر فرد به مستطیلی محدود گشته و در اندازه

۱۲۸×۴۸ پیکسل می‌باشد. این مجموعه داده سخت‌ترین مجموعه داده‌ای است که در بازشناسی مورد استفاده قرار گرفته است. از آنجایی که در روش‌های مقایسه شده از یادگیری معیار فاصله در بخشی از الگوریتم استفاده شده است، احتیاج به یک فرآیند آموزش داریم. بدین منظور داده‌های ۶۳۲ جفت تصویر که تصاویر افراد در دو دوربین هست بصورت تصادفی به دو بخش تقسیم می‌شود. در نتیجه ۳۱۶ جفت تصویر برای آموزش و ۳۱۶ جفت تصویر برای آزمون در نظر گرفته می‌شود. سپس با استفاده از جفت تصاویر مجموعه آموزشی معیار فاصله آموزش می‌یابد که با استفاده از آن فاصله هر یک از ۳۱۶ تصویر دوربین اول از ۳۱۶ تصویر دوربین دوم از مجموعه آزمون محاسبه می‌گردد. تمامی این فاصله‌ها درون ماتریسی با ابعاد ۳۱۶ در ۳۱۶ ذخیره می‌شود. بنابراین انتظار می‌رود اطلاعات بر روی قطر اصلی ماتریس که افراد یکسانی را در دو دوربین نشان می‌دهند نسبت به اعداد دیگر ماتریس در هر ستون کوچکترین مقدار باشد. به عبارتی الگوریتم بازشناسی باید برای یک شخص خاص در دو دوربین، مقدار فاصله را کوچکترین مقدار محاسبه نماید. شکل ۶ تقسیم‌بندی مجموعه داده‌ها را برطبق توضیحات بالا نشان می‌دهد.



شکل ۶: طریقه تقسیم‌بندی مجموعه داده به مجموعه‌های آموزشی و آزمون و ساخت ماتریس شباهت برای مجموعه آزمون

به منظور اطمینان از معنادار بودن میزان بهبود در الگوریتم‌های استفاده شده با اعمال روش پیشنهادی، فرآیند بالا را که شامل تقسیم‌بندی تصادفی ۶۳۲ جفت تصویر است را دفعات زیاد انجام می‌دهیم و در نهایت نتایج گزارش شده میانگین نتایج برای کلیه آزمایش‌ها می‌باشد. نتایج حاصل شده از اعمال روش پیشنهادی بر هر دو الگوریتم بهبود مشابهی را نشان می‌دهد که حاکی از معنادار



شکل ۷: نتایج بکارگیری رویکرد پیشنهادی در الگوریتم GoG+XQDA در قالب نمودار CMC

آزمایش دیگری به منظور نشان دادن نحوه عملکرد الگوریتم پیشنهادی انجام گردید. در این آزمایش، تصاویری که الگوریتم معرفی شده در [۳۷] به غلط بازشناسی نموده و در مقابل، الگوریتم بهبود داده شده توسط رویکرد پیشنهادی (روش ماسک‌گذاری)، به درستی بازشناسی کرده در شکل ۸ نشان داده شده است. در این شکل ستون اول تصویری از دوربین یک است که باید تصویر متناظر آن در دوربین دوم بازشناسی گردد. همانطور که در شکل مشاهده می‌شود الگوریتم [۳۷] در حالت پایه تصویر درست را در رتبه‌ای غیر از یک بازشناسی می‌کند. در مقابل رویکرد پیشنهادی با استفاده از ماسک‌گذاری تصویر درست را در رتبه یک بازشناسی می‌کند. با دقت در این شکل در می‌یابیم که پوشش جلوی شخص در تصویر، زمانی که تغییرات زاویه‌ای بین تصاویر دوربین‌ها وجود داشته باشد منجر به گمراهی و تشخیص غلط الگوریتم می‌گردد. همچنین با دقت در ستون پنجم این شکل دیده می‌شود که فرآیند نمونه‌افزایی در برخی موارد منجر به بهم‌ریختگی جزئی در تصویر می‌گردد. به همین دلیل نرخ تشخیص در بازشناسی به روش نمونه‌افزایی نسبت به روش ماسک‌گذاری پایین‌تر است.

