

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



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

رساله دکتری

بازشناسایی انسان در سیستم‌های نظارت ویدیویی
با در نظر گرفتن انسداد ناشی از حمل اشیاء

نگارنده: زهرا مرتضایی

استاد راهنما

دکتر حمید حسن پور

استاد مشاور

Prof. Azeddine Beghdadi

اسفند ۱۴۰۰

شماره: ۸۱۹
تاریخ: ۸۵۹
ویرایش:

باسمه تعالی



مدیریت تحصیلات تکمیلی

پیوست شماره ۲

دانشکده:

گروه:

رساله دکتری خانم زهرا مرتضایی
تحت عنوان: بازشناسایی انسان در سیستم‌های نظارت ویدیویی با در نظر گرفتن انسداد ناشی از حمل اشیاء

در تاریخ ۱۴۰۰/۱۲/۳ 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 عزالدین بغدادی		نام و نام خانوادگی: دکتر حمید حسن پور
	نام و نام خانوادگی:		نام و نام خانوادگی:

امضاء	نماینده تحصیلات تکمیلی	امضاء	اساتید داور
	نام و نام خانوادگی: دکتر علیرضا تجوی		نام و نام خانوادگی: دکتر منصور فاتح
			نام و نام خانوادگی: دکتر علیرضا احمدی فرد
			نام و نام خانوادگی: دکتر فرزین یغمایی

تقدیم به خدایی که آفرید

جهان را، انسان را، عقل را، علم را، معرفت را، عشق را

و تقدیم به کسانی که عشقشان را در وجودم دمید؛

مهربان فرشتگانی که لحظات ناب باور بودن، لذت و غرور دانستن، جسارت خواستن، عظمت رسیدن و تمام تجربه‌های یکتا و زیبای زندگی‌م، مدیون حضور سبز
آنهاست؛

پدر و مادر عزیز و مهربانم که در سختی‌ها و دشواری‌های زندگی، همواره یاور و دل‌سوز و خداکار و پشتیبان محکم و مطمئن برایم بوده‌اند؛

برادر و خواهرم که وجودشان شادی، بخش و صفایشان مایه آرامش من است.

پاس خدای را که سخوران، در ستودن او بماند و شامندگان، شردن نعمت های او ندانند و کوشندگان، حق او را گردن نتوانند. و سلام و دور بر محمد و خاندان پاک او
طاهران مصوم، هم آنان که وجودان و لیدار وجودشان است؛ و نفرین پیوسته بردشمنان ایشان تا روز تاختنیز...

بدون شک جایگاه و منزلت معلم، آبل از آن است که در مقام قدردانی از زحمات بی سائبه او، بازبان قاصر و دست ناتوان، چیزی بنجاریم.

اما از آنجایی که تجلیل از معلم، پاس از انسانی است که هدف و غایت آفرینش را تائین می کند و سلامت امانت های را که به دستش سپرده اند، تقصین؛ بر حسب
و خیفه و از باب «من لم یسکر المنعم من المخلوقین لم یسکر الله عزوجل»:

از پدر و مادر عزیزم... این دو معلم بزرگوارم... که همواره بر کوتاهی و درستی من، قلم عنو کشیده و کرمیانه از کنار غفلت هایم گذشته اند و در تمام عرصه های زندگی یار و
یاوری بی چشم داشت برای من بوده اند؛

از استاد با کالات و شایسته؛ جناب آقای دکتر حسن پور که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ گلی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این
رساله را بر عهده گرفتند؛

کمال شکر و قدردانی را دارم.

پنچین لازم می دانم از مسئول محترم آموزش دانشکده سرکار خانم سرایبی و از دوست و بهکلاسی مهربانم سرکار خانم جاوید و خانواده محترم آقای مهندس حامد
عبداللهی به خاطر مساعدت ها و بهرایی هایشان قدردانی نمایم.

باشد که این خردترین، بخشی از زحمات آنان را پاس گوید.

تعهد نامه

اینجانب **زهرا مرتضائی** دانشجوی دوره دکتری رشته هوش مصنوعی دانشکده کامپیوتر و فناوری اطلاعات دانشگاه صنعتی شاهرود نویسنده پایان نامه بازناسایی انسان در سیستم‌های نظارت ویدیویی با در نظر گرفتن انسداد ناشی از حمل اشیاء تحت راهنمایی **آقای دکتر حمید حسن پور** متعهد می‌شوم.

- تحقیقات در این پایان نامه توسط اینجانب 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 پایان نامه 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 است 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 پایان نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

امضای دانشجو

تاریخ

مالکیت نتایج و حق نشر

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

چکیده

از مراحل مهم و تاثیرگذار در عملکرد سیستم‌های نظارتی، مرحله بازشناسایی افراد است. یکی از فرض‌های اولیه در بیشتر سیستم‌های بازشناسایی این است که ویژگی‌های افراد حین فرآیند ردیابی و با گذشت زمان تغییر نمی‌کند. این در حالی است که عواملی مانند تغییر زاویه افراد نسبت به دوربین، تغییر در روشنایی محیط، انسداد تمام یا بخشی از بدن افراد به دلیل حمل اشیاء و پس‌زمینه‌های شلوغ، منجر به ایجاد برخی تغییرات در ویژگی‌های ظاهری افراد می‌شوند. از این‌رو، فرض عدم تغییر ویژگی‌های ظاهری افراد در سیستم‌های بازشناسایی منجر به کاهش دقت این سیستم‌ها می‌شود.

در این پژوهش با در نظر گرفتن مسئله انسداد ناشی از حمل اشیاء و پس‌زمینه‌های شلوغ، عملکرد سیستم‌های بازشناسایی انسان بهبود می‌یابد. در این راستا، راهکارهایی برای بهبود عملکرد توصیفگرهای مبتنی بر ویژگی‌های ظاهری و در نتیجه بهبود عملکرد سیستم‌های بازشناسایی انسان، ارائه می‌شود. در تمامی راهکارهای پیشنهادی تاثیر پیکسل‌های تصویر بر بردار ویژگی استخراجی توسط توصیفگرها، متناسب با تعلق پیکسل‌ها به پس‌زمینه و پیش‌زمینه و با توجه به میزان اهمیت آنها در مسئله بازشناسایی تعیین می‌شود. در راهکار اول، با استفاده از توزیع گوسین و تابع چگالی احتمال، روشی برای قطعه‌بندی معنایی تصویر ارائه می‌شود که در آن پیکسل‌ها با اهمیت یکسان در یک قطعه قرار می‌گیرند. سپس از تصویر قطعه‌بندی شده به‌عنوان نقشه وزن برای ارزش‌گذاری ویژگی‌های استخراجی از پیکسل‌های تصویر استفاده می‌شود. به‌منظور افزایش دقت قطعه‌بندی معنایی ارائه شده، در راهکار دوم ابتدا تصویر ورودی با استفاده از روش قطعه‌بندی مبتنی بر شبکه عمیق DeepLabv3+، به‌صورت معنایی به سه ناحیه بدن فرد، اشیاء و پس‌زمینه قطعه‌بندی می‌شود. سپس پیکسل‌های مرتبط با هر ناحیه با توجه به ارزش آن در مسئله بازشناسایی، به‌صورت دستی وزن‌دهی می‌شوند. برای اجتناب از وزن‌دهی دستی پیکسل‌های تصویر، ویژگی‌های استخراجی از پیکسل‌های مرتبط با اشیاء در حال حمل و پس‌زمینه، به‌طور خودکار با توجه به اندازه هر یک از نواحی و فاصله پیکسل‌های مربوطه تا مرکز تصویر، در راهکار سوم وزن‌دهی می‌شود. در ادامه با توجه به اهمیت بازیابی نواحی مسدود بدن در بهبود عملکرد سیستم‌های بازشناسایی، در راهکار چهارم، فرآیندی به نام فرآیند یکسان‌سازی¹، برای بازیابی پیکسل‌های نواحی مسدود با توجه به پیکسل‌های همسایه (نواحی غیرمسدود) ارائه می‌شود. پس از انجام عمل یکسان‌سازی روی تصویر ورودی، وزن ویژگی‌های استخراجی از پیکسل‌های مرتبط با اشیاء در حال حمل، نواحی یکسان‌سازی شده و پس‌زمینه، به‌طور خودکار با توجه به فاصله پیکسل‌های مربوطه تا مرکز تصویر، تنظیم می‌شود. ارزیابی و مقایسه نتایج حاصل از سیستم‌های بازشناسایی پیشنهادی روی پایگاه‌های داده VIPeR، PRID450s، CUHK01 و CUHK03، نشان‌دهنده دقت بالاتر سیستم بازشناسایی پیشنهادی نسبت به سیستم‌های بازشناسایی مورد مقایسه است.

کلمات کلیدی: سیستم نظارتی، سیستم بازشناسایی، انسداد ناشی از حمل اشیاء، پس‌زمینه‌های شلوغ.

¹ Unification process

لیست مقالات مستخرج از رساله

- [1] Zahra Mortezaie, Hamid Hassanpour, Azeddine Beghdadi, People Re-Identification under Occlusion and Crowded Background, *Multimedia Tools and Applications*, pp. 1-21, (2022), doi:10.1007/s11042-021-11868-y.
- [2] Zahra Mortezaie, Hamid Hassanpour, Azeddine Beghdadi, Sensing Image Regions for Enhancing Accuracy in People Re-Identification, *Iranian (Iranica) Journal of Energy & Environment*, 13(3), pp. 295-304, (2022), doi:10.5829/IJEE.2022.13.03.09.
- [3] Zahra Mortezaie, Hamid Hassanpour, Azeddine Beghdadi, A Color-Based Re-Ranking Process for People Re-Identification, *9th European Workshop on Visual Information Processing (EUVIP)*, pp. 1-5, (2021), doi:10.1109/EUVIP50544.2021.9484056.
- [4] Zahra Mortezaie, Hamid Hassanpour, A Survey on People Re-Identification Approaches Considering Occlusion, *Majlesi Journal of Electrical Engineering (MJEE)*, Accepted on Jun 28, 2022.
- [5] Zahra Mortezaie, Hamid Hassanpour, Azeddine Beghdadi, Re-Identification in Video Surveillance Systems Considering Appearance Changes, *International Journal of Science and Technology (SCIENTIA IRANICA)*, Submitted in the revised form on May 31, 2022.
- [6] Zahra Mortezaie, Hamid Hassanpour, Azeddine Beghdadi, People Re-Identification Using Part-Based Weighing Mechanisms, *Iranian Journal of Electrical and Electronic Engineering (IJEET)*, Submitted on Jan 19, 2022.

فهرست مطالب

شماره صفحه

عناوین

فصل اول: بازشناسایی انسان در سیستم‌های نظارتی

- ۱-۱- مقدمه ۲
- ۲-۱- سیستم بازشناسایی انسان ۳
- ۳-۱- مشکلات موجود در زمینه بازشناسایی انسان ۵
- ۴-۱- محدودیت‌های موجود در سیستم‌های بازشناسایی انسان ۸
- ۵-۱- پایگاه‌های داده موجود در زمینه بازشناسایی انسان ۹
- ۶-۱- اهداف این پژوهش ۲۱
- ۷-۱- نوآوری‌ها و دستاوردهای رساله ۲۲
- ۸-۱- فرض‌های اولیه سیستم پیشنهادی ۲۴
- ۹-۱- جمع‌بندی ۲۵

فصل دوم: مروری بر کارهای پیشین

- ۱-۲- مقدمه ۲۸
- ۲-۲- بازشناسایی انسان بدون توجه به انسداد بدن و پس‌زمینه شلوغ ۲۹
- ۳-۲- بازشناسایی انسان با توجه به انسداد و پس‌زمینه شلوغ ۴۲
- ۴-۲- جمع‌بندی ۴۷

فصل سوم: روش پیشنهادی

- ۱-۳- مقدمه ۵۰
- ۲-۳- سیستم بازشناسایی پیشنهادی ۵۰
- ۱-۲-۳- توصیفگر گوسین گوسین و گوسین سلسله‌مراتبی ۵۲
- ۲-۲-۳- بهبود عملکرد سیستم بازشناسایی ۵۵
- ۱-۲-۲-۳- وزن‌دهی ویژگی‌ها با استفاده از توزیع گوسین و تابع چگالی احتمال ۵۶
- ۲-۲-۲-۳- وزن‌دهی ویژگی‌ها با توجه به قطعه‌بندی معنایی ۶۰
- ۳-۲-۲-۳- وزن‌دهی ویژگی‌ها با توجه به اندازه نواحی و فاصله پیکسل‌ها تا مرکز تصویر ۶۳

۶۷-۳-۲-۲-۴- وزن دهی ویژگی‌ها با توجه به نواحی بازبایی شده.....

۷۴-۳-۳- جمع‌بندی.....

فصل چهارم: ارزیابی راهکارهای پیشنهادی و بررسی نتایج

۷۸-۴-۱- مقدمه.....

۷۸-۴-۲- معیار ارزیابی مورد استفاده.....

۷۹-۴-۳- پایگاه‌های داده مورد استفاده.....

۷۹-۴-۴- نتایج حاصل از راهکارهای پیشنهادی.....

۸۰-۴-۴-۱- نتایج حاصل از راهکار پیشنهادی اول و دوم.....

۸۸-۴-۴-۲- نتایج حاصل از راهکار پیشنهادی سوم و چهارم.....

۹۷-۴-۵- بررسی سربار محاسباتی راهکارهای پیشنهادی بر سیستم‌های بازناسایی موجود.....

۹۸-۴-۶- جمع‌بندی.....

فصل پنجم: جمع‌بندی و راهکارهای پیشنهادی برای ادامه کار

۱۰۰-۵-۱- جمع‌بندی.....

۱۰۱-۵-۲- راهکارهای پیشنهادی برای ادامه کار.....

۱۰۵-مراجع.....

فهرست شکل‌ها

عناوین	شماره صفحه
شکل (۱-۱): تصویری از پس‌زمینه یک فرد مقابل دوربین‌های متفاوت از پایگاه داده VIPeR.....	۶
شکل (۲-۱): تصویری از پوشش ظاهری یک فرد در زاویه دیدهای متفاوت از پایگاه داده GRID.....	۷
شکل (۳-۱): تصویر مربوط به یک فرد با پوشش ظاهری ثابت، در دو محیط با شدت روشنایی متفاوت از پایگاه داده 3DPeS.....	۷
شکل (۴-۱): نمونه تصویر افراد مسدود شده به دلایل مختلف.....	۸
شکل (۵-۱): نمونه تصاویری از پایگاه داده VIPeR.....	۱۰
شکل (۶-۱): نمونه تصاویری از پایگاه داده GRID.....	۱۱
شکل (۷-۱): نمونه تصاویری از پایگاه داده 3DpeS.....	۱۱
شکل (۸-۱): نمونه تصاویری از پایگاه داده WARD.....	۱۲
شکل (۹-۱): نمونه تصاویری از پایگاه داده CUHK01.....	۱۲
شکل (۱۰-۱): نمونه تصاویری از پایگاه داده CUHK02.....	۱۲
شکل (۱۱-۱): نمونه تصاویری از پایگاه داده CUHK03.....	۱۳
شکل (۱۲-۱): نمونه تصاویری از پایگاه داده PRID450s.....	۱۳
شکل (۱۳-۱): نمونه تصاویری از پایگاه داده RAiD.....	۱۴
شکل (۱۴-۱): نمونه تصاویری از پایگاه داده HAD.....	۱۵
شکل (۱۵-۱): نمونه تصاویری از پایگاه داده Marklet-1501.....	۱۵
شکل (۱۶-۱): نمونه تصاویری از پایگاه داده PRW.....	۱۶
شکل (۱۷-۱): نمونه تصاویری از پایگاه داده Large Scale Person Search.....	۱۷
شکل (۱۸-۱): نمونه تصاویری از پایگاه داده MARS.....	۱۷

- شکل (۱-۱۹): نمونه تصاویری از پایگاه داده DukeMTMC4ReID..... ۱۸
- شکل (۱-۲۰): نمونه تصاویری از پایگاه داده Airport..... ۱۹
- شکل (۱-۲۱): نمونه تصاویری از پایگاه داده MSMT17..... ۲۰
- شکل (۳-۱): یک نمونه تصویر ورودی، تصویر W_{LAB} و نقشه وزن (\hat{W}) ۵۸
- شکل (۳-۲): نمونه تصاویر اخذ شده توسط دوربین (الف) و (ب) و نقشه‌های وزن حاصل از روش پیشنهادی (\hat{W}) ۵۹
- شکل (۳-۳): دو نمونه تصویر به همراه نقشه‌های وزن پیشنهادی اول (\hat{W}) ۶۰
- شکل (۳-۴): نمونه‌هایی از تصاویر پایگاه‌های داده VIPeR و CUHK01 به همراه نقشه‌های قطعه‌بندی معنایی حاصل از شبکه عمیق DeepLabv3+..... ۶۱
- شکل (۳-۵): تعدادی نمونه تصویر به همراه نقشه‌های وزن مربوطه (W_s) ۶۴
- شکل (۳-۶): تعدادی نمونه تصویر به همراه نقشه‌های وزن مربوطه (W_d) ۶۶
- شکل (۳-۷): تعدادی نمونه تصویر به همراه تصویر قطعه‌بندی شده دستی (I_s) ، تصاویر حاصل از فرآیند یکسان‌سازی (I_u) و نقشه وزن پیشنهادی (W) ۷۰
- شکل (۴-۱): تعدادی نمونه تصویر از پایگاه داده PRID450s..... ۹۳