جدول ۱ نتایج مقایسه را به صورت عددی برای رتبه‌های ۱، ۱۰ و ۲۰ نشان می‌دهد. در این جدول، روش پیشنهادی در دو حالت ماسک‌گذاری و نمونه‌افزایی بر روی توصیفگرهای GoG و LOMO مورد مقایسه قرار گرفته است. توصیفگر GoG به همراه الگوریتم یادگیری معیار فاصله XQDA از جمله بالاترین نرخ بازشناسی را برای مجموعه داده ViPer گزارش نموده است.

بودن این اختلاف 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 صحت تشخیص از نمودار CMC^{۲۸} استفاده می‌شود. در این نمودار محور افقی نشان‌دهنده رتبه و محور عمودی نرخ تشخیص را به درصد نشان می‌دهد. از آنجایی که الگوریتم‌های بازشناسی برای هر تصویر آزمون، مجموعه گالری را رتبه‌بندی می‌کنند، در نتیجه الگوریتمی که تصاویر مرتبط را در رتبه بالاتری قرار دهد از کارآمدی بیشتری برخوردار است.

ساخت این نمودار با استفاده از ماتریس فاصله‌ای است که پیش‌تر به آن اشاره شد. بدین منظور برای هر شخص از دوربین اول اندیس اشخاص در دوربین دوم بر اساس مقدار فاصله محاسبه شده از کوچک به بزرگ مرتب می‌شود. در این صورت الگوریتم ایده‌آل الگوریتمی است که برای هر شخص از دوربین یک، اندیس همان شخص در دوربین دوم را در رتبه اول در لیست مرتب‌شده قرار دهد. بنابراین در نمودار CMC رتبه یک، درصد مواردی از ۳۱۶ مورد است که الگوریتم، اندیس درست را در رتبه یک تشخیص داده است. برای مثال رتبه ۱۰ در این نمودار بیانگر درصد مواردی از ۳۱۶ مورد است که الگوریتم مذکور اندیس درست را در رتبه‌ای کوچکتر از ۱۰ تشخیص داده است. به بیانی دیگر زمانی که تصویری به عنوان ورودی به سیستم بازشناسی داده می‌شود و سیستم بازشناسی لیست مرتب‌شده‌ای بر اساس مقدار فاصله از تصاویر موجود در پایگاه می‌سازد، رتبه یک نشان‌دهنده احتمال تشخیص درست الگوریتم در اولین انتخابش می‌باشد. با توجه به جدول ۱، الگوریتم معرفی شده در [۳۷]، در ۸۸،۹۹ درصد موارد، اندیس درست را در ۱۰ مورد ابتدایی از لیست مرتب‌شده قرار می‌دهد. بنابراین جنس اعداد در نمودار CMC و جدول ۱ از جنس درصد تشخیص می‌باشند.

به منظور بررسی تاثیر استفاده از مرحله یکسان‌سازی بر روی صحت بازشناسی، تلفیقی از روش پیشنهادی و توصیفگرهای ارائه شده در [۲۹] و [۳۷] انجام گردید. جزئیات و نحوه این تلفیق در زیربخش ۲-۳ تشریح گردید. مقایسات در قالب نمودار CMC و جدولی از مقایسه روش‌های مختلف براساس رتبه‌های مختلف می‌باشد. شکل ۷ نمودار CMC تا رتبه ۲۰ را برای توصیفگر GoG و ترکیب آن با رویکرد پیشنهادی نشان می‌دهد.

جدول ۱: نتایج بازشناسی برای تلفیق رویکرد پیشنهادی با الگوریتم‌های LOMO, GoG و مقایسه با روش‌های دیگر برای سه رتبه ۱، ۲ و ۳

روش	رتبه ۱	رتبه ۲	رتبه ۳
GoG+XQDA [۳۷]	۴۹/۷۱	۸۸/۶۷	۹۴/۵۲
LOMO+ XQDA [۳۰]	۴۰/۰۰	۸۰/۵۱	۹۱/۰۸
SCNCD [۲۱]	۳۷/۸۰	۸۱/۲۰	۹۰/۴۰
kBiCov [۱۴]	۳۱/۱۱	۷۰/۷۱	۸۲/۴۵
Mid-level Filter [۲۷]	۲۹/۱۱	۶۵/۹۵	۷۹/۸۷
SDALF [۱۸]	۱۹/۸۷	۴۹/۳۷	۶۵/۷۳
KISSME [۴۰]	۱۹/۶۰	۶۲/۲۰	۷۷/۰۰
روش پیشنهادی (نمونه‌افزایی - ماسک‌گذاری)			
GoG+XQDA + ماسک‌گذاری	۵۱/۰۴	۸۸/۹۹	۹۵/۰۶
GoG+XQDA + نمونه‌افزایی	۵۰/۴۱	۹۰/۳۸	۹۵/۰۰
LOMO+ XQDA + ماسک‌گذاری	۴۰/۹۸	۸۰/۸۹	۹۱/۰۴
LOMO+ XQDA + نمونه‌افزایی	۴۰/۰۹	۸۱/۳۰	۹۱/۰۴

این نتایج نشان می‌دهد که استفاده از یکسان‌سازی در مرحله تطبیق از الگوریتم‌های بازشناسی منجر به بهبود در نرخ تشخیص می‌شود. این نتایج نشان می‌دهد که ترکیب توصیفگر GoG با روش پیشنهادی منجر به بهبود ۱/۳ درصد تشخیص در رتبه یک برای ۳۱۶ شخص شده است. همچنین برای توصیفگر LOMO نیز بهبود ۱ درصدی در رتبه یک حاصل شده است. این نتایج نشان می‌دهد که فارغ از نوع الگوریتم بازشناسی، استفاده از رویکرد پیشنهادی منجر به بهبود در نتایج می‌گردد.

مقایسه دو روش نمونه‌افزایی و ماسک‌گذاری نشان می‌دهد که روش ماسک‌گذاری بهبود بیشتری را حاصل می‌کند. دلیل ضعف روش نمونه‌افزایی نسبت به روش ماسک‌گذاری، بهم‌ریختگی‌هایی است که در برخی موارد در تصویر ایجاد می‌شود. در الگوریتم نمونه‌افزایی سعی شده با ساده‌ترین و سریع‌ترین روش ناحیه ماسک هم‌رنگ نواحی مجاور آن شود. اما این سادگی در برخی موارد مشکل‌زا می‌شود و بر روی خصوصیات بافتی تصویر تأثیرات منفی مصنوعی ایجاد می‌کند. ستون پنجم از سطر شش در شکل ۸ مثالی از نمونه‌افزایی است که اختلالاتی در تصویر پدید آمده است. همانطور که قبلاً اشاره شد دلیل معرفی روش نمونه‌افزایی سادگی بکارگیری آن در الگوریتم‌های بازشناسی است. هرچقدر بتوان الگوریتم کارآمدتری را در نمونه‌افزایی معرفی کرد، بهبود بازشناسی نیز افزایش خواهد یافت.



شکل ۸: مقایسه الگوریتم [۳۷] در دو حالت استاندارد و با اعمال رویکرد پیشنهادی

و فناوری اطلاعات، دانشگاه صنعتی نوشیروانی بابل، جلد ۱، شماره ۲، صفحات ۱۹-۲۸، ۱۳۹۱.

[۳] علی سبطی، حمید حسن‌پور، "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.