فهرست جدول‌ها

عناوین	شماره صفحه
جدول (۱-۱): تعدادی از پایگاه‌های داده مورد استفاده در زمینه بازشناسایی انسان.....	۲۱
جدول (۱-۴): نتایج حاصل از بازشناسایی انسان توسط توصیفگرهای GoG و HGD قبل و بعد از اعمال نقشه وزن پیشنهادی در راهکارهای اول و دوم.....	۸۲
جدول (۲-۴): مقایسه عملکرد سیستم بازشناسایی پیشنهادی با برخی از سیستم‌های بازشناسایی موجود، با استفاده از پایگاه داده (Labeled) CUHK03.....	۸۴
جدول (۳-۴): مقایسه عملکرد سیستم‌های بازشناسایی پیشنهادی اول و دوم با برخی از سیستم‌های بازشناسایی موجود، با استفاده از پایگاه داده VIPeR.....	۸۴
جدول (۴-۴): مقایسه عملکرد سیستم‌های بازشناسایی پیشنهادی اول و دوم با برخی از سیستم‌های بازشناسایی موجود، با استفاده از پایگاه داده CUHK01(M=2).....	۸۵
جدول (۵-۴): مقایسه عملکرد سیستم‌های بازشناسایی پیشنهادی اول و دوم با برخی از سیستم‌های بازشناسایی موجود، با استفاده از پایگاه داده PRID450s.....	۸۵
جدول (۶-۴): نتایج حاصل از توصیفگرهای GoG و HGD قبل و بعد از اعمال نقشه وزن پیشنهادی سوم.....	۹۰
جدول (۷-۴): نتایج حاصل از توصیفگرهای GoG و HGD قبل و بعد از اعمال راهکار پیشنهادی چهارم.....	۹۲
جدول (۸-۴): مقایسه عملکرد سیستم‌های بازشناسایی پیشنهادی سوم و چهارم با برخی از سیستم‌های بازشناسایی موجود، با استفاده از پایگاه‌های داده VIPeR و PRID450s.....	۹۴

فهرست اختصارات

A		LBP..... Local Binary Pattern
ADGC.....Adaptive Direction Graph Convolutional		LSLDG.....Least-Squares Log-Density Gradient
B		M
BED..... Bounded Exponential Distance		MARS.....Motion Analysis and Re-identification Set
BCN..... Background_bias Constraint Net		MLGD..... Multi-Level Gaussian Descriptor
C		N
CNN..... Convolutional Neural Network		nRnG.....Normalized Red-normalized Green
CAN.....Comparative Attention Component		O
CGEA.....Cross-Graph Embedded Alignment		OCSVM.....One-Class Support Vector Machine
CGAN.....Conditional Generative adversarial networks		P
D		PeS..... People Surveillance dataset
DMP..... Deformable Part Models		PRID450s.....Person Re-Identification 450s
DukeMTMC.....Duke Multi-Target Multi-Camera		PRW..... Person Re-identification in the Wild
DNN..... Deep Decompositional Network		PAN..... Pedestrian Alignment Network
FPSN.....Foreground Partial Segmentation Net		PCA..... Principle Component Analysis
G		PHOG..... Pyramid Histogram of Oriented Gradients
GRID.....QMUL underGround Re-Identification		PDF..... Probability Density Function
GCT.....Graph Correspondence Transfer		R
GoG..... Gaussian of Gaussian		RAiD...Re-identification Across indoor-outdoor Dataset
GLBN.....Global- Local Background_bias Net		RGB.....Red-Green-Blue
GASN..... Global Aligned Supervision Net		S
H		SVD.....Singular Vector Decomposition
HSV..... Hue-Saturation-Value		SILTP.....Scale Invariant Local Ternary Pattern
HGD..... Hierarchical Gaussian Descriptor		SIFT..... Scale Invariant Feature Transform
I		SPC..... Spatial-Patch Contrastive
IALM.....Inexact Augmented Lagrange Multiplier		SRNet.....Shrinking and Reweighting Network
IGoG..... Improved GoG		SPD..... Symmetric Positive Definite
IHGD..... Improved HGD		SSM..... Semantic Segmented Map
K		V
KNN.....K-Nearest Neighbor		VIPeR.....Viewpoint Invariant Pedestrian Recognition
L		W
LSPS..... Large Scale Person Search		WARD..... Wide Area Re-identification Dataset
LOMO..... Local Maximal Occurrence		X
Lab..... Lightness-a-b		XQDA....Cross-view Quadratic Discriminant Analysis

فصل اول

بازشناسایی انسان در سیستم‌های نظارتی

۱-۱- مقدمه

امروزه به منظور نظارت بر موقعیت و رفتار افراد در مکان‌های عمومی مانند ساختمان‌های اداری و آموزشی، فرودگاه‌ها، ایستگاه‌های قطار و مراکز خرید، از شبکه‌ای از دوربین‌ها استفاده می‌شود [۱ و ۲]. محیط تحت پوشش این دوربین‌ها می‌تواند فضای داخلی یا خارجی یک ساختمان، یا ترکیبی از این دو باشد. داخلی یا خارجی بودن محیط تحت نظارت دوربین‌ها، هر کدام مسائل مختلفی را در سیستم‌های نظارتی ایجاد می‌کنند. از جمله این مسائل می‌توان به تفاوت در میزان روشنایی فضای داخلی و فضای خارجی، وجود عوامل مسدود کننده متفاوت در هر کدام از این محیط‌ها، و ویژگی‌های پس‌زمینه متفاوت اشاره کرد. از این‌رو، نظارت افراد در هر کدام از محیط‌های داخلی و خارجی چالش‌های مخصوص به خود را دارد.

همچنین، همپوشان یا ناهمپوشان بودن محدوده تحت پوشش دوربین‌ها، در سیستم نظارتی مسائل مختلفی را ایجاد می‌کند. نظارت و شناسایی افراد در سیستم‌های نظارتی با دوربین‌های همپوشان، نسبت به سیستم‌های نظارتی با دوربین‌های ناهمپوشان راحت‌تر انجام می‌شود، زیرا در سیستم‌های نظارتی با دوربین‌های همپوشان، از اطلاعات موجود در نواحی همپوشان می‌توان استفاده کرد تا بدون گم کردن افراد، بر آنها نظارت صورت گیرد. لازم به ذکر است که در سیستم‌های نظارتی معمولاً محدوده تحت پوشش دوربین‌ها با هم همپوشانی ندارد [۳]. دلیل این امر این است که با حداقل دوربین‌های ممکن، بتوان محدوده وسیعی از محیط مورد نظر را پوشش داد.

برای نظارت مؤثر بر محیط، هر یک از دوربین‌های نظارتی باید به‌طور مداوم به جمع‌آوری اطلاعات تصویری از محدوده تحت پوشش خود بپردازند. از این‌رو، نظارت بر محیط مورد نظر نیازمند بررسی حجم وسیعی از اطلاعات است [۴]. در این شرایط، نظارت انسانی بر موقعیت و رفتار افراد کاری زمان‌بر و پرهزینه بوده و فواید نظارت بر محیط را تحت تاثیر قرار می‌دهد. بنابراین به‌منظور افزایش سرعت و بهبود عملکرد سیستم‌های نظارتی، بهتر است نظارت بر موقعیت و رفتار افراد در این سیستم‌ها به‌صورت خودکار انجام شود [۵]. لذا با استفاده از سیستم‌های نظارت ویدیویی خودکار در زمینه‌های مختلفی

مانند سیستم‌های امنیتی [۶]، ساختمان‌های هوشمند [۷] و مراقبت از افراد سالمند [۸]، می‌توان نظارت بر موقعیت و رفتار افراد را به‌طور مؤثرتری انجام داد.

یکی از نکات مهم در سیستم‌های نظارتی خودکار، بازشناسایی صحیح انسان در دوربین‌های مختلف است. در واقع بازشناسایی افراد به‌عنوان یکی از مؤثرترین اجزاء در سیستم‌های نظارتی خودکار و پیش‌نیازی ضروری جهت پردازش‌های سطح بالاتر از قبیل شناسایی و تحلیل فعالیت و رفتار افراد، رویدادها و تجزیه و تحلیل صحنه است [۹]. بنابراین، بازشناسایی صحیح افراد منجر به بهبود پردازش‌های سطح بالاتر و در نهایت بهبود عملکرد سیستم‌های نظارتی خودکار می‌شود.

با توجه به اهمیت و تاثیر سیستم‌های بازشناسایی در عملکرد سیستم‌های نظارتی خودکار، در این پژوهش به ارائه راهکاری برای بهبود عملکرد سیستم‌های بازشناسایی می‌پردازیم. از این‌رو، در ادامه این فصل، ابتدا سیستم بازشناسایی انسان به‌صورت اجمالی بررسی می‌شود. سپس به بررسی برخی از مشکلات موجود در زمینه بازشناسایی انسان پرداخته می‌شود. همچنین با توجه به اهداف این پژوهش، در این فصل تعدادی از پایگاه‌های داده رایج در زمینه بازشناسایی انسان معرفی می‌شود. سپس برخی از فرض‌های محدود کننده که بسیاری از سیستم‌های بازشناسایی برای ساده‌سازی مسئله بازشناسایی در نظر می‌گیرند، به‌صورت اجمالی مورد بررسی قرار می‌گیرد. در نهایت به ارائه برخی از اهداف این پژوهش، فرض‌های اولیه مورد نظر در سیستم پیشنهادی و همچنین جمع‌بندی مطالب عنوان شده در این فصل می‌پردازیم.

۱-۲- سیستم بازشناسایی انسان

بازشناسایی انسان به معنی مشاهده یک فرد مقابل یک دوربین و شناسایی مجدد آن بعد از گذشت مدت زمانی اندک مقابل همان دوربین یا سایر دوربین‌های متصل به شبکه است [۱۰]. در واقع سیستم بازشناسایی مشابه سیستم شناسایی است. این سیستم شامل مجموعه‌ای از تصاویر افراد شناخته شده

به نام مجموعه گالری^۱ و مجموعه‌ای از تصاویر افراد ناشناخته و تازه وارد به زاویه دید یک دوربین، به نام مجموعه آزمون^۲ است. البته باید به این نکته توجه داشت که در دنیای واقعی تعداد افراد موجود در مجموعه‌های گالری و آزمون در طول زمان متغیر است. اما معمولاً برای ساده‌سازی مسئله بازشناسایی، در یک بازه زمانی خاص تعداد افراد موجود در این دو مجموعه ثابت در نظر گرفته می‌شود.

فرض کنید در سیستم بازشناسایی انسان مجموعه گالری به صورت رابطه (۱-۱) نشان داده شود:

$$G = \{g_1, g_2, \dots, g_N\} \quad (1-1)$$

به طوری که g_i نشان‌دهنده فرد شناخته شده i ام است. همچنین برچسب‌های مربوط به تصاویر افراد شناخته شده توسط رابطه (۲-۱) بیان می‌شود:

$$id(G) = \{id(g_1), id(g_2), \dots, id(g_N)\} \quad (2-1)$$

به طوری که $id(G)$ برچسب تصاویر موجود در مجموعه گالری (G) را مشخص می‌کند. همچنین فرض کنید که مجموعه تصاویر آزمون به صورت رابطه (۳-۱) تعریف شود:

$$P = \{p_1, p_2, \dots, p_m\} \quad (3-1)$$

در سیستم‌های بازشناسایی انسان، هدف تعیین برچسب اعضای مجموعه آزمون یا به عبارت دیگر تعیین $id(P)$ است. این کار به وسیله بررسی میزان شباهت تصویر هر یک از افراد ناشناخته (مجموعه آزمون) با تصاویر همه افراد شناخته شده (مجموعه گالری)، به کمک رابطه (۴-۱) انجام می‌شود:

$$sim(p, g_i) = similarity(p, g_i), \quad i = 1, 2, \dots, N$$

$$id(p) = \begin{cases} id(g_i) & argmax_i(sim(p, g_i)), \quad s.t \quad sim(p, g_i) > T, \quad i = 1, 2, \dots, N \\ id^*(p) & o.w \end{cases} \quad (4-1)$$

در واقع با استفاده از یک معیار شباهت ($similarity(p, g_i)$)، میزان شباهت بین هر یک از اعضای مجموعه آزمون و همه اعضای مجموعه گالری محاسبه می‌شود. در صورتی که شباهت بدست آمده ($sim(p, g_i)$) از حد آستانه مشخص (T) بیشتر باشد، برچسب شبیه‌ترین عضو مجموعه گالری به عنوان

¹ Gallery Set

² Probe Set

برچسب فرد ناشناخته، در نظر گرفته می‌شود. در غیر این صورت، فرد ناشناخته، به‌عنوان فرد تازه وارد در زاویه دوربین‌ها در نظر گرفته شده و برچسب جدیدی ($id^*(p)$) به آن اختصاص می‌یابد.

۱-۳- مشکلات موجود در زمینه بازشناسایی انسان

همان‌طور که در بخش قبل اشاره شد، در یک سیستم بازشناسایی انسان، باید میزان شباهت بین تصویر آزمون و تمامی تصاویر مجموعه گالری، بر اساس یک معیار شباهت، محاسبه شود. نکته بسیار مهم در بحث بازشناسایی انسان، استفاده از ویژگی‌های مناسب است، تا مبنایی برای مقایسه و تعیین میزان شباهت بین افراد باشد. در واقع یکی از مؤلفه‌های بسیار تاثیرگذار در عملکرد سیستم‌های بازشناسایی انسان، استفاده از ویژگی‌های مناسب برای بررسی میزان شباهت بین افراد است.

برخی از سیستم‌های بازشناسایی، برای بررسی و مقایسه میزان شباهت افراد، از ویژگی‌های بیومتریک انسان‌ها مانند الگوی راه رفتن [۱۱] یا الگوی چهره [۱۲] افراد استفاده می‌کنند. این سیستم‌ها مشکلات و محدودیت‌هایی دارند. به‌عنوان مثال، استفاده از الگوی راه رفتن افراد در سیستم بازشناسایی انسان، نیازمند نگهداری فریم‌های بسیار زیادی از راه رفتن افراد است تا بتوان با استفاده از آنها الگوی راه رفتن هر یک از افراد را شناسایی کرد. همچنین در صورتی که سیستم بازشناسایی انسان بخواهد از الگوهای چهره افراد برای بازشناسایی آنها استفاده کند، نیازمند تصاویر با کیفیت بالا خواهد بود. این در حالی است که در سیستم‌های بازشناسایی انسان، نگهداری و بررسی تعداد زیادی از فریم‌ها به ازای هر فرد و همچنین استفاده از دوربین‌ها با کیفیت بالا، مقرون به صرفه نیست.

همچنین در سیستم‌هایی که برای بازشناسایی افراد از ویژگی‌های بیومتریک دیگر مانند عنبیه چشم استفاده می‌کنند [۱۳]، نیازمند مشارکت و همکاری تمامی افراد موجود در محیط مورد نظر هستند و در واقع به نوعی جزء سیستم‌های بازشناسایی تهاجمی محسوب می‌شوند. بنابراین، استفاده از سیستم‌های بازشناسایی که در آنها از ویژگی‌های ظاهری افراد استفاده می‌شود، امری مطلوب‌تر خواهد بود.

از این‌رو، بسیاری از سیستم‌های بازشناسایی انسان، از ویژگی‌های ظاهری افراد مانند رنگ و طرح لباس استفاده می‌کنند. استفاده از این‌گونه ویژگی‌ها هم باعث ایجاد مشکلات و محدودیت‌هایی در سیستم‌های

بازشناسایی می‌شود. در این سیستم‌ها حرکت افراد مقابل یک دوربین یا بین دوربین‌های مختلف ممکن است باعث ایجاد تغییراتی در ظاهر افراد شود. در این حالت سیستم بازشناسایی عملکرد مناسبی نخواهد داشت.

در ادامه به بررسی برخی از عوامل تغییردهنده ظاهر افراد در سیستم‌های بازشناسایی می‌پردازیم:

● تغییرات پس‌زمینه

در برخی از سیستم‌های بازشناسایی انسان که در آنها ویژگی‌های ظاهری از کل تصویر ورودی استخراج می‌شود، وجود پس‌زمینه‌های شلوغ و تغییر در پس‌زمینه افراد منجر به کاهش عملکرد سیستم می‌شود [۱۴]. در واقع با حرکت افراد مقابل یک دوربین یا بین دوربین‌های مختلف ممکن است فضای پس‌زمینه دچار تغییرات اساسی شود. در چنین حالتی استفاده از ویژگی‌های ظاهری موجود در فضای پس‌زمینه، عملکرد سیستم بازشناسایی را مختل می‌کند. در شکل (۱-۱) نمونه تصویر مربوط به یک فرد با دو پس‌زمینه متفاوت نشان داده شده است.



شکل (۱-۱): تصویری از پس‌زمینه یک فرد مقابل دوربین‌های متفاوت از پایگاه داده VIPeR [۱۵]

● تغییرات زاویه دید

حرکت افراد مقابل دوربین یا بین دوربین‌های مختلف، ممکن است منجر به تغییر زاویه افراد نسبت به دوربین‌های مختلف شود. در سیستم‌های بازشناسایی که از ویژگی‌های ظاهری افراد مانند رنگ و طرح لباس استفاده می‌کنند، تغییر زاویه فرد نسبت به دوربین ممکن است به صورت ناخواسته منجر به تغییر در رنگ و طرح لباس افراد شود. بنابراین، تغییر زاویه افراد نسبت به دوربین، یکی از عواملی است که باعث کاهش عملکرد سیستم‌های بازشناسایی انسان می‌شود [۱۶]. در شکل (۲-۱) نمونه تصویری از پوشش ظاهری یک فرد در زاویه دیدهای متفاوت نشان داده شده است.



شکل (۲-۱): تصویری از پوشش ظاهری یک فرد در زاویه دیدهای متفاوت از پایگاه داده GRID [۱۷]

● تغییرات روشنایی محیط

شدت نور در مکان نصب دوربین‌های نظارتی مختلف ممکن است متفاوت باشد. بنابراین ممکن است، حرکت افراد بین دوربین‌های مختلف، باعث تغییر در شدت روشنایی تصاویر افراد شود [۱۸]. این امر منجر به کاهش دقت عملکرد سیستم‌هایی می‌شود که در آنها برای بازشناسایی افراد از ویژگی‌هایی مانند هیستوگرام رنگ لباس استفاده می‌کنند. در شکل (۳-۱) نمونه تصویر مربوط به یک فرد با پوشش ظاهری ثابت، در دو محیط با شدت روشنایی متفاوت نشان داده شده است.



شکل (۳-۱): تصویر مربوط به یک فرد با پوشش ظاهری ثابت، در دو محیط با شدت روشنایی متفاوت از پایگاه داده 3DPeS [۱۹]

● انسداد

یکی دیگر از مشکلات موجود در سیستم‌های بازشناسایی انسان، انسداد تمام یا بخشی از بدن افراد موجود در محیط تحت نظارت است [۲۰]. این امر ممکن است در اثر حمل یک وسیله خاص توسط افراد، همپوشانی توسط بخشی از ساختار محیط، یا در محیط‌های شلوغ به دلیل همپوشانی برخی از افراد توسط برخی دیگر، رخ دهد. از این رو، در سیستم‌های بازشناسایی که بر مبنای ویژگی‌های ظاهری افراد عمل می‌کنند، ممکن است افراد به درستی بازشناسایی نشوند. در شکل (۴-۱)، نمونه تصویر مربوط به افرادی که به دلایل مختلف دچار انسداد شده‌اند، نشان داده شده است.



شکل (۱-۴): نمونه تصویر افراد مسدود شده به دلایل مختلف

۱-۴- محدودیت‌های موجود در سیستم‌های بازشناسایی انسان

با توجه به مشکلات و محدودیت‌های اشاره شده، در سیستم‌های بازشناسایی موجود، معمولاً برای ساده‌سازی مسئله بازشناسایی یکسری فرض‌های اولیه در نظر گرفته می‌شود. برخی از این فرض‌ها علاوه بر اینکه فضای مسئله را از دنیای واقعی دور می‌کند، بر عملکرد نهایی سیستم‌های بازشناسایی تاثیر گذاشته و باعث کاهش دقت آنها می‌شود. برخی از محدودیت‌های موجود در سیستم‌های بازشناسایی انسان به شرح زیر است:

الف) بسیاری از سیستم‌های بازشناسایی انسان این‌طور فرض می‌کنند که در طول زمان و با حرکت افراد مقابل یک دوربین یا مجموعه‌ای از دوربین‌ها، ویژگی‌های ظاهری افراد تغییر نمی‌کند. همان‌طور که اشاره شد، با حرکت فرد مقابل یک دوربین یا بین دوربین‌ها و در نتیجه آن تغییر زاویه فرد نسبت به دوربین یا تغییرات روشنایی محیط، ویژگی‌های ظاهری فرد معمولاً تغییر می‌کند.

به‌عنوان مثال فرض کنید در یک سیستم بازشناسایی بخواهیم از هیستوگرام رنگ لباس افراد به‌عنوان ویژگی متمایز کننده استفاده کنیم. در این سیستم فرض کنید فردی که مقابل یک دوربین در حرکت است، کتی با رنگ متفاوت از پیراهن خود به تن دارد. در این صورت لباس این فرد وقتی که دقیقاً روبروی دوربین باشد، دارای یک رنگ خاص است و در نتیجه دارای یک هیستوگرام رنگ مشخص خواهد بود. این در حالی است که با حرکت فرد نسبت به دوربین، به‌عنوان مثال وقتی که فرد مورد نظر پشت به دوربین قرار بگیرد، رنگ لباسش تغییر می‌کند. نتیجه این امر این است که هیستوگرام رنگ مربوط به فرد مورد نظر نیز دچار تغییراتی می‌شود. از این‌رو، این سیستم بازشناسایی توانایی خود را در

شناسایی مجدد فرد مذکور از دست می‌دهد. در این صورت ممکن است سیستم بازناسایی به اشتباه برچسب فردی دیگر را به فرد مذکور اختصاص دهد یا اینکه ممکن است این فرد را به‌عنوان یک فرد تازه وارد در نظر گرفته و برچسبی جدید به آن تخصیص دهد.

ب) در بسیاری از سیستم‌های بازناسایی انسان، مسئله حمل اشیاء توسط افراد، در نظر گرفته نمی‌شود. همان‌طور که در این فصل اشاره شد، در صورتی که فردی در حال حمل یک شیء باشد، با حرکت فرد در زاویه دید یک دوربین یا بین دوربین‌های مختلف، این شیء ممکن است منجر به انسداد تمام یا بخشی از بدن فرد شود. نتیجه این امر این است که ویژگی‌های ظاهری استخراج شده از فرد دچار تغییر اساسی خواهد شد. از این‌رو، ممکن است سیستم بازناسایی قادر به شناسایی درست فرد نباشد. این در حالی است که مسئله انسداد ناشی از حمل اشیاء در بسیاری از سیستم‌های بازناسایی موجود رفع نشده است.

ج) بسیاری از سیستم‌های بازناسایی در مرحله استخراج ویژگی، ویژگی‌های ظاهری را از کل تصویر ورودی (شامل پوشش افراد و فضای پس‌زمینه) استخراج می‌کنند. همان‌طور که در ابتدای این فصل بررسی شد، در چنین سیستم‌هایی تغییر در پس‌زمینه تصاویر، منجر به کاهش دقت بازناسایی می‌شود. با توجه به مشکلات و فرض‌های محدود کننده مطرح شده در سیستم‌های بازناسایی انسان، در این پژوهش سعی می‌شود سیستمی برای بازناسایی افراد پیشنهاد شود که در آن علاوه بر توجه به ویژگی‌های ظاهری افراد، به مسئله انسداد ناشی از حمل اشیاء و تغییرات در پس‌زمینه توجه شود.

۱-۵- پایگاه‌های داده موجود در زمینه بازناسایی انسان

به دلیل اهمیت استفاده از سیستم‌های نظارتی، در دهه اخیر مسئله بازناسایی انسان مورد توجه بسیاری از محققان حوزه بینایی ماشین قرار گرفته است. در این زمینه تعداد زیادی پایگاه داده با ویژگی‌ها و مشخصات مختلف، فراهم شده است. از این‌رو، در این بخش به بررسی برخی از پایگاه‌های داده موجود در زمینه بازناسایی انسان می‌پردازیم.

● پایگاه داده VIPeR¹

پایگاه داده VIPeR شامل ۱۲۶۴ تصویر از ۶۳۲ فرد مختلف است. در این پایگاه داده برای هر فرد دو تصویر وجود دارد که توسط دو دوربین با محدوده دید متفاوت تهیه شده است [۱۵]. این پایگاه داده به دلیل دارا بودن داده‌هایی با تغییرات در موقعیت وضعیت ظاهری افراد، تغییر در روشنایی محیط، پس‌زمینه شلوغ، و کیفیت پایین تصاویر یکی از چالش برانگیزترین پایگاه‌های داده است. تعدادی از تصاویر این پایگاه داده در شکل (۵-۱) نشان داده شده است.



شکل (۵-۱): نمونه تصویری از پایگاه داده VIPeR

● پایگاه داده GRID²

تصاویر این پایگاه داده توسط هشت دوربین با محدوده تحت پوشش مجزا، در یک ایستگاه زیر زمینی تهیه شده است. این پایگاه داده شامل تصویری از ۲۵۰ فرد مختلف است که از هر فرد دو تصویر توسط دوربین‌های مختلف تهیه شده است. از این‌رو، ۲۵۰ تصویر از افراد مختلف در مجموعه گالری و ۲۵۰ تصویر دیگر از این افراد در مجموعه آزمون قرار دارد. همچنین این پایگاه داده شامل ۷۷۵ تصویر دیگر نیز هست که به تصاویر مجموعه آزمون متعلق نیستند. از این‌رو، از این تصاویر می‌توان برای ارزیابی متقابل استفاده کرد. همچنین به دلیل تغییرات در وضعیت ظاهری افراد، رنگ لباس، روشنایی محیط و کیفیت پایین تصاویر، کار کردن با این پایگاه داده چالش برانگیز است [۱۷]. در شکل (۶-۱) تعدادی از تصاویر موجود در این پایگاه داده نشان داده شده است.

¹ Viewpoint Invariant Pedestrian Recognition (VIPeR)

² QMUL underGround Re-IDentification (GRID)



شکل (۱-۶): نمونه تصاویری از پایگاه داده GRID

● پایگاه داده 3DPeS¹

تصاویر این پایگاه داده در محیط بیرونی توسط هشت دوربین مجزا طی چند روز تهیه شده است. روشنایی محیط قرارگیری دوربین‌ها تفاوت زیادی ندارد. اما با توجه به اینکه تصاویر تهیه شده از افراد در طول روز، و در مناطق مختلفی با شدت نور زیاد و همچنین سایه، تهیه شده است، در نتیجه آن در بعضی موارد تغییرات شدید نور در تصاویر این پایگاه داده وجود دارد. علاوه بر تصاویر افراد، این پایگاه داده شامل اطلاعاتی درباره تنظیمات دوربین‌ها، ساختار سه بعدی و تعداد زیادی ویدیو از محیط مورد نظر است [۱۹]. در شکل (۱-۷)، نمونه تصاویری از این پایگاه داده نشان داده شده است.



شکل (۱-۷): نمونه تصاویری از پایگاه داده 3DPeS

● پایگاه داده WARD²

تصاویر این پایگاه داده توسط سه دوربین با محدوده تحت پوشش مجزا تهیه شده است. در این پایگاه داده هر یک از دوربین‌ها، چندین تصویر از هر فرد ثبت کرده اند. این پایگاه داده شامل تصاویری از ۷۰ فرد مختلف بوده و در آن از کل این افراد تعداد ۴۷۸۶ تصویر موجود است [۲۱]. در شکل (۱-۸) نمونه‌ای از تصاویر این پایگاه داده نشان داده شده است.

¹ 3D People Surveillance dataset (3DPeS)

² Wide Area Re-identification Dataset (WARD)



شکل (۸-۱): نمونه تصاویری از پایگاه داده WARD

● پایگاه داده CUHK01

تصاویر این پایگاه داده مربوط به ۹۷۱ فرد مختلف است که توسط دو دوربین مجزا تهیه شده است. در این پایگاه داده، به ازای هر فرد، دو تصویر در هر دوربین وجود دارد. از این رو، تعداد کل تصاویر پایگاه داده CUHK01، ۳۸۸۴ تصویر می‌باشد. در شکل (۹-۱) نمونه تصاویر از این پایگاه داده نشان داده شده است [۲۲].



شکل (۹-۱): نمونه تصاویری از پایگاه داده CUHK01

● پایگاه داده CUHK02

این پایگاه داده نسخه توسعه یافته‌ای از پایگاه داده CUHK01 است. تصاویر پایگاه داده CUHK02 توسط ۱۰ دوربین مختلف تهیه شده است. این پایگاه داده شامل ۷۲۶۴ تصویر از ۱۸۱۶ فرد مختلف است. تصاویر این پایگاه داده از کیفیت نسبتاً مناسبی برخوردار هستند [۲۳]. در شکل (۱۰-۱) نمونه‌ای از تصاویر این پایگاه داده نشان داده شده است.



شکل (۱۰-۱): نمونه تصاویری از پایگاه داده CUHK02

● پایگاه داده CUHK03

تصاویر این پایگاه داده مشابه پایگاه داده CUHK02، توسط ۱۰ دوربین تهیه شده است. تعداد تصاویر موجود در این پایگاه داده ۱۳۱۶۴ تصویر مربوط به ۱۳۶۰ فرد مختلف است. در این پایگاه داده به طور میانگین ۴/۸ تصویر از هر فرد در هر نمای دوربین گرفته شده است. از این رو، این پایگاه داده اولین پایگاه داده با شناسایی انسان است که به دلیل تعداد تصاویر مناسب، قابلیت استفاده در شبکه‌های عصبی عمیق را دارد.

لازم به ذکر است که تعیین افراد موجود در تصاویر این پایگاه داده به روش دستی و روش خودکار¹ DMP انجام شده است [۲۴]. در شکل (۱۱-۱) نمونه تصویری از این پایگاه داده نشان داده شده است.



ب) تعیین افراد به روش خودکار



الف) تعیین افراد به روش دستی

شکل (۱۱-۱): نمونه تصویری از پایگاه داده CUHK03

● پایگاه داده PRID450s²

پایگاه داده PRID450s شامل ۹۰۰ تصویر از ۴۵۰ فرد مختلف است که توسط دو دوربین نظارتی ثابت با زاویه دید مجزا اخذ شده اند. در این پایگاه داده پیش‌زمینه و پس‌زمینه تصاویر، به صورت خودکار و با توجه به حرکت افراد، قطع‌بندی شده است. همچنین نواحی مختلف بدن افراد شامل سر، نیم‌تنه، پاها و اشیاء در حال حمل، به صورت دستی برچسب‌گذاری شده است [۲۵]. در شکل (۱۲-۱)، نمونه تصویری از پایگاه داده PRID450s نشان داده شده است.



شکل (۱۲-۱): نمونه تصویری از پایگاه داده PRID450s

¹ Deformable Part Models (DMP)

² Person Re-Identification 450s (PRID450s)

● پایگاه داده RAiD¹

تصاویر این پایگاه داده توسط چهار دوربین با محدوده تحت پوشش مجزا تهیه شده است. برای تهیه تصاویر این پایگاه داده، دو دوربین در محیط بیرونی و تعداد دو دوربین در محیط داخلی یک ساختمان قرار داده شده است. تعداد کل تصاویر موجود در این پایگاه داده ۶۹۲۰ مربوط به ۴۳ فرد مختلف است. هر یک از افراد موجود در این پایگاه داده از مقابل تمامی دوربین‌ها عبور کرده و تصاویر آنها ثبت شده است.

با توجه به اینکه دو دوربین در محیط بیرونی و دو دوربین در محیط داخلی تصاویر این پایگاه داده را ثبت کرده اند، تغییرات شدت روشنایی در تصاویر نسبتاً زیاد است [۲۶]. در شکل (۱-۱۳)، نمونه‌ای از تصاویر این پایگاه داده نشان داده شده است.



شکل (۱-۱۳): نمونه تصاویری از پایگاه داده RAiD

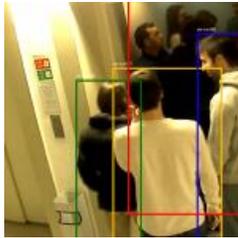
● پایگاه داده HDA

تصاویر این پایگاه داده توسط ۱۳ دوربین در محیط داخلی یک اداره تهیه شده است. این دوربین‌ها در ساعات پر رفت و آمد اداره به مدت ۳۰ دقیقه تصاویری از ۸۵ فرد تهیه کرده اند. کیفیت تصاویر ثبت شده توسط این دوربین‌ها بالا است. لازم به ذکر است که تعیین افراد موجود در تصاویر این پایگاه داده هم به صورت دستی و هم با استفاده از روش ویژگی‌های هرمی^۲ انجام شده است. همچنین در این پایگاه داده تصاویری افرادی که به صورت جزئی انسداد دارند یا در آن بخشی از بدن فرد، خارج از محدوده تصویر است، به عنوان تصاویر مسدود شده برچسب گذاری شده اند. همچنین تصاویری که در آن به دلیل

¹ Re-identification Across indoor-outdoor Dataset (RAiD)

² Pyramid Features (ACF)

از دیاد جمعیت، برچسب گذاری افراد به صورت مجزا امکان پذیر نیست، به عنوان تصاویر شلوغ برچسب گذاری شده اند [۲۷]. در شکل (۱-۱۴)، نمونه ای از تصاویر این پایگاه داده نشان داده شده است.



ج) تصویر شلوغ



ب) تصویر مسدود شده



الف) تصویر بدون انسداد و با برچسب مجزا

شکل (۱-۱۴): نمونه تصاویری از پایگاه داده HAD

● پایگاه داده Marklet-1501

تصاویر این پایگاه داده با استفاده از پنج دوربین با کیفیت تصویربرداری بالا و یک دوربین با کیفیت تصویربرداری پایین، از محیط بیرونی یک فروشگاه جمع آوری شده اند. محدوده تحت پوشش بعضی از این دوربین ها با هم همپوشانی دارد. این پایگاه داده شامل ۳۲۶۶۸ تصویر برچسب گذاری شده از ۱۵۰۱ فرد مختلف است. هر یک از افراد برچسب گذاری شده، ممکن است دارای چندین تصویر از هر دوربین باشند و حداقل مقابل دو دوربین ظاهر شده اند. از این رو این پایگاه داده را می توان در بررسی عملکرد سیستم های بازشناسایی که هدف آنها علاوه بر بازشناسایی افراد مقابل یک دوربین، بازشناسایی افراد متحرک بین دوربین های مختلف نیز هست، استفاده کرد.