- [4] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image and Vision Computing*, vol. 32, no. 4, pp. 270-286, 2014.
- [5] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 29, 2013.
- [6] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, "3D ellipsoid fitting for multi-view gait recognition," in *8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pp. 355-360, 2011.
- [7] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 406-415, 2010.
- [8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886-893, 2005.
- [9] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 1491-1498, 2006.
- [10] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 246-252, 1999.
- [11] N. Jovic, A. Perina, M. Cristani, V. Murino, and B. Frey, "Stel component analysis: Modeling spatial correlations in image class structure," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2044-2051, 2009.
- [12] 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, no. 16, pp. 2401-2411, 2007.
- [13] J. Metzler, "Appearance-based re-identification of humans in low-resolution videos using means of covariance descriptors," in *IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pp. 191-196, 2012.
- [14] B. Ma, Y. Su, and F. Jurie, "Bicov: a novel image representation for person re-identification and face verification," in *British Machine Vision Conference*, p. 11, 2012.
- [15] 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*, pp. 806-820, 2012.
- [16] N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 1528-1535, 2006.
- [17] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," *Pattern Recognition Letters*, vol. 33, no. 7, pp. 898-903, 2012.
- [18] 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, no. 2, pp. 130-144, 2013.
- [19] A. D'Angelo and J.-L. Dugelay, "People re-identification in camera networks based on probabilistic color histograms," in *Visual Information Processing and Communication*, p. 78820K, 2011.
- [20] Z. J. Xiang, Q. Chen, and Y. Liu, "Person re-identification by fuzzy space color histogram," *Multimedia tools and applications*, vol. 73, no. 1, pp. 91-107, 2014.
- [21] 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*, pp. 536-551, 2014.
- [22] S. T. Birchfield and S. Rangarajan, "SpatioGrams versus histograms for region-based tracking," in *IEEE Computer Society Conference*

همچنین در بخش‌های مختلف رویکرد پیشنهادی سعی شده کمترین پیچیدگی محاسباتی را داشته باشیم. در بخش تخمین زاویه یک‌بار محاسبه توصیفگر HOG^{۲۹} را خواهیم داشت. همچنین دسته‌بندی‌های استفاده شده در آن به صورت خطی بوده که سر بار محاسباتی بسیار پایینی دارند. در بخش استخراج ناحیه سر تنها ۲۰ سطر بالای تصویر مورد جستجو قرار می‌گیرد. همانطور که عنوان شد استخراج ناحیه سر با کمک شبکه عصبی کانولوشنی انجام شده است. ابزارهای موجود برای آموزش و تست این شبکه‌ها از پردازنده گرافیکی به منظور موازی سازی و سرعت در اجرا بهره برده اند. در این پژوهش نیز به منظور پایین آوردن هرچه بیشتر زمان پردازش سیستم از چنین ابزارهایی استفاده شده است.

۵- نتیجه‌گیری و کارهای آتی

هدف از این پژوهش ارائه رویکردی مبتنی بر نحوه قضاوت ناظر انسانی در مسئله بازشناسی است. نتایج حاصله نشان می‌دهند که استفاده از اطلاعات زاویه‌ای در تصویر می‌تواند به فرآیند بازشناسی در گام تطبیق کمک نماید. بدین منظور در روش پیشنهادی بخشی تحت عنوان یکسان‌سازی طراحی گردید. در این بخش ابتدا زاویه تصویر ورودی محاسبه می‌شود، سپس با استفاده از موقعیت ناحیه سر در تصویر، بخشی از پوشش بالانته فرد که تحت چرخش زاویه‌ای آشکار یا نهان می‌گردد، انتخاب و در مرحله تطبیق، وزن دهی یا نمونه‌افزایی می‌شود. عمل یکسان‌سازی تنها در مواقعی اعمال می‌گردد که تصویر آزمون و تصویر مورد بررسی دارای شرایط خاص زاویه‌ای باشند. در نهایت روش پیشنهادی با دو الگوریتم از کارآمدترین الگوریتم‌های بازشناسی ترکیب گردید. آزمایش‌های صورت گرفته بهبود در نرخ بازشناسی را نشان می‌دهند. شایان ذکر است که روش پیشنهادی قابلیت استفاده در بسیاری از الگوریتم‌های بازشناسی را داراست و می‌تواند به عنوان یک مرحله پیش‌پردازشی مورد استفاده قرار گیرد. در راستای تکمیل این پژوهش می‌توان بخش‌های دیگری از پوشش فرد (مانند کوله‌پشتی و کیف همراه) که تحت چرخش زاویه‌ای منجر به تغییرات ظاهری زیادی در فرد می‌شود را شناسایی و فرآیند یکسان‌سازی مشابهی با آنچه در این پژوهش معرفی شد، اعمال نمود.