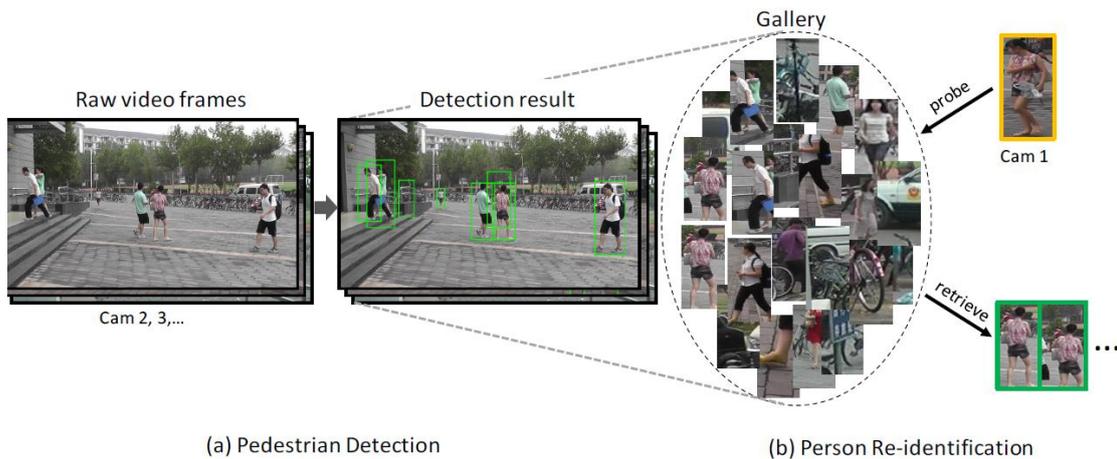
لازم به ذکر است که برای تشخیص افراد موجود در تصاویر این پایگاه داده، از روش DPM استفاده شده است. برای تطابق پایگاه داده با شرایط واقعی، از بین تصاویر افراد تشخیص داده شده توسط روش DPM، افراد موجود در ۲۷۹۳ تصویر به درستی تشخیص داده نشده اند [۲۸]. در شکل (۱-۱۵)، نمونه تصاویری از این پایگاه داده نشان داده شده است.



شکل (۱-۱۵): نمونه تصاویری از پایگاه داده Marklet-1501

● پایگاه داده PRW¹

این پایگاه داده، نسخه توسعه یافته‌ای از پایگاه داده Marklet-1501 است. از بین ۳۴۳۰۴ تصویر موجود در این پایگاه داده، تعداد ۱۱۸۱۶ تصویر از ۹۳۲ فرد مختلف، برچسب‌گذاری شده است. در پایگاه داده Marklet-1501، فضایی که توسط فریم‌های ثبت شده به صورت کامل، در اختیار نیست؛ بلکه افراد موجود در تصاویر، ابتدا توسط روش DPM، تشخیص داده شده و سپس برچسب‌گذاری شده اند. اما در پایگاه داده PRW، تمامی فضای موجود در فریم‌ها، وجود دارد. در واقع در این پایگاه داده، افراد موجود در کل یک تصویر، بدون جداسازی افراد از فضای اطراف، برچسب‌گذاری شده اند. از این رو از پایگاه داده PRW، می‌توان برای ارزیابی عملکرد روش‌های تشخیص افراد موجود در تصویر نیز استفاده کرد [۲۹]. در شکل (۱-۱۶)، تصاویری از این پایگاه داده نشان داده شده است.



شکل (۱-۱۶): نمونه تصاویری از پایگاه داده PRW

● پایگاه داده LSPS²

در این پایگاه داده نیز مشابه پایگاه داده PRW، افراد از تصاویر جدا نشده اند و تمامی فضای موجود در تصاویر در دسترس است. در این پایگاه داده تعداد ۱۸۱۸۴ تصویر از ۸۴۳۲ فرد مختلف و ۹۹۸۰۹ مستطیل برچسب‌گذاری شده محاط بر افراد^۳، وجود دارد. برای واقعی‌تر شدن شرایط این پایگاه داده،

¹ Person Re-identification in the Wild (PRW)

³ Bounding box

² Large Scale Person Search (LSPS)

تصاویر ثبت شده توسط دوربین‌های دستی در مجموعه تصاویر گالری موجود است. همچنین در این پایگاه داده تصاویری با کیفیت پایین و تصاویر دارای انسداد، موجود است [۳۰]. در شکل (۱۷-۱) نمونه تصاویری از این پایگاه داده نشان داده شده است.



شکل (۱۷-۱): نمونه تصاویری از پایگاه داده Large Scale Person Search

● پایگاه داده MARS¹

این پایگاه داده نیز نسخه توسعه یافته‌ای از پایگاه داده Marklet-1501 است. در این پایگاه داده تعداد ۲۰۴۷۸ مسیر حرکت از ۱۲۶۱ فرد مختلف وجود دارد. این پایگاه داده با تعداد کل ۱۱۹۱۰۰۳ تصویر، جزء اولین پایگاه‌های داده ویدیویی با تعداد تصاویر بسیار زیاد محسوب می‌شود [۳۱]. در شکل (۱۸-۱) نمونه‌ای از تصاویر این پایگاه داده نشان داده شده است.



شکل (۱۸-۱): نمونه تصاویری از پایگاه داده MARS

● پایگاه‌های داده DukeMTMC-reID و DukeMTMC4ReID

این دو پایگاه داده، نسخه‌های توسعه یافته‌ای از پایگاه داده DukeMTMC^۲ هستند. پایگاه داده DukeMTMC، پایگاه داده‌ای بسیار بزرگ برای ردیابی چند هدفه در چندین دوربین است. این پایگاه داده شامل ۸۵ دقیقه ویدیویی ثبت شده توسط هشت دوربین از بیش از ۲۷۰۰ فرد مختلف است [۳۲].

¹ Motion Analysis and Re-identification Set (MARS)

² Duke Multi-Target Multi-Camera (DukeMTMC)

برای ایجاد پایگاه داده DukeMTMC-reID به‌عنوان یکی از نسخه‌های توسعه یافته پایگاه داده DukeMTMC، به ازای هر ۱۲۰ فریم ویدیویی، افراد موجود در تصاویر به‌صورت دستی مشخص و برچسب‌گذاری شده‌اند. بنابراین پایگاه داده DukeMTMC-reID، شامل ۳۶۴۱۱ تصویر افراد برچسب‌گذاری شده است. تعداد افراد مجزای موجود در این پایگاه داده ۱۸۱۲ نفر است که ۱۴۰۴ فرد در زاویه دید بیش از دو دوربین و ۴۰۸ فرد تنها در زاویه دید یک دوربین قرار داشته‌اند. از بین ۱۴۰۴ فرد مجزا، تعداد ۷۰۲ نفر در داده‌های آموزشی و ۷۰۲ نفر در داده‌های آزمون قرار دارد. از هر فرد مجزا، به ازای هر یک از دوربین‌ها، یک تصویر در مجموعه آزمون و بقیه تصاویر در مجموعه گالری قرار داده شده است. در نتیجه مجموعه داده آموزشی این پایگاه داده شامل ۱۶۵۲۲ تصویر از ۷۰۲ فرد مجزا است. همچنین مجموعه آزمون خود شامل دو مجموعه است؛ به‌طوری که مجموعه تصاویر گالری دارای ۱۷۶۶۱ تصویر از ۱۱۱۰ فرد مجزا (۷۰۲ فرد مشاهده شده در حداقل دو دوربین و ۴۰۸ فرد مشاهده شده در فقط یک دوربین) و مجموعه تصاویر آزمون دارای ۲۲۲۸ تصویر از ۷۰۲ فرد هستند [۳۳].

پایگاه داده DukeMTMC4ReID، مانند پایگاه داده DukeMTMC-reID، نسخه توسعه یافته‌ای از پایگاه داده DukeMTMC است. این پایگاه داده دارای ۱۸۵۲ فرد مجزا است که از بین این افراد تعداد ۱۴۱۳ نفر با تعداد کل ۲۲۵۱۵ تصویر در زاویه دید بیش از دو دوربین و تعداد ۴۳۹ نفر با تعداد کل ۲۱۹۵ تصویر فقط در زاویه دید یکی از دوربین‌ها ظاهر شده‌اند. لازم به ذکر است که تفاوت اصلی این دو پایگاه داده در شیوه تشخیص افراد تصویر است. در واقع پایگاه داده DukeMTMC-reID، تشخیص افراد در تصاویر به‌صورت دستی و در پایگاه داده DukeMTMC4ReID، تشخیص افراد توسط روش Doppia انجام می‌شود [۳۴]. در شکل (۱-۱۹)، نمونه تصاویری از این پایگاه داده نشان داده شده است.



شکل (۱-۱۹): نمونه تصاویری از پایگاه داده DukeMTMC4ReID

● پایگاه داده Airport

تصاویر این پایگاه داده توسط شش دوربین از شبکه نظارت داخلی یک فرودگاه تهیه شده است. هر یک از این دوربین‌ها در مدت ۱۲ ساعت، از محل مخصوص بازرسی و همچنین سالن اجتماعات، فیلم‌هایی با سرعت ۳۰ فریم در ثانیه ثبت کرده‌اند.

با فرض اینکه هر فرد مدت زمان محدودی را در مقابل این دوربین‌ها تردد می‌کند، هر کدام از فیلم‌های تهیه شده به صورت تصادفی به ۴۰ ویدیو کلیپ پنج دقیقه‌ای تقسیم شده است. در این ویدیو کلیپ‌ها، تمامی افراد به وسیله روش‌های تشخیص و ردیابی خودکار موجود، تشخیص و ردیابی شده‌اند. در نهایت تمامی افراد تشخیص داده شده و ردیابی شده، به صورت دستی برچسب‌گذاری شده‌اند. تعداد کل تصاویر موجود در این پایگاه داده ۳۹۹۰۲ تصویر از ۹۶۵۱ فرد مختلف است [۳۵]. در شکل (۱-۲۰) نمونه‌ای از تصاویر این پایگاه داده نشان داده شده است.



شکل (۱-۲۰): نمونه تصاویری از پایگاه داده Airport

● پایگاه داده MSMT17

این پایگاه داده شامل ۱۲۶۴۴۱ تصویر از ۴۱۰۱ فرد مختلف است از این‌رو، از بزرگترین پایگاه داده‌های موجود در زمینه بازشناسایی انسان محسوب می‌شود. برای تهیه تصاویر این پایگاه داده از فریم‌های ثبت شده توسط ۱۲ دوربین در محیط بیرونی و سه دوربین در محیط داخلی یک دانشگاه، استفاده شده است.

برای انتخاب تصاویر از بین فریم‌های ثبت شده توسط این ۱۵ دوربین، ابتدا چهار روز از یک ماه با وضعیت آب و هوایی متفاوت انتخاب شده است. سپس با توجه به سه ساعت از فریم‌های ثبت شده در

صبح، ظهر و بعد از ظهر هر روز، تصاویر افراد مشخص و برچسب‌گذاری شده‌اند. در شکل (۱-۲۱) نمونه تصویری از این پایگاه داده نشان داده شده است [۳۶].



شکل (۱-۲۱): نمونه تصویری از پایگاه داده MSMT17

● پایگاه داده RPIfield

این پایگاه داده شامل ۶۰۱۵۸۱ تصویر از ۱۱۲ فرد مختلف است که توسط ۱۲ دوربین تهیه شده‌اند. در این پایگاه مجموعه گالری تغییر کرده و با گذشت زمان تعداد افراد موجود در این مجموعه افزایش می‌یابد به طوری که بعضی از افرادی که قبلاً جلوی دوربین‌ها مشاهده شده‌اند، ممکن است مجدداً چندین بار دیگر نیز مقابل دوربین‌ها قرار گیرند. این امر باعث شده است که برای بررسی و مقایسه عملکرد سیستم‌های بازشناسایی در گذشت زمان بتوان از این پایگاه داده استفاده کرد [۳۷].

در جدول (۱-۱)، خلاصه‌ای از اطلاعات مربوط به پایگاه‌های داده معرفی شده در این بخش، ارائه شده است. در این جدول به ازای هر پایگاه داده، ویژگی‌هایی از قبیل نام پایگاه داده، تاریخ انتشار آن، تعداد افراد مجزا، تعداد دوربین‌ها، تعداد کل تصاویر موجود در پایگاه داده، اندازه تصاویر، وجود یا عدم وجود بیش از یک تصویر از هر فرد در پایگاه داده، ارائه شده است.

جدول (۱-۱): تعدادی از پایگاه‌های داده مورد استفاده در زمینه بازشناسایی انسان

پایگاه داده	تاریخ انتشار	تعداد افراد	تعداد دوربین	تعداد تصاویر	ابعاد تصاویر	وجود چند تصویر از هر فرد
VIPeR	۲۰۰۹	۶۳۲	۲	۱۲۶۴	۱۲۸×۴۸	-
GRID	۲۰۰۹	۱۰۲۵	۸	۱۲۷۵	متنوع	-
3DPeS	۲۰۱۱	۱۹۲	۸	۱۰۱۱	متنوع	✓
WARD	۲۰۱۲	۷۰	۳	۴۷۸۶	۱۲۸×۴۸	✓
CUHK01	۲۰۱۲	۹۷۱	۲	۳۸۸۴	۱۶۰×۶۰	✓
CUHK02	۲۰۱۳	۱۸۱۶	۱۰	۷۲۶۴	۱۶۰×۶۰	✓
CUHK03	۲۰۱۴	۱۴۶۷	۱۰	۱۳۱۶۴	متنوع	✓
PRID450s	۲۰۱۴	۴۵۰	۲	۹۰۰	متنوع	-
RAiD	۲۰۱۴	۴۳	۴	۶۹۲۰	۱۲۸×۶۴	✓
HDA Person	۲۰۱۴	۸۵	۱۳	۲۹۷۶	متنوع	✓
Market1501	۲۰۱۵	۱۵۰۱	۶	۳۲۲۱۷	۱۲۸×۶۴	✓
PRW	۲۰۱۶	۹۳۲	۶	۳۴۳۰۴	متنوع	✓
Large scale person search	۲۰۱۶	۸۴۳۲	-	۱۸۱۸۴	متنوع	-
MARS	۲۰۱۶	۱۲۶۱	۶	۱۱۹۱۰۰۳	۲۵۶×۱۲۸	✓
DukeMTMC-reID	۲۰۱۷	۱۸۱۲	۸	۳۶۴۴۱	متنوع	✓
DukeMTMC4ReID	۲۰۱۷	۱۸۵۲	۸	۴۶۲۶۱	متنوع	✓
Airport	۲۰۱۷	۹۶۵۱	۶	۳۹۹۰۲	۱۲۸×۶۴	✓
MSMT17	۲۰۱۸	۴۱۰۱	۱۵	۱۲۶۴۴۱	متنوع	✓
RPIfield	۲۰۱۸	۱۱۲	۱۲	۶۰۱۵۸۱	متنوع	✓

۱-۶- اهداف این پژوهش

همان‌طور که اشاره شد، سیستم‌های نظارتی در بسیاری از مکان‌های امنیتی و ساختمان‌های هوشمند مورد استفاده گسترده‌ای قرار می‌گیرند. با توجه به اهمیت و تاثیر استفاده از این سیستم‌ها در مکان‌های امنیتی و نظارت هوشمند، بهبود روز افزون عملکرد سیستم‌های نظارتی امری ضروری است. از طرفی، یکی از مهم‌ترین بخش‌های یک سیستم نظارتی، سیستم بازشناسایی انسان است. این اهمیت تا جایی

است که عملکرد سیستم‌های بازشناسایی به‌طور مستقیم بر عملکرد سیستم‌های نظارتی تاثیر می‌گذارد. بنابراین با بهبود عملکرد سیستم‌های بازشناسایی، عملکرد سیستم‌های نظارتی هم بهبود می‌یابد. همان‌طور که اشاره شد، فرض عدم تغییر در ویژگی‌های ظاهری افراد در طول زمان، عدم توجه به مسئله انسداد تمام یا بخشی از بدن افراد به‌دلیل حمل اشیاء، و همچنین استخراج ویژگی‌های ظاهری از کل تصویر، از جمله فرض‌های محدود کننده و مشکلات موجود در سیستم‌های بازشناسایی انسان هستند. با توجه به مسائل و مشکلات مطرح شده در سیستم‌های بازشناسایی موجود، در این پژوهش راهکاری ارائه می‌شود تا توسط آن عملکرد سیستم‌های بازشناسایی انسان بهبود یابد. در این راستا، سیستمی برای بازشناسایی انسان پیشنهاد می‌شود که در آن علاوه بر توجه به ویژگی‌های ظاهری موجود در سایر روش‌های بازشناسایی، مسئله انسداد ناشی از حمل اشیاء و تغییرات در پس‌زمینه نیز مورد توجه قرار می‌گیرد. نتیجه این امر افزایش دقت سیستم بازشناسایی پیشنهادی خواهد بود.

۱-۷- نوآوری‌ها و دستاوردهای رساله

با توجه به نکات عنوان شده در بخش قبل، نتایج حاصل از راهکار پیشنهادی در این پژوهش، به‌صورت تعدادی مقاله پذیرش شده و در دست داوری در مجله‌های بین‌المللی داخلی و خارجی مختلف ارائه شده است. در ادامه این بخش به‌طور مختصر به معرفی هر یک از این مقالات و نوآوری‌های انجام گرفته در آنها می‌پردازیم.

- با توجه به اهمیت ویژگی‌های استخراجی از بدن فرد در مسئله بازشناسایی، ابتدا یک روش قطعه‌بندی بدون ناظر و بدون نیاز به مرحله آموزش، به‌منظور قطعه‌بندی معنایی نواحی پس‌زمینه و پیش‌زمینه تصویر ارائه شده است. در نهایت از تصویر قطعه‌بندی شده به‌عنوان نقشه وزن برای وزن‌دهی ویژگی‌های استخراجی از هر پیکسل تصویر استفاده می‌شود. نتایج حاصل از این راهکار در مقاله زیر ارائه شده است.

Zahra Mortezaie, Hamid Hassanpour, Azeddine Beghdadi, Re-Identification in Video Surveillance Systems Considering Appearance Changes, *International Journal of Science and Technology (SCIENTIA IRANICA)*, Submitted in the revised form on May 31, 2022.

● در ادامه برای بهبود راهکار پیشنهادی اول، از قابلیت‌های شبکه عصبی عمیق برای قطعه‌بندی معنایی تصاویر استفاده می‌شود. پس از قطعه‌بندی تصویر، با توجه به میزان اهمیت هر ناحیه در مسئله بازشناسایی، برای ویژگی‌های استخراجی از پیکسل‌های نواحی، به صورت دستی وزنی اختصاص می‌یابد. نتایج حاصل از این راهکار در مقاله زیر ارائه شده است.

Zahra Mortezaie, Hamid Hassanpour, Azeddine Beghdadi, Sensing Image Regions for Enhancing Accuracy in People Re-Identification, *Iranian (Iranica) Journal of Energy & Environment*, 13(3), pp. 295-304, (2022), doi:10.5829/IJEE.2022.13.03.09.

● به منظور جلوگیری از وزن‌دهی دستی نواحی تصویر، نقشه وزنی برای پیکسل‌های مربوط به اشیاء در حال حمل و پس‌زمینه، با توجه به اندازه هر ناحیه و فاصله پیکسل‌ها تا مرکز تصویر پیشنهاد می‌شود. نتایج حاصل از این راهکار در مقاله زیر ارائه شده است.

Zahra Mortezaie, Hamid Hassanpour, Azeddine Beghdadi, People Re-Identification Using Part-Based Weighing Mechanisms, *Iranian Journal of Electrical and Electronic Engineering (IJEET)*, Submitted on Jan 19, 2022.

● در ادامه، با توجه به اهمیت بازیابی نواحی مسدود بدن در مسئله بازشناسایی، ابتدا تصاویر ورودی توسط فرآیندی به نام فرآیند یکسان‌سازی، پردازش می‌شوند. در این فرآیند، نواحی بدن که توسط اشیاء در حال حمل مسدود شده است، با توجه به پیکسل‌های نواحی غیرمسدود بازیابی می‌شود. سپس ویژگی‌های استخراجی از پیکسل‌های مربوط به اشیاء در حال حمل، نواحی بازیابی شده و نواحی پس‌زمینه، با توجه به فاصله پیکسل‌ها تا مرکز تصویر وزن-دهی می‌شوند. نتایج حاصل از این راهکار در مقاله زیر ارائه شده است.

Zahra Mortezaie, Hamid Hassanpour, Azeddine Beghdadi, People Re-Identification under Occlusion and Crowded Background, *Multimedia Tools and Applications*, pp. 1-21, (2022), doi:10.1007/s11042-021-11868-y.

● در نهایت با توجه به مطالعات انجام شده برای گردآوری این پژوهش، با تمرکز بر مسئله انسداد، مروری بر سیستم‌های بازشناسایی موجود در این زمینه انجام شده است. نتایج این مطالعات در مقاله زیر ارائه شده است.

Zahra Mortezaie, Hamid Hassanpour, A Survey on People Re-Identification Approaches Considering Occlusion, *Majlesi Journal of Electrical Engineering (MJEE)*, Accepted on Jun 28, 2022.

۱-۸- فرض‌های اولیه سیستم پیشنهادی

می‌خواهیم در سیستم‌های نظارت ویدیویی بین افرادی که در محیطی عمومی (فضای داخلی یا خارجی ساختمان) قرار دارند، کار بازشناسی را انجام دهیم. فرض‌های اولیه که برای شرایط موجود در این محیط در نظر گرفته شده است، به صورت زیر است:

- این محیط حاوی چندین دوربین با دیدهای مجزا است.
- افراد جدید فقط از طریق درگاه‌های ورودی (یعنی دوربین‌های مشخص) وارد سامانه نظارتی می‌شوند.
- در این محیط ممکن است چندین فرد به‌طور همزمان در مقابل یک دوربین قرار گیرند.
- محیط مورد نظر، محیط شلوغی نیست و بدن فرد حداقل در بعضی از فریم‌ها به‌طور کامل دیده می‌شود.
- عمل بازشناسایی در این محیط، برای یک بازه زمانی کوتاه (مثلاً یک روز) انجام می‌شود.

۱-۹- جمع‌بندی

در این فصل ابتدا به بررسی اجمالی سیستم‌های نظارتی پرداختیم. در ادامه با توجه به تاثیرگذاری سیستم‌های بازشناسایی در عملکرد سیستم‌های نظارتی، نحوه عملکرد سیستم‌های بازشناسایی به صورت کلی مورد بررسی قرار گرفت. با توجه به برخی مشکلات موجود، بسیاری از سیستم‌های بازشناسایی برای ساده‌سازی مسئله، فرض‌های اولیه و محدود کننده‌ای در نظر می‌گیرند. اما در شرایط دنیای واقعی، این فرض‌ها باعث کاهش دقت سیستم‌های بازشناسایی خواهد شد. از این‌رو، در این فصل تعدادی از مشکلات موجود در سیستم‌های بازشناسایی و همچنین برخی از فرض‌های محدود کننده این سیستم‌ها بررسی شد.

فصل دوم

مروری بر کارهای پیشین

۲-۱- مقدمه

در فصل قبل اشاره شد که بسیاری از سیستم‌های بازشناسایی موجود با در نظر گرفتن فرض‌های اولیه، سعی دارند تا به نحوی مسئله بازشناسایی را ساده‌تر کنند. اما از آنجایی که برخی از این فرض‌ها با شرایط دنیای واقعی مطابقت ندارد، منجر به ایجاد مشکلاتی در عملکرد سیستم‌های بازشناسایی می‌شود. به دلیل اهمیت این موضوع، در فصل قبل تعدادی از فرض‌ها و مشکلات موجود در سیستم‌های بازشناسایی به صورت اجمالی مورد بررسی قرار گرفت.

در این فصل، با توجه به فرض‌ها و مشکلات مطرح شده درباره سیستم‌های بازشناسایی انسان، به صورت اجمالی به بررسی و نقد کارهای انجام شده در این زمینه می‌پردازیم.

تاکنون رویکردهای زیادی برای بهبود عملکرد سیستم‌های بازشناسایی معرفی شده است. این رویکردها را می‌توان به دو دسته تقسیم کرد.

در دسته اول، رویکردهای موجود، تأثیرات مخرب انسداد بدن و پس‌زمینه‌های شلوغ را بر ویژگی‌های استخراج شده در نظر نمی‌گیرند. برخی از این رویکردها با استفاده از ویژگی‌های ظاهری استخراج شده از روش‌های مبتنی بر یادگیری عمیق، تلاش می‌کنند عملکرد سیستم‌های بازشناسایی را بهبود دهند [۳۸-۴۱]. برخی دیگر از این رویکردها، صرفاً مسائل چالش برانگیزی مانند تغییرات ظاهری ناشی از تغییر در حالت و موقعیت بدن افراد نسبت به دوربین و تغییر در روشنایی محیط [۴۲-۵۰]، شرایط مختلف دوربین‌ها [۵۱، ۵۲]، ناهماهنگی توزیع داده‌ها در نمای دوربین‌ها [۵۳]، موضوع کمبود داده‌های آموزشی [۵۴-۵۶]، و همچنین وجود ویژگی‌های ظاهری متمایز^۱ در تصاویر [۵۷، ۵۸] را در نظر می‌گیرند. این در حالی است که سایر رویکردهای موجود در این دسته سعی می‌کنند برای بهبود عملکرد سیستم‌های بازشناسایی، از ویژگی‌های ظاهری منحصربفرد^۲ استفاده کنند [۵۹-۶۸].

^۱ Discriminative Appearance Features

^۲ Salient Appearance Features

در دسته دوم، رویکردهای موجود تلاش می‌کنند سیستم‌های بازشناسایی را در برابر مسائل مربوط به انسداد جزئی بدن [۷۷-۸۶] و پس‌زمینه‌های شلوغ [۸۷-۹۰] مقاوم کنند. اما رویکردهای موجود در دسته دوم هنگام استخراج ویژگی‌های ظاهری افراد، مسائل انسداد ناشی از حمل اشیاء و پس‌زمینه‌های شلوغ را به‌صورت همزمان در نظر نمی‌گیرند.

در ادامه این فصل به بررسی برخی از روش‌های بازشناسایی موجود در دو دسته عنوان شده می‌پردازیم.

۲-۲- بازشناسایی انسان بدون توجه به انسداد بدن و پس‌زمینه شلوغ

همانطور که ذکر شد، برخی از رویکردهای بازشناسایی انسان مسئله انسداد بدن و پس‌زمینه شلوغ در استخراج ویژگی‌ها را در نظر نمی‌گیرند [۳۸-۶۸] در این بخش، برخی از این رویکردها بررسی می‌شود. در سال‌های اخیر روش‌های یادگیری عمیق در بسیاری از مسائل حوزه بینایی ماشین، مورد استفاده قرار می‌گیرد. از این‌رو برخی از روش‌های بازشناسایی و ردیابی انسان برای استخراج ویژگی‌های مناسب، از قابلیت‌های روش‌های یادگیری عمیق استفاده کرده‌اند. در [۳۸]، برای استخراج ویژگی‌های ظاهری سراسری از تصاویر از شبکه‌های عصبی عمیق استفاده شده است. در این راستا، آموزش شبکه عصبی پیچشی^۱، براساس بهینه‌سازی فرآیند یادگیری بازنمایی عمیق^۲، توسط تجزیه بردار منفرد^۳ انجام شده است. در مرجع [۳۹]، شبکه عمیقی با عنوان شبکه تراز عابر پیاده^۴ ارائه شده است که در آن بر اساس نقشه‌های ویژگی شبکه عصبی پیچشی، علاوه بر تعیین محدوده افراد در تصاویر، ویژگی‌های ظاهری افراد به‌طور همزمان یاد گرفته می‌شود. در [۴۰]، ویژگی‌های ظاهری افراد با استفاده از ترکیب برخی از رویکردهای استخراج ویژگی مبتنی بر شبکه‌های عمیق و توصیفگرهای دستی^۵ استخراج می‌شود. در این مرجع، بازشناسایی افراد براساس ویژگی‌های ترکیبی، توسط تعدادی از روش‌های یادگیری متریک^۶ انجام می‌شود. سپس، فهرست‌های رتبه‌بندی به‌دست‌آمده از روش‌های یادگیری متریک، با هم جمع

¹ Convolutional Neural Network (CNN)

² Deep representation learning process

³ Singular Vector Decomposition (SVD)

⁴ Pedestrian Alignment Network (PAN)

⁵ Hand crafted features

⁶ Metric learning

می‌شود. در مرجع [۴۱] برای یادگیری ویژگی‌های متمایز کننده از قسمت‌های مختلف بدن، یک ماژول از شبکه‌های مبتنی بر تابع هزینه قسمت‌های بدن^۱ پیشنهاد شده است. در این ماژول، ابتدا قسمت‌های مختلف بدن با توجه به تابع هزینه مربوطه تعیین می‌شوند. سپس، برای هر قسمت بدن، تابع هزینه دسته‌بندی به صورت جداگانه بهینه می‌شود.

در مرجع [۴۲] به منظور بهبود عملکرد سیستم‌های بازشناسایی در برابر تغییرات موقعیت و حالت بدن^۲ فرد نسبت به دوربین، روشی مبتنی بر انتقال تناظر گراف^۳ ارائه شده است. در این روش ابتدا با استفاده از یک فرآیند تطبیق گراف مبتنی بر تکه^۴، الگوهای تناظر مبتنی بر تکه^۵، از زوج تصاویر مربوط به یک فرد^۶ یادگرفته می‌شوند. سپس، تعدادی از شبیه‌ترین جفت‌های آموزشی از نظر حالت بدن، به عنوان تصاویر مرجع برای هر یک از جفت تصاویر آزمون انتخاب می‌شوند. در ادامه تناظر بین تصاویر مرجع، به جفت تصاویر آزمون نگاشت داده می‌شود تا شباهت بین تصاویر مشخص شود. همچنین، برای بهبود انتقال تناظر مورد استفاده در [۴۲]، یک توصیفگر حالت بدن مبتنی بر ساختار توپولوژی نقاط اتصال^۷ قسمت‌های مختلف بدن [۶۹]، در [۴۳] استفاده می‌شود.

همچنین در مرجع [۴۴] برای مقابله با مسئله انسداد و تغییرات در حالت بدن افراد، از ویژگی‌های مدل رنگی^۸ HSV و الگوی سه‌گانه محلی با مقیاس ثابت^۹ [۷۰] استفاده شده است. در این روش با تقسیم تصاویر به صورت افقی و عمودی، ویژگی‌های مورد نظر به صورت محلی استخراج می‌شوند. در [۴۵] روشی به نام توصیفگر وقوع بیشینه محلی^{۱۰} ارائه شده است. در این روش نیز با تقسیم‌بندی تصاویر به تکه‌های همپوشان^{۱۱}، ویژگی‌های مدل رنگی HSV و الگوی سه‌گانه محلی با مقیاس ثابت به صورت محلی استخراج می‌شوند.

¹ Part loss networks

² Viewpoint and pose variations

³ Graph Correspondence Transfer (GCT)

⁴ Patch-wise

⁵ Patch-wise correspondence templates

⁶ Positive image pairs

⁷ Joint locations

⁸ Hue-Saturation-Value (HSV)

⁹ Scale Invariant Local Ternary Pattern (SILTP)

¹⁰ Local Maximal Occurrence (LOMO) representation

¹¹ Overlapping patches

در [۴۶] و [۴۷] راهکاری برای مقاوم‌سازی سیستم بازشناسایی در برابر مشکلاتی مانند تغییر در روشنایی محیط و تغییر زاویه افراد ارائه شده است. در سیستم‌های بازشناسایی ارائه شده ابتدا تصویر افراد به صورت سطری به هفت ناحیه همپوشان تقسیم می‌شود. سپس، هر کدام از این ناحیه‌ها به صورت تعدادی تکه متراکم در نظر گرفته می‌شود. هر یک از پیکسل‌های موجود در تکه‌ها، با استفاده از بردار ویژگی هشت‌تایی شامل مختصات پیکسل مورد نظر در راستای محور عمودی، اندازه گرادینان در جهت چهار زاویه مختلف و مقدار پیکسل مربوطه در سه کانال رنگی قرمز، سبز و آبی معرفی می‌شود. در این مراجع، توزیع گوسین تکه‌های موجود در هر یک از هفت ناحیه از تصویر، با استفاده از بردار ویژگی ایجاد شده برای هر پیکسل از تصویر، محاسبه می‌شود. با این کار هر یک از هفت ناحیه سطری تصویر، توسط تعدادی توزیع گوسین توصیف می‌شود. لازم به ذکر است که فضای توزیع‌های گوسین فضایی غیرخطی است. بنابراین عملیات اقلیدسی را نمی‌توان به طور مستقیم روی این فضا اعمال کرد. از این رو، در [۴۶] و [۴۷]، فضای گوسین به فضای مماسی^۱ اقلیدسی نگاشت می‌شود. سپس از توزیع‌های گوسین نگاشت شده تکه‌های هر ناحیه، برای محاسبه توزیع گوسین ناحیه مربوطه استفاده می‌شود. پس از محاسبه توزیع گوسین هر یک از هفت ناحیه موجود در تصویر، فضای گوسین به فضای مماسی اقلیدسی نگاشت می‌شود. در نهایت با کنار هم قرار دادن فضای گوسین نگاشت شده هر یک از هفت ناحیه تصویر، بردار ویژگی نهایی برای کل تصویر حاصل می‌شود. لازم به ذکر است که در مراجع [۴۶] و [۴۷]، به منظور افزایش دقت بردار ویژگی حاصل، علاوه بر مدل رنگی RGB^۲، از مدل‌های رنگی HSV، Lab^۳ و nRnG^۴ نیز استفاده می‌شود. همچنین در این مراجع برای این دو توصیفگر به ترتیب نام توصیفگر گوسین گوسین^۵ (GoG) و توصیفگر گوسین سلسله‌مراتبی^۶ (HGD) انتخاب شده است.

^۱ Tangent space

^۲ Red-Green-Blue (RGB)

^۳ Lightness-a-b (Lab)

^۴ Normalized Red-normalized Green (nRnG)

^۵ Gaussian of Gaussian (GoG)

^۶ Hierarchical Gaussian Descriptor (HGD)

همچنین در [۴۸]، از توصیفگری با نام توصیفگر گوسین چند سطحی^۱ استفاده می‌شود. عملکرد این توصیفگر مشابه با توصیفگرهای ارائه شده در [۴۶] و [۴۷] است با این تفاوت که در آن از برخی از دیگر ویژگی‌هایی رنگ و بافت مانند مقادیر لحظه رنگ^۲ موجود در مؤلفه‌های مدل رنگی RGB و پاسخ‌های فیلتر^۳ Schmid استفاده می‌شود.

در رویکرد بازشناسایی ارائه شده در [۴۹]، ویژگی‌های استخراج شده توسط توصیفگر وقوع بیشینه محلی و همچنین ویژگی‌های استخراج شده از شبکه‌های عصبی پیچشی، برای بهبود عملکرد سیستم‌های بازشناسایی ترکیب شده است.

در [۵۰]، یک روش رتبه‌بندی مجدد تصاویر بر مبنای Density-Adaptive Smooth Kernel ارائه شده است. در این روش، ارتباط بین نمونه داده‌ها توسط یک تابع فرمول‌بندی می‌شود. لازم به ذکر است که در این روش از توصیفگرهای وقوع بیشینه محلی و گوسین گوسین برای نمایش تصاویر استفاده شده است.

در مرجع [۵۱] برای مقابله با تغییرات ظاهری افراد ناشی از شرایط مختلف دوربین‌ها، سعی می‌شود تا تصاویر توسط اطلاعات متمایز حاصل از زاویه دید دوربین‌های متقابل^۴ نمایش داده شود. در این رویکرد، ابتدا از توصیفگر گوسین گوسین برای استخراج ویژگی‌های ظاهری تصاویر استفاده می‌شود. سپس، ویژگی‌های استخراج شده به زیر فضاها یادگرفته شده نگاشت می‌شوند. همچنین در مرجع [۵۲]، یک مدل رگرسیون غیرخطی برای نگاشت نمونه‌های داده به زیر فضایی با ابعاد پایین‌تر ارائه شده است. مشابه با روش [۵۱]، در این روش نیز ویژگی‌های استخراج شده توسط توصیفگر گوسین گوسین، به زیر فضای یادگرفته شده نگاشت می‌شوند.