مراجع

- [۱] علی سبطی، حمید حسن‌پور، "بازشناسی انسان در سیستم‌های نظارت ویدئویی"، *مجله محاسبات نرم، دانشگاه کاشان، دوره سوم، شماره اول، صفحات ۶۲-۸۱، بهار ۱۳۹۳.*
- [۲] حمید حسن‌پور، علی قنبری، "ردیابی انسان در شبکه‌ای از دوربین‌ها با دید مجزا بر اساس هیستوگرام قسمت‌های مختلف بدن"، *مجله رایانش نرم*

- [33] 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, no. 2, pp. 146-162, 2008.
- [34] 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)*, vol. 3, no. 5, pp. 1-7, 2007. ViPer dataset: accessed 22 July 2019, <<https://vision.soe.ucsc.edu/node/178>>
- [35] 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*, pp. 2893-2901, 2015
- [36] A. Sebt and H. Hamid, "Body orientation estimation with the ensemble of logistic regression classifiers," *Multimedia Tools and Applications*, pp. 1-17, 2016.
- [37] 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*, pp. 1363-1372, 2016.
- [38] M. W. Khan, "A survey: image segmentation techniques," *International Journal of Future Computer and Communication*, vol. 3, no. 2, p. 89, 2014.
- [39] 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 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1301-1306, 2010.
- [40] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints". In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2288-2295, 2012.
- on *Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 1158-1163, 2005.
- [23] R. Layne, T. M. Hospedales, and S. Gong, "Towards person identification and re-identification with attributes," in *European Conference on Computer Vision*, pp. 402-412, 2012.
- [24] 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 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1-8, 2016.
- [25] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," *Computer Vision-ECCV*, pp. 262-275, 2008.
- [26] R. Satta, G. Fumera, and F. Roli, "Fast person re-identification based on dissimilarity representations," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1838-1848, 2012.
- [27] 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*, pp. 144-151, 2014.
- [28] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 3, pp. 653-668, 2013.
- [29] T. D'Orazio, P. L. Mazzeo, and P. Spagnolo, "Color brightness transfer function evaluation for non overlapping multi camera tracking," in *Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, pp. 1-6, 2009.
- [30] 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*, pp. 2197-2206, 2015.
- [31] 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*, pp. 3908-3916, 2015.
- [32] W. Lin et al., "Learning Correspondence Structures for Person Re-Identification," *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2438-2453, 2017.

¹ Accuracy

² Gaussian mixture model

³ Covariance Descriptor

⁴ Epitome Analysis

⁵ Maximally Stable Color Regions (MSCR)

⁶ Recurrent High-Structured Patches (RHSP)

⁷ Ensemble of Localized Feature (ELF)

⁸ Scale-Invariant Feature Transform

⁹ L*a*b color space

¹⁰ Correspondence structure matrix

¹¹ Probe image

¹² Gallery set

¹³ Upsampling

¹⁴ Deep convolutional neural network

¹⁵ Segmentation

¹⁶ Color space

¹⁷ Local Maximal Occurrence (LOMO)

¹⁸ Crossview Quadratic Discriminant Analysis (XQDA)

¹⁹ Hierarchical Gaussian Descriptor

²⁰ Multiscale

²¹ Local Binary Pattern

²² Scale Invariant Local Ternary Pattern

²³ Quantization

²⁴ Unimodal

²⁵ Tangent space

²⁶ Gaussian of Gaussian (GoG)

²⁷ Close set

²⁸ Cumulative Matching Characteristic (CMC)

²⁹ Histogram of Oriented Gradients (HOG)