یک روش یادگیری شباهت در مرجع [۵۳] ارائه شده است. در این روش سعی می‌شود تا سیستم‌های بازشناسایی در برابر مسئله ناهماهنگی توزیع داده‌ها در نمای دوربین‌ها مقاوم شود. در این رویکرد،

¹ Multi-Level Gaussian Descriptor (MLGD)

² Color moment values

³ Schmid filter responses

⁴ Cross-view discriminative information

مسئله بازشناسایی افراد به عنوان یک مسئله بهینه‌سازی یادگیری انتقالی مشترک با چند دید تکراری^۱ در نظر گرفته می‌شود. در نهایت این مسئله بهینه‌سازی با استفاده از الگوریتم ضرب‌کننده افزایش یافته لاگرانژ تضعیف شده^۲ [۷۱] حل می‌شود.

به منظور بهبود عملکرد سیستم‌های بازشناسایی، مسئله کمبود نمونه‌های آموزشی در مراجع [۵۴-۵۶] مورد توجه قرار گرفته است. در [۵۴]، ماژولی ارائه شده است که شامل دو لایه غیرپارامتری^۳ با عنوان لایه یکپارچگی میانگین تقارن بدن^۴ و لایه نرمال‌سازی محلی^۵ می‌باشد. ماژول ارائه شده در این مرجع، بین ماژول‌های یادگیری ویژگی‌های عمیق و یادگیری شباهت تعبیه شده است. همچنین در [۵۵]، یک روش یادگیری هم‌متریک نیمه نظارت شده^۶ ارائه شده است. در این روش یک ماتریس فاصله شبیه به فاصله ماهالانوبیس^۷، با استفاده از چند نمونه آموزشی یاد گرفته می‌شود. در این رویکرد، ابتدا ویژگی‌هایی مانند اسامی رنگ‌های منحصر بفرده^۸ موجود در ظاهر افراد [۷۲] و ویژگی‌های استخراج شده از شبکه عصبی پیچشی سیامز^۹ [۷۳] برای نمایش تصاویر در یک نما^{۱۰} استفاده می‌شود. در همین حال، یک روش یادگیری وزن دوتایی^{۱۱} برای تجزیه ویژگی‌های تصاویر در یک نما به نماهای شبه دوتایی^{۱۲} استفاده می‌شود. در نهایت ویژگی‌های تجزیه شده برای یادگیری مدل‌های متریک استفاده می‌شود. در مرجع [۵۶] مشابه با روش‌های [۵۴] و [۵۵]، سعی می‌شود با استفاده از داده‌های کم آموزشی، زیر فضایی برای نگاشت داده‌های اخذ شده توسط هر یک از دوربین‌ها بدست آید.

روش بازشناسایی ارائه شده در [۵۷]، شامل سه قسمت رتبه‌بندی، رتبه‌بندی مجدد تصاویر آموزشی و بهینه‌سازی بعد از رتبه‌بندی است. در قسمت رتبه‌بندی این روش، ابتدا ویژگی‌های ظاهری مانند بافت، رنگ و شکل بدن افراد به عنوان ویژگی‌های ورودی سیستم بازشناسایی استخراج می‌شوند. سپس با

¹ Consistent iterative multi-view joint transfer learning optimal problem

² Inexact Augmented Lagrange Multiplier (IALM)

³ Nonparametric

⁴ Body symmetry average pooling layer

⁵ Local normalization layer

⁶ Semi-supervised co-metric learning

⁷ Mahalanobis

⁸ Salient color name

⁹ Siamese

¹⁰ Single-view images

¹¹ Binary-weight learning

¹² Pseudo binary views

فرض اینکه تصاویر پایگاه داده فقط توسط دو دوربین الف و ب، تهیه شده اند، مجموعه‌ای از زوج بردارهای ویژگی از این دو دوربین، برای آموزش کلاس‌بندی که زوج تصاویر مربوط به یک فرد^۱ و همچنین زوج تصاویر مربوط به دو فرد مجزا^۲ را تشخیص می‌دهد، مورد استفاده قرار می‌گیرند. پس از آموزش کلاس‌بند، پارامترهای آن برای محاسبه فاصله بین تصویر آزمون و تصاویر مجموعه گالری استفاده شده و در واقع رتبه‌بندی اولیه تصاویر انجام می‌شود. در قسمت رتبه‌بندی مجدد تصاویر آموزشی، هدف آموزش مجدد کلاس‌بند با استفاده از ویژگی‌های جزئی و متمایزکننده افراد است. در این قسمت ابتدا هر زوج تصویر آموزشی از دو دوربین الف و ب، توسط کلاس‌بند آموزش دیده در قسمت قبل، رتبه‌بندی می‌شوند. سپس با استفاده از یک روش تحلیل اطلاعات متمایز کننده براساس محتوا و اطلاعات زمینه تصویر، اطلاعات ظاهری و سراسری از بردار ویژگی تصاویر افراد، حذف می‌شود. در نهایت در این قسمت بردارهای ویژگی جدیدی برای تصاویر افراد بدست می‌آید که در آن اطلاعات متمایز کننده افراد قرار دارد. در قسمت سوم این روش، کلاس‌بند این بار با استفاده از بردارهای ویژگی جدید آموزش می‌بیند تا بتواند بردارهای ویژگی جدید را مجدداً رتبه‌بندی کند. پس از آموزش کلاس‌بندها، با ورود تصویر آزمون، ابتدا بردار ویژگی اولیه تصویر آزمون به بردار ویژگی جدید تبدیل می‌شود که در آن ویژگی‌های متمایز کننده تصویر آزمون موجود است. سپس با استفاده از کلاس‌بند آموزش دیده در قسمت سوم سیستم، تصاویر مجموعه گالری رتبه‌بندی می‌شوند.

در [۵۸]، برای بدست آوردن ویژگی‌های متمایز کننده از تصاویر افراد، روش یادگیری دیکشنری^۳ متمایز کننده ارائه شده است. در این سیستم برای مقابله با مشکل تغییر زاویه افراد و در نتیجه تغییرات ظاهری آنها، دو نگاهت نامتقارن برای نگاهت ویژگی‌های ورودی به زیر فضای دیگر ارائه شده است. همچنین برای افزایش خاصیت متمایزکنندگی دیکشنری آموزش دیده، محدودیت‌های لیستی و انطباق

^۱ Positive Pairs
^۲ Negative Pairs

^۳ Dictionary Learning

افراد^۱، در فرآیند یادگیری دیکشنری مورد استفاده قرار گرفته است. لازم به ذکر است که در این روش ویژگی‌های اولیه مورد استفاده، توسط توصیفگر وقوع بیشینه محلی [۴۵] استخراج می‌شود.

باید به این نکته توجه داشت که در مسئله بازنشاسایی انسان، علاوه بر استفاده از ویژگی‌های ظاهری افراد مانند رنگ و طرح لباس، می‌توان از ویژگی‌های سطح بالاتری نیز بهره برد. ویژگی‌های منحصربفرد یا متمایزکننده در مجموعه داده در حال بررسی و تشخیص آن می‌تواند به‌عنوان یک ویژگی سطح بالا در سیستم‌های بازنشاسایی مورد استفاده قرار گیرد. در واقع ویژگی‌های منحصربفرد با توجه به قدرت متمایزکنندگی بالاتر نسبت به سایر ویژگی‌های ظاهری، می‌توانند به‌عنوان یک ویژگی مهم در بازنشاسایی افراد مورد استفاده قرار گیرند. به‌عنوان مثال، پوشیدن لباسی با رنگ خاص، داشتن علامت یا شکل خاص روی لباس و حمل برخی اشیاء می‌تواند به‌عنوان یک ویژگی منحصربفرد برای افراد در نظر گرفته شده و در نهایت با استفاده از این ویژگی‌ها کار شناسایی افراد دقیق‌تر و راحت‌تر انجام شود. در واقع با در نظر گرفتن ویژگی‌های منحصربفرد می‌توان فضای مورد جستجو را کاهش داد. نتیجه کاهش فضای مورد جستجو، افزایش دقت بازنشاسایی افراد خواهد بود.

در ادامه به بررسی تعدادی از روش‌های بازنشاسایی موجود خواهیم پرداخت که در آنها از ویژگی‌های ظاهری و ویژگی‌های منحصربفرد افراد استفاده شده است. در واقع در تمامی روش‌های مورد بررسی در این بخش، برای تعیین وجود یا عدم وجود ویژگی‌های منحصربفرد در تصاویر، از ویژگی‌های ظاهری استفاده شده است.

روش بازنشاسایی ارائه شده در [۵۹]، با توجه به خصیصه‌های استخراج شده از تصاویر، سعی دارد تا ویژگی‌های سطح پایین را وزن‌دهی کند. در این روش ابتدا تصاویر برچسب نخورده بر اساس پارتیشن-بندی گراف، خوشه‌بندی می‌شود. خروجی خوشه‌بندی، مجموعه‌ای از نماینده‌ها است که هر کدام نشان-دهنده یک خصیصه هستند. سپس برای هر نماینده یک بردار وزن متناظر با ویژگی‌های سطح پایین

¹ Listwise and Identity Consistency Constraints

محاسبه می‌شود. در این روش، وزن‌دهی هر یک از ویژگی‌ها به این صورت انجام می‌شود که ابتدا به کمک یک ویژگی، داده‌های نماینده‌های مختلف دسته‌بندی می‌شود. هر ویژگی که بهتر بتواند یک نماینده خاص را از سایر نماینده‌ها جدا کند، از اهمیت بالاتری برای نماینده مورد نظر برخوردار خواهد بود. در مرحله بازنسازایی، با ورود هر فرد جدید به سیستم، ابتدا خوشه یا نماینده تصویر ورودی تعیین می‌شود. سپس در خوشه مورد نظر، با استفاده از بردار وزنی متناظر با ویژگی‌های آن، تصاویر موجود در خوشه مذکور رتبه‌بندی می‌شود.

یکی از ایراداتی که به این روش وارد می‌شود این است که خروجی خوشه‌بندی، تعداد از پیش تعیین شده‌ای خوشه خواهد بود. این در حالی است که ممکن است ورودی سیستم فردی با خصیصه و ویژگی منحصربفردی باشد و در هیچ یک از خوشه‌های موجود قرار نگیرد.

لاین و همکاران در مرجع [۶۰]، با در نظر گرفتن ویژگی‌های سطح میانی که برای انسان نیز قابل فهم است، سعی کردند تا کیفیت بازنسازایی بهبود یابد. در این روش ابتدا تعداد محدودی از ویژگی‌های سطح میانی مانند شلوار کوتاه، موی بلند، موی کوتاه، قد بلند، قد کوتاه، حمل شیء، حمل کوله پشتی و ... توسط یک فرد خبره تعیین می‌شود. سپس با استفاده از یک پایگاه داده برای هر یک از ویژگی‌های سطح میانی تعیین شده، یک کلاس‌بند طراحی می‌شود. سپس برای بازنسازایی فرد جدید، با استفاده از هر یک از کلاس‌بندها مشخص می‌شود که کدام یک از ویژگی‌های سطح میانی مورد نظر در فرد جدید وجود دارد. در نهایت برای هر فرد برداری از وجود یا عدم وجود ویژگی‌های میانی مورد نظر، ایجاد می‌شود.

یکی از ایرادات وارد شده به این روش هم مانند ایراد روش قبلی است. در واقع این روش تعداد محدودی ویژگی منحصربفرد را برای افراد در نظر می‌گیرد، در حالی که ممکن است فرد ورودی دارای ویژگی منحصربفردی غیر از ویژگی‌های مورد نظر در این مقاله باشد. به‌عنوان مثال ممکن است فردی در فضای مورد جستجو، جارو به دست گرفته باشد. این ویژگی می‌تواند به‌عنوان یک ویژگی منحصربفرد برای بازنسازایی فرد مورد نظر، استفاده شود. این در حالی است که در روش [۶۰]، ویژگی‌هایی چون حمل

جارو بین ویژگی‌های منحصربفرد در نظر گرفته نشده است. بنابراین با در نظر گرفتن تعداد محدودی ویژگی منحصربفرد، ممکن است سایر ویژگی‌های منحصربفرد را از دست بدهیم. از این رو باید به دنبال ارائه راهکاری بود که ویژگی‌های منحصربفرد در افراد را بدون هیچگونه پیش فرض و تصمیم‌گیری قبلی و به صورت کاملاً پویا، تعیین کند.

در [۶۱] و [۶۲] راهکاری برای بازشناسایی افراد با استفاده از ویژگی‌های منحصربفرد ارائه شده است. در این دو مرجع، وجود یا عدم وجود ویژگی‌های منحصربفرد در تصاویر، به صورت پویا تعیین شده و در آنها ویژگی منحصربفرد از پیش تعیین شده‌ای در نظر گرفته نمی‌شود. در این دو مرجع ابتدا با استفاده از یک رویکرد بدون نظارت، نواحی مختلف هر کدام از تصاویر آزمون از نظر منحصربفرد بودن یا نبودن، مورد بررسی قرار می‌گیرد. در این رویکرد بدون نظارت، ابتدا تصاویر آزمون به صورت تکه‌های^۱ کوچک با ابعاد مساوی، در نظر گرفته می‌شوند. سپس میزان منحصربفرد بودن هر تکه با استفاده از مجموعه تصاویری به نام مجموعه تصاویر مرجع تعیین می‌شود. در واقع میزان منحصربفرد بودن هر تکه از تصویر آزمون، با بررسی محدوده مشخصی از هر تصویر مجموعه مرجع، و با استفاده از روش‌های k نزدیکترین همسایه^۲ و ماشین بردار پشتیبان یک کلاسه^۳، تعیین می‌شود. پس از تعیین میزان منحصربفرد بودن هر تکه از تصویر آزمون، برای بررسی میزان تطابق آن با تصاویر مجموعه گالری، مرجع [۶۱] از روش تطابق بدون نظارت و مرجع [۶۲] از دو روش تطابق بدون نظارت مشابه [۶۱] و روش تطابق با نظارت استفاده می‌شود.

از آنجایی که هدف در هر دو روش تطابق بدون نظارت و تطابق با نظارت، کمینه کردن تابع هزینه مربوطه است، به ازای هر تکه از تصویر آزمون، محدوده مشخصی از تصویر گالری بررسی شده و مقدار تابع هزینه تعیین می‌شود. در صورتی که در محدوده مورد نظر میزان منحصربفرد بودن تصویر آزمون و تصویر گالری با هم تطابق نداشته باشد، مقدار تابع هزینه افزایش و در غیراین صورت مقدار این تابع

¹ Patches

² K-Nearest Neighbor (KNN)

³ One-Class Support Vector Machine (OCSVM)

کاهش می‌یابد. بنابراین در دو روش انطباق تصاویر آزمون و گالری به‌صورت بدون نظارت و با نظارت، با کمینه کردن تابع هزینه، تصاویر مشابه با تصویر آزمون مشخص خواهد شد.

همان‌طور که اشاره شد، در روش‌های [۶۱] و [۶۲] برای تعیین میزان منحصربفرد بودن هر تکه از تصویر، از مجموعه مرجع استفاده می‌شود. در [۶۳] و [۶۴] راهکاری بر مبنای پیکسل‌های مجاور ارائه شده است تا برای تعیین میزان منحصربفرد بودن نواحی تصویر، نیاز به استفاده از مجموعه مرجع نباشد. این روش از پنج مرحله تشکیل شده است. در فرآیند آموزش، ابتدا مدلی برای تعیین میزان منحصربفردی هر یک از پیکسل‌های تصاویر افراد، استفاده می‌شود. خروجی این مدل، نقشه‌ای از هر یک از تصاویر است که میزان منحصربفرد بودن هر یک از پیکسل‌ها را با توجه به پیکسل‌های مجاورش در تصویر اولیه، نشان می‌دهد. در مرحله دوم، ویژگی‌هایی مانند میانگین رنگ تصاویر،^۱ SIFT و^۲ LBP از تصاویر اولیه استخراج می‌شود و با توجه به نقشه بدست آمده در مرحله اول، این ویژگی‌ها وزن‌دهی می‌شوند. در مرحله سوم این روش، ابعاد ویژگی‌های استخراج شده و وزن‌دار، توسط روش تحلیل مؤلفه اصلی^۳ کاهش می‌یابد. سپس ویژگی‌های کاهش بعد یافته، در مرحله چهارم، به یک مدل یادگیری متریک چندگانه^۴ اعمال می‌شود. در مرحله پنجم، برای بازشناسایی افراد، ویژگی‌های کاهش بعد یافته، تصویر ورودی، به مدل ترکیبی فاصله‌ای^۵ و مدل متریک چندگانه آموزش دیده اعمال می‌شود تا شباهت و عدم شباهت بین تصاویر تعیین شود.

در [۶۵] یک روش بازشناسایی مبتنی بر مجموعه مرجع ارائه شده است. در این روش به‌جای اینکه برای بازشناسایی افراد، به‌طور مستقیم از ویژگی‌های ظاهری آنها استفاده شود، عمل مقایسه و انطباق افراد در زیر فضایی جدید و با استفاده از توصیفگری به نام توصیفگر مرجع انجام می‌شود. در این روش ابتدا یک مجموعه از تصاویر به نام مجموعه مرجع، با استفاده از روش تجزیه و تحلیل همبستگی کانونی^۶، به زیر فضای دیگری نگاشت داده می‌شود. این نگاشت به زیر فضای جدید به نحوی انجام می‌شود که

¹ Scale Invariant Feature Transform (SIFT)

² Local Binary Pattern (LBP)

³ Principle Component Analysis (PCA)

⁴ Multiple metric learning module

⁵ Distance fusion module

⁶ Regularized canonical correlation analysis

در مجموعه مرجع، همبستگی بین تصاویری که مقابل دوربین‌های مختلف قرار دارند، بیشینه شود. در ادامه تصاویر مجموعه گالری و آزمون نیز به زیر فضای مورد نظر نگاشت می‌شوند. سپس در این زیر فضا، برای ایجاد توصیفگر مرجع، از نظر ویژگی‌های ظاهری مانند رنگ و بافت، میزان شباهت هر کدام از تصاویر مجموعه گالری و مجموعه آزمون با هر یک از تصاویر مجموعه مرجع محاسبه می‌شود. در نهایت برای بازشناسایی تصویری از مجموعه آزمون، توصیفگر مرجع این تصویر با توصیفگرهای مرجع تمامی تصاویر مجموعه گالری، مقایسه شده و برحسب میزان شباهت بین تصویر آزمون و تصاویر گالری، تعدادی از شبیه‌ترین تصاویر مجموعه گالری به صورت نزولی، رتبه‌بندی می‌شود.

در روش ارائه شده در [۶۶] برای بهبود عملکرد سیستم بازشناسایی پیشنهادی در [۶۵]، تصاویر رتبه‌بندی شده را با توجه به وجود یا عدم وجود ویژگی‌های منحصربفرد در تصاویر گالری و آزمون، مجدداً رتبه‌بندی می‌کنند. در این روش، راهکار مورد استفاده برای تشخیص وجود یا عدم وجود ویژگی‌های منحصربفرد در تصاویر، مشابه راهکار ارائه شده در [۶۱] و [۶۲] است. بعد از اینکه وجود ویژگی‌های منحصربفرد در تصویر آزمون و تصاویر رتبه‌بندی شده اولیه مشخص شد، براساس میزان شباهت ویژگی‌های منحصربفرد موجود در تصاویر رتبه‌بندی شده اولیه و تصویر آزمون، تصاویر رتبه‌بندی شده اولیه، مجدداً رتبه‌بندی می‌شوند.

در سیستم بازشناسایی پیشنهادی در [۶۷]، برای بازشناسایی انسان بر اساس نواحی منحصربفرد آنها، از روش یادگیری عمیق استفاده شده است. در این سیستم، برای تمرکز خودکار روی نواحی متمایز کننده تصویر افراد از شبکه عصبی End-to-End CAN^۱ استفاده شده است. در این روش، شبکه عمیق مورد نظر به منظور شبیه‌سازی فرآیند ادراکی ناظر انسانی در بازشناسایی، به نحوی آموزش می‌بیند که بتواند به طور خودکار نواحی منحصربفرد تصویر افراد را تشخیص داده و برای مقایسه بین تصاویر، روی اطلاعات استخراج شده از این نواحی تمرکز کند. این شبکه از دو قسمت تشکیل شده است. در قسمت

^۱ End-to-End Comparative Attention Network

اول این شبکه، یک شبکه عصبی پیچشی قرار دارد که ویژگی‌های متمایزکننده سراسری تصاویر ورودی را یاد می‌گیرد. قسمت دوم این شبکه از یک مؤلفه مقایسه توجه^۱ تشکیل می‌شود. در مرحله آموزش، هر بار سه تصویر اولیه شامل دو تصویر از یک فرد و یک تصویر از فرد دیگر به‌طور همزمان به سه شبکه CAN اعمال می‌شود. در قسمت اول این سه شبکه، ویژگی‌های سراسری تصویر ورودی، توسط شبکه عصبی پیچشی یاد گرفته می‌شوند به‌طوری که این سه شبکه عصبی پیچشی، وزن‌های خود را در حین آموزش با هم به اشتراک می‌گذارند. سپس ویژگی‌های سراسری استخراج شده از سه تصویر ورودی، به قسمت مؤلفه مقایسه توجه از سه شبکه CAN مورد نظر اعمال می‌شود. در این قسمت با مقایسه بین ویژگی‌های سراسری دو تصویر ورودی مربوط به یک فرد و همچنین مقایسه بین ویژگی‌های سراسری تصویر فرد مورد نظر با فرد دیگر، ویژگی‌هایی محلی و متناسب با سیستم بینایی انسان استخراج می‌شود به‌طوری که از بین سه تصویر ورودی، تصاویر مربوط به یک فرد به هم نزدیک و تصاویر غیرمرتبط از هم دور باشند. در نهایت در مرحله آزمون، تصویر آزمون به‌صورت جداگانه با هر یک از تصاویر مجموعه گالری به دو شبکه CAN اعمال شده و بر اساس ویژگی‌های استخراج شده، میزان فاصله آنها با هم تعیین می‌شود.

در سیستم بازشناسایی ارائه شده در [۶۸]، روشی برای خوشه‌بندی نواحی منحصربفرد هر یک از افراد و بازشناسایی انسان با استفاده از تطابق نواحی منحصربفرد تصاویر معرفی شده است. در واقع در این سیستم بازشناسایی، این‌طور فرض شده است که مجموعه گالری شامل دنباله‌ای از فریم‌های ویدیویی از افراد مختلف است. از این‌رو هدف اصلی این سیستم این است که نواحی منحصربفرد در تصاویر مربوط به هر فرد در دنباله‌ای از فریم‌های ویدیویی، در یک خوشه قرار بگیرد. این سیستم شامل سه قسمت است. در قسمت اول این سیستم، از روشی بر اساس متوسط انتقال^۲ [۷۴] برای استخراج نواحی منحصربفرد تصویر افراد استفاده می‌شود. در قسمت دوم این سیستم، از روش LSLDG^۳ [۷۵] برای

¹ Comparative Attention Component

² Mean shift

³ Least-Squares Log-Density Gradient (LSLDG)

خوشه‌بندی نواحی منحصربفرد تصویر افراد استفاده می‌شود. همچنین در قسمت سوم این سیستم، تطابق نواحی منحصربفرد بر اساس هیستوگرام رنگ نواحی انجام می‌شود. قسمت اول این سیستم برای استخراج نواحی منحصربفرد، ابتدا تصویر افراد با استفاده از متوسط انتقال، به نواحی مختلفی قطعه‌بندی می‌شود. سپس، میزان منحصربفرد بودن هر ناحیه توسط روش تشخیص نواحی منحصربفرد بر اساس کنتراست سراسری^۱ [۷۶] محاسبه می‌شود. در قسمت دوم این سیستم، نواحی منحصربفرد مربوط به هر فرد در دنباله‌ای از فریم‌های ویدیویی، توسط روش LSLDG، خوشه‌بندی می‌شود. بنابراین برای هر فرد خوشه‌ای از نواحی منحصربفرد در دنباله‌ای از فریم‌های ویدیویی بدست می‌آید. سپس در قسمت سوم این سیستم، برای بازشناسایی افراد، میزان تطابق نواحی منحصربفرد تصویر آزمون با نواحی منحصربفرد تصاویر گالری که توسط روش LSLDG، خوشه‌بندی شده‌اند، بررسی می‌شود.

لازم به ذکر است که در بعضی از شرایط، سیستم‌های بازشناسایی که براساس ویژگی‌های منحصربفرد عمل می‌کنند، عملکرد مناسبی نخواهند داشت. به‌عنوان مثال، ممکن است با حرکت افراد مقابل یک دوربین یا بین دوربین‌های مختلف و در نتیجه تغییر زاویه افراد نسبت به دوربین، ویژگی منحصربفرد موجود در افراد ناپدید شود. همچنین ممکن است حمل اشیاء توسط افراد منجر به انسداد ویژگی‌های منحصربفرد افراد شود. بنابراین لازم است تا در بازشناسایی افراد مسائل دیگر مانند مسئله انسداد ناشی از حمل اشیاء هم مورد نظر قرار بگیرد.

همچنین، همانطور که در فصل قبل اشاره شد، با حرکت افراد مقابل یک دوربین یا بین دوربین‌های مختلف ممکن است فضای پس‌زمینه دچار تغییرات اساسی شود. در چنین حالتی استفاده از ویژگی‌های ظاهری موجود در فضای پس‌زمینه، عملکرد سیستم بازشناسایی را کاهش می‌دهد.

^۱ Global contrast based salient region detection

۲-۳- بازشناسی انسان با توجه به انسداد و پس‌زمینه شلوغ

رویکردهای موجود در این دسته تلاش می‌کنند سیستم‌های بازشناسایی را در برابر مسائل مربوط به انسداد جزئی بدن [۷۷-۸۶] و پس‌زمینه‌های شلوغ [۸۷-۹۰] مقاوم کنند. در این بخش به بررسی اجمالی تعدادی از این رویکردها می‌پردازیم.

برخی از مراجع برای بهبود عملکرد سیستم‌های بازشناسایی افراد در برابر مسئله انسداد جزئی از روش‌های مبتنی بر یادگیری عمیق استفاده می‌کنند [۷۷] و [۷۸]. در مراجع‌های [۷۷] و [۷۸]، ابتدا با حرکت پنجره‌ای لغزان روی تعدادی از نمونه تصاویر موجود، تعدادی تصویر با انسداد جزئی ایجاد می‌شود. سپس، به‌منظور آموزش یک سیستم بازشناسایی مقاوم در برابر مسئله انسداد جزئی، شبکه عصبی ResNet-50 [۹۱] با استفاده از نمونه تصاویر اصلی و تصاویر ایجاد شده حاوی انسداد جزئی، آموزش می‌بیند.

در رویکرد استخراج ویژگی پیشنهاد شده در مرجع [۷۹]، ابتدا به ازای هر تصویر ورودی نقشه‌ای با عنوان نقشه توجه^۱ با توجه به نواحی کلیدی بدن انسان^۲، بدست می‌آید. سپس از این نقشه به‌منظور هدایت مدل استخراج ویژگی برای نادیده گرفتن نواحی مسدود تصویر استفاده می‌شود.

وانگ و همکاران در مرجع [۸۰] راهکاری برای مقابله با مسئله انسداد در سیستم‌های بازشناسایی ارائه کرده‌اند. در این راهکار، از یک شبکه عصبی پیچشی و یک مدل تخمین نقاط کلیدی به‌منظور استخراج ویژگی‌های محلی معنایی استفاده می‌شود. در این مرجع، با در نظر گرفتن ویژگی‌های محلی استخراجی به‌عنوان گره‌های یک گراف، یک لایه پیچشی گراف جهت تطبیقی^۳، برای انتقال اطلاعات ارتباطی بین گره‌ها معرفی شده است. در این راهکار، بررسی وجود شباهت بین ویژگی‌های محلی دو تصویر مختلف، به‌عنوان یک مسئله تطبیق گراف در نظر گرفته می‌شود. از این‌رو، در [۸۰]، اطلاعات توپولوژی به‌طور

¹ Attention map

² Human landmarks

³ Adaptive Direction Graph Convolutional (ADGC)

مشترک و با استفاده از یک لایه هم ترازوی تعبیه شده گراف متقابل^۱ در ویژگی‌های محلی، آموخته و تعبیه می‌شود.

در [۸۱]، یک شبکه عصبی برای مقابله با مسئله انسداد در سیستم‌های بازنشاسایی پیشنهاد شده است. این شبکه دارای یک بخش برای استخراج ویژگی‌های محلی تصویر، یک بخش برای استخراج ویژگی‌های سراسری تصویر و یک بخش برای قطعه‌بندی معنایی تصویر است. در بخش استخراج کننده ویژگی‌های سراسری، برای استخراج ویژگی‌های مقاوم در برابر انسداد، از تابع هدفی با عنوان SPC^۲ استفاده می‌شود. همچنین، قسمت‌های غیرمسدود بدن با استفاده از بخش معنایی این شبکه بدست می‌آید. در نهایت ویژگی‌های محلی استخراجی از قسمت‌های غیرمسدود بدن و ویژگی‌های استخراجی از سراسر تصویر برای تطبیق بین تصاویر استفاده می‌شود.

در مرجع [۸۲] از شبکه عصبی ResNet-50 برای مقابله با مسئله انسداد در بازنشاسایی انسان استفاده شده است. در این مرجع تلاش شده است بدون نیاز به یادگیری ویژگی‌های ظاهری از قسمت‌های مختلف بدن، مدلی برای بازنشاسایی انسان ارائه شود که در آن با تمرکز بر ویژگی‌های استخراجی از سراسر بدن، مسئله انسداد کاهش یابد. از این‌رو، در این مرجع ابتدا یک روش افزایش داده مبتنی بر انسداد با هدف افزایش داده‌های آموزشی و همچنین ایجاد انسداد مصنوعی در تصاویر ورودی، ارائه شده است. همچنین به منظور تمرکز شبکه عصبی مورد استفاده بر بدن افراد، در این مرجع تغییراتی در لایه ادغام^۳ شبکه عصبی ایجاد می‌شود. علاوه بر این، برای آموزش شبکه مورد نظر، تابع هزینه‌ای به نام تابع فاصله نمایی محدود^۴ پیشنهاد شده است. در نهایت با استفاده از داده‌های آموزشی اولیه و داده‌های تولید شده و با کمینه‌سازی تابع هزینه پیشنهادی، شبکه عصبی مورد نظر با هدف تمرکز بر نواحی غیرمسدود و متمایز افراد آموزش می‌بیند.

¹ Cross-Graph Embedded Alignment (CGEA) layer

² Spatial-Patch Contrastive loss (SPC)

³ Pooling layer

⁴ Bounded Exponential Distance (BED)

در مرجع [۸۳] یک شبکه عصبی با عنوان SRNet^۱ ارائه شده است که در آن تلاش می‌شود ویژگی‌های سراسری متمایز از تصویر استخراج شود. در ساختار شبکه عصبی ارائه شده در این مرجع، برای حذف ویژگی‌های بی‌اهمیت استخراجی از نواحی مسدود تصویر، از ماژولی به نام Deep Residual Shrinkage استفاده شده است. همچنین در مرحله تطبیق بین نواحی مختلف تصویر، به‌منظور کاهش اثر ویژگی‌های استخراجی از نواحی مسدود، از ماژولی به نام ماژول وزن‌دهی مجدد، استفاده می‌شود. این ماژول برای وزن‌دهی مجدد ویژگی‌های استخراجی از هر ناحیه به میزان تمرکزی که شبکه به نواحی مختلف تصویر داشته است، توجه می‌کند. با این کار در مرحله تطبیق بین نواحی مختلف تصویر، به نواحی تصویر که شبکه تمرکز بیشتری بر آنها دارد وزن بیشتری تعلق می‌گیرد. در ادامه با در نظر گرفتن مسئله تطبیق بین ویژگی‌های استخراجی از تصاویر، به‌عنوان مسئله تطبیق بین گراف‌ها، میزان شباهت بین ویژگی‌های وزن‌دار تصاویر بررسی می‌شود.

در [۸۴]، سیستمی برای بازشناسایی انسان با توجه به مسئله انسداد ارائه شده است. این سیستم به‌منظور ادغام و استفاده از چندین ویژگی استخراجی توسط شبکه‌های عصبی مختلف، از سه قسمت تشکیل شده است. قسمت اول و دوم این سیستم هر کدام شامل یک شبکه عصبی برای استخراج ویژگی از تصاویر است. با این تفاوت که در قسمت دوم این سیستم استخراج ویژگی از تصویر با استفاده از اطلاعات موقعیت بدن انجام می‌شود. قسمت سوم این سیستم شامل یک شبکه عصبی است که از آن برای ادغام ویژگی‌های استخراجی از نواحی مختلف تصویر استفاده می‌شود.

در مرجع [۸۵]، به‌منظور استخراج ویژگی‌های مناسب و متمایز از تصاویر بدون انسداد و تصاویر حاوی انسداد، یک چارچوب یادگیری تخصصی^۲ ارائه شده است. در این چارچوب از دو شبکه عصبی پیچشی برای یادگیری ویژگی‌های تصویر اصلی و ویژگی‌های تصویر حاوی انسداد استفاده می‌شود. در این مرجع با توجه به برجسته‌ترین ناحیه موجود در نقشه توجه حاصل از شبکه عصبی عمیق، نواحی متمایز تصویر

^۱ Shrinking and Reweighting Network (SRNet)

^۲ Adversarial learning

شناسایی می‌شود. سپس با ایجاد انسداد در نواحی متمایز، تصاویر دارای نواحی مسدود تولید می‌شود. در ادامه به منظور تمرکز شبکه عصبی به سایر نواحی تصویر، از تصاویر مسدود تولید شده، حین فرآیند آموزش استفاده می‌شود. همچنین از یک ماژول تفکیک کننده^۱ برای تمایز بین ویژگی‌های استخراجی از تصاویر اصلی و تصاویر حاوی انسداد استفاده می‌شود که در آن ویژگی‌های استخراجی از تصاویر مسدود برای آموزش بهتر ماژول تفکیک کننده استفاده می‌شود.

در مرجع [۸۶] سیستمی برای ردیابی برخط^۲ و همزمان چند فرد در تصویر ارائه شده است. این سیستم شامل دو ماژول کلیدی با عنوان ماژول بازشناسایی بدون نظارت و ماژول تخمین انسداد در تصاویر است. در این مرجع، با توجه به این نکته که افراد با هویت یکسان در فریم‌های مجاور، ظاهری مشابه دارند، تابع هزینه‌ای مبتنی بر تطابق بدون نظارت بین فریم‌های مجاور ارائه می‌شود. از این تابع هزینه، برای آموزش ماژول بازشناسایی بدون نظارت استفاده می‌شود. همچنین با فرض اینکه در محیط تحت نظارت انسدادهای شدیدی وجود دارد، ماژول تخمین انسداد با تولید نقشه‌ای به نام نقشه انسداد، نواحی احتمالی وقوع انسداد در تصویر را پیش‌بینی می‌کند.

در مرجع [۸۷] برای بهبود عملکرد سیستم‌های بازشناسایی، مسئله پس‌زمینه‌های شلوغ در نظر گرفته شده است. در این رویکرد ابتدا مناطق پیش‌زمینه و پس‌زمینه تصاویر با استفاده از روش‌های قسمت‌بندی بدن انسان^۳ از هم جدا می‌شوند. سپس با ارائه یک شبکه عصبی عمیق مبتنی بر قسمت‌های مختلف بدن افراد سعی می‌شود تا ویژگی‌های متمایزی از پیش‌زمینه تصاویر (بدن افراد) استخراج شود. با این رویکرد شبکه عصبی وادار می‌شود به تمرکز روی نواحی از تصویر که حاوی اطلاعات سودمند است.

در روش ارائه شده در مرجع [۸۸]، با استفاده از یک روش انتخاب تکه مبتنی بر تجزیه و تشخیص برجستگی^۴ تکه‌ها، تکه‌های مربوط به پس‌زمینه تصاویر حذف می‌شوند. در این روش ابتدا از شبکه تجزیه

¹ Discriminator

² Online

³ Human parsing approaches

⁴ Parsing and saliency detection

عمیق^۱ برای قسمت‌بندی مفهومی^۲ تصاویر استفاده می‌شود. سپس، از دو روش پنجره لغزان و تطبیق رنگ برای حذف تکه‌های پس‌زمینه از تصاویر قسمت‌بندی استفاده می‌شود. در این رویکرد، تکه‌های مناسب با استفاده از رویکرد تشخیص برجستگی از بین تکه‌های باقی مانده انتخاب می‌شوند. همچنین ویژگی‌هایی مانند هیستوگرام هرمی گرادیان‌های جهت‌دار^۳ [۹۲]، مؤلفه‌های مدل رنگی HSV و ویژگی SIFT از تکه‌های انتخاب شده استخراج می‌شوند. سپس، ویژگی‌های استخراج شده محلی با ویژگی سراسری استخراج شده توسط توصیفگر وقوع بیشینه محلی ترکیب می‌شوند.

در [۸۹]، راهکاری برای یادگیری ویژگی‌های متمایز از مناطق بدن انسان ارائه شده است. در این راهکار از اطلاعات ساختاری بدن (اطلاعات خوشه مانند) استفاده می‌شود. برای این منظور، با در نظر گرفتن ویژگی‌های ظاهری به‌صورت بردار، ارتباط بین بردارهای ویژگی مدل می‌شود. سپس، این ارتباطات به‌عنوان یک بردار در نظر گرفته می‌شود که اطلاعات ساختاری سراسری را نشان می‌دهد. در نهایت ویژگی‌های ظاهری و ساختاری بدن برای تعیین اهمیت ویژگی‌های استخراجی استفاده می‌شود.

در مرجع [۹۰]، برای مقابله با مسئله پس‌زمینه شلوغ، شبکه‌ای با عنوان شبکه سوگیری پس‌زمینه محلی-سراسری^۴ ارائه شده است. ساختار این شبکه شامل سه قسمت است. قسمت اول ماژولی برای قطعه‌بندی جزئی نواحی پیش‌زمینه^۵، قسمت دوم ماژولی با عنوان ماژول نظارت بر هم‌ترازی سراسری^۶ تصویر و قسمت سوم ماژولی با عنوان ماژول محدودیت سوگیری پس‌زمینه^۷ است. در این مرجع برای کاهش اثرات نواحی پس‌زمینه بر ویژگی‌های استخراجی از تصویر، با استفاده از قسمت اول شبکه نواحی پیش‌زمینه تصویر به‌صورت محلی قطعه‌بندی می‌شود. سپس در قسمت دوم شبکه، ویژگی‌های سراسری، ویژگی‌های محلی و ویژگی‌های استخراجی از پس‌زمینه و پیش‌زمینه با هم ادغام می‌شوند. از ویژگی‌های ادغام شده برای هدایت شبکه در یادگیری ویژگی‌های محلی استفاده می‌شود. در نهایت با استفاده از

¹ Deep Decompositional Network (DNN)

² Semantically segmentation

³ Pyramid Histogram of Oriented Gradients (PHOG)

⁴ Global-Local Background_bias Net (GLBN)

⁵ Foreground Partial Segmentation Net (FPSN)

⁶ Global Aligned Supervision Net (GASN)

⁷ Background_bias Constraint Net (BCN) module

قسمت سوم شبکه، با در نظر گرفتن محدودیت روی فاصله و ارتباط بین ویژگی‌های استخراجی از پیش‌زمینه و ویژگی‌های استخراجی از پس‌زمینه، تاثیر ویژگی‌های استخراج شده از پس‌زمینه بر ویژگی‌های استخراج شده نهایی کاهش می‌یابد.

لازم به ذکر است که رویکردهای موجود در دسته دوم هنگام استخراج ویژگی‌های ظاهری افراد، مسائل انسداد ناشی از حمل اشیاء و پس‌زمینه‌های شلوغ را به‌صورت همزمان در نظر نمی‌گیرند. با توجه به نکات مطرح شده، در این پژوهش می‌خواهیم برای بهبود عملکرد سیستم‌های بازشناسی انسان، علاوه بر توجه به ویژگی‌های ظاهری افراد، مسئله انسداد ناشی از حمل اشیاء و همچنین مسئله پس‌زمینه‌های شلوغ نیز مورد توجه قرار گیرد. در فصل بعد ابعاد مسئله و توضیحات بیشتر درباره راهکار پیشنهادی ارائه می‌شود.

۲-۴- جمع‌بندی

در این فصل به بررسی و نقد اجمالی برخی از سیستم‌های بازشناسایی موجود پرداختیم. سیستم‌های بازشناسایی مورد بررسی به دو دسته تقسیم شدند. در دسته اول، رویکردهای موجود، تأثیرات مخرب انسداد بدن و پس‌زمینه‌های شلوغ را بر ویژگی‌های استخراجی کرده در نظر نمی‌گیرند. در حالی که در دسته دوم، رویکردهای موجود تلاش می‌کنند سیستم‌های بازشناسایی را در برابر تغییرات ظاهری مربوط به انسداد جزئی و پس‌زمینه‌های شلوغ مقاوم کنند.

ضمناً رویکردهای بازشناسایی موجود در دسته اول با توجه به ویژگی‌های ظاهری سعی در بهبود عملکرد سیستم‌های بازشناسایی در برابر مسائلی مانند تغییرات ظاهری ناشی از تغییر در حالت و موقعیت بدن افراد نسبت به دوربین و تغییر در روشنایی محیط، شرایط مختلف دوربین‌ها، ناهماهنگی توزیع داده‌ها در نمای دوربین‌ها و همچنین موضوع کمبود داده‌های آموزشی دارند. علاوه بر این تعدادی از روش‌های موجود در این دسته با تمرکز روی ویژگی‌های منحصربفرد افراد سعی در بهبود عملکرد سیستم‌های بازشناسایی دارند. این در حالی است که حرکت افراد مقابل یک دوربین یا بین دوربین‌های مختلف، همچنین انسداد تمام یا بخشی از بدن افراد ناشی از حمل اشیاء، ممکن است منجر به ناپدید شدن

ویژگی‌های ظاهری و ویژگی‌های منحصر بفرد در افراد شود. بنابراین بهتر است برای بهبود عملکرد سیستم‌های بازشناسایی به مسئله انسداد ناشی از حمل اشیاء هم توجه شود.

همچنین، اشاره شد که فضای پس‌زمینه ممکن است با حرکت افراد مقابل یک دوربین یا بین دوربین‌های مختلف دچار تغییرات اساسی شود. از این‌رو، استفاده از ویژگی‌های ظاهری موجود در فضای پس‌زمینه، منجر به کاهش عملکرد سیستم بازشناسایی می‌شود.

در روش‌های بازشناسایی موجود در دسته دوم، تلاش می‌شود سیستم‌های بازشناسایی در برابر مسائل مربوط به انسداد جزئی بدن و پس‌زمینه‌های شلوغ مقاوم شوند. اما رویکردهای موجود در دسته دوم هنگام استخراج ویژگی‌های ظاهری افراد، مسائل انسداد ناشی از حمل اشیاء و پس‌زمینه‌های شلوغ را به‌صورت همزمان در نظر نمی‌گیرند.

با توجه به موارد ذکر شده در این فصل، در فصل بعدی به ارائه راهکاری برای بهبود سیستم بازشناسایی انسان می‌پردازیم که در آن علاوه بر ویژگی‌های ظاهری افراد، مسئله انسداد ناشی از حمل اشیاء و همچنین مسئله پس‌زمینه‌های شلوغ، نیز در نظر گرفته می‌شود.

فصل سوم

روش پیشنهادی

۳-۱- مقدمه

با توجه به نکات مطرح شده و نقدهای وارد بر روش‌های بازشناسایی موجود، در این فصل به معرفی سیستم بازشناسایی پیشنهادی می‌پردازیم. همان‌طور که اشاره شد، در این سیستم برای بازشناسایی افراد، علاوه بر توجه به ویژگی‌های ظاهری مورد استفاده در سایر سیستم‌های بازشناسایی، مسئله انسداد ناشی از حمل اشیاء در افراد و همچنین مسئله پس‌زمینه شلوغ هم مورد توجه قرار خواهد گرفت. در ادامه این فصل ابتدا مجدداً فرض‌های اولیه موجود در سیستم پیشنهادی ارائه می‌شود. سپس با توجه به فرض‌های اولیه، جزئیات سیستم بازشناسایی پیشنهادی بررسی خواهد شد.

۳-۲- سیستم بازشناسایی پیشنهادی

همان‌طور که در فصل اول اشاره شد، در سیستم بازشناسایی پیشنهادی می‌خواهیم بین افرادی که در یک محیط عمومی (فضای داخلی یا خارجی ساختمان) قرار دارند، در یک بازه زمانی کوتاه عمل بازشناسی افراد را انجام دهیم. وجود چندین دوربین با دیدهای مجزا، ورود افراد جدید از طریق درگاه‌های ورودی (دوربین‌های مشخص) و وجود چندین فرد به‌طور هم‌زمان مقابل دوربین، از جمله فرض‌های اولیه سیستم بازشناسایی پیشنهادی است.

همان‌طور که قبلاً اشاره شد، در سیستم بازشناسایی انسان، حرکت فرد مقابل زاویه دید یک دوربین یا بین زاویه دید دوربین‌های مختلف، ممکن است باعث تغییر زاویه فرد نسبت به دوربین شود. تغییر زاویه فرد نسبت به دوربین، ممکن است منجر به تغییر در ویژگی‌های ظاهری فرد شود.

به‌عنوان مثال فرض کنید در یک سیستم بازشناسایی بخواهیم از هیستوگرام رنگ لباس به‌عنوان ویژگی متمایزکننده افراد از هم استفاده کنیم. در این سیستم فرض کنید فردی که مقابل یک دوربین در حرکت است، در حال حمل یک شیء با رنگ متفاوت از پوشش خود باشد. در این صورت پوشش این فرد وقتی که توسط شیء مسدود نشده باشد، دارای یک رنگ خاص و در نتیجه دارای یک هیستوگرام رنگ مشخص خواهد بود. این در حالی است که با حرکت فرد نسبت به دوربین و مسدود شدن بخشی

از پوشش فرد توسط شیء، ویژگی‌های ظاهری فرد تغییر می‌کند. نتیجه این امر این است که هیستوگرام رنگ مربوط به فرد مورد نظر نیز دچار تغییراتی می‌شود.

با تغییر در ویژگی‌های ظاهری افراد، توانایی سیستم بازناسایی کاهش یافته و ممکن است برچسب افراد را به درستی تعیین نکند یا برخی افراد را به عنوان افراد تازه وارد در نظر بگیرد. در واقع تغییر در ویژگی‌های ظاهری فرد ممکن است باعث گم شدن فرد حین فرآیند بازناسایی شود. از این رو برای بهبود سیستم بازناسایی انسان باید از ویژگی‌هایی استفاده کرد که نسبت به تغییرات ظاهری افراد مقاوم باشد.

در این راستا می‌توان از توصیفگرهای مناسبی مانند توصیفگر گوسینِ گوسین (GoG) [۴۶] و توصیفگر گوسینِ سلسله‌مراتبی (HGD) [۴۷] استفاده کرد. در فصل دوم، اشاره شد که با استفاده از توصیفگر GoG و توصیفگر HGD تصاویر افراد براساس میانگین و واریانس پیکسل‌ها توصیف می‌شوند. ویژگی‌های اولیه مورد استفاده در این توصیفگرها شامل اطلاعاتی مانند مختصات هر یک از پیکسل‌های تصویر در راستای محور عمودی، اندازه گرادیان تصویر در زاویه‌های مختلف و مقدار پیکسل‌های تصویر در مدل‌های رنگی مختلف RGB، Lab، HSV، و nRnG هستند.

در این روش توجه به مختصات پیکسل‌های تصویر در راستای محور عمودی، منجر به مقاومت توصیفگرهای GoG و HGD در برابر تغییرات مکان افراد در راستای محور عمودی می‌شود. همچنین استفاده از اندازه گرادیان تصویر در زاویه‌های مختلف، مقاومت این توصیفگرها را در برابر تغییرات زاویه‌ای افراد افزایش می‌دهد. از طرف دیگر با استفاده از مدل‌های رنگی مختلف در بردار ویژگی اولیه، تاثیر تغییرات روشنایی محیط در ویژگی‌های استخراج شده از تصویر افراد کاهش می‌یابد. از این رو در این پژوهش برای استخراج ویژگی از تصاویر افراد، از شکل بهبود یافته توصیفگرهای GoG و HGD استفاده می‌شود.

در ادامه این بخش ابتدا به بررسی نحوه عملکرد این دو توصیفگر می‌پردازیم. سپس در راستای بهبود عملکرد این توصیفگرها و در نتیجه بهبود عملکرد سیستم بازشناسی، راهکارهای پیشنهادی این رساله ارائه می‌شود.

۳-۲-۱- توصیفگر گوسین گوسین و گوسین سلسله‌مراتبی

در سیستم‌های بازشناسایی ارائه شده در [۴۶] و [۴۷]، ابتدا تصویر فرد به صورت سطری به هفت ناحیه همپوشان تقسیم می‌شود. سپس، هر کدام از این ناحیه‌ها به صورت تعدادی تکه متراکم در نظر گرفته می‌شود. هر یک از پیکسل‌های موجود در تکه‌ها، با استفاده از بردار ویژگی هشت‌تایی به صورت رابطه (۳-۱)، معرفی می‌شود:

$$f_i = (y, M_{0^\circ}, M_{90^\circ}, M_{180^\circ}, M_{270^\circ}, R, G, B)^T \quad (۳-۱)$$

در این رابطه، y مختصات پیکسل مورد نظر در راستای محور عمودی، M_{i° اندازه گرادیان در جهت زاویه i° و مؤلفه‌های R ، G و B ، به ترتیب مقدار پیکسل مربوطه در سه کانال رنگی قرمز، سبز و آبی را نشان می‌دهد.

در این مراجع، توزیع گوسین تکه‌های موجود در هر یک از هفت ناحیه از تصویر، با استفاده از بردار ویژگی ایجاد شده برای هر پیکسل از تصویر، توسط رابطه (۳-۲) محاسبه می‌شود:

$$\mathcal{N}(f; \mu_s, \Sigma_s) = \frac{\exp\left(-\frac{1}{2}(f - \mu_s)^T \Sigma_s^{-1}(f - \mu_s)\right)}{(2\pi)^{\frac{d}{2}} |\Sigma_s|} \quad (۳-۲)$$

در این رابطه، μ_s و Σ_s به ترتیب ماتریس کوواریانس و میانگین بردارهای ویژگی استخراجی از پیکسل‌های تکه s است. همچنین d اندازه بردار ویژگی می‌باشد، و عملگر $|\cdot|$ نشان‌دهنده عمل دترمینان است. این کار هر یک از هفت ناحیه سطری تصویر، توسط تعدادی توزیع گوسین توصیف می‌شود. این توزیع‌های گوسین روی یک خمینه ریمانی قرار دارد و عملیات اقلیدسی را نمی‌توان مستقیماً روی این

فضا اعمال کرد. از طرفی یکی از فضاهاى ریمانی، فضای ریمانی معین مثبت متقارن^۱ است [۹۳، ۹۴]. هر نقطه از این فضا را می‌توان با استفاده از لگاریتم ماتریس اصلی^۲ به فضای مماسی^۳ اقلیدسی نگاشت داد. از این‌رو در مراجع [۴۶] و [۴۷]، ابتدا فضای d بُعدی گوسین توسط رابطه (۳-۳) به فضای $d + 1$ بُعدی ماتریس‌های معین مثبت متقارن نگاشت می‌شود [۹۵]:

$$\mathcal{N}(f; \mu_s, \Sigma_s) \sim P_s = |\Sigma_s|^{-\frac{1}{d+1}} \begin{bmatrix} \Sigma_s + \mu_s \mu_s^T & \mu_s \\ \mu_s^T & 1 \end{bmatrix} \quad (3-3)$$

در واقع توسط رابطه (۳-۳) بردار میانگین و ماتریس کوواریانس هر تکه s ، به ماتریس معین مثبت متقارن نگاشت می‌شود. سپس ماتریس P_s حاصل، توسط رابطه (۴-۳) به فضای مماسی اقلیدسی نگاشت شده و به برداری با تعداد $m = \frac{(d^2+3d)}{2} + 1$ مؤلفه تبدیل می‌شود:

$$g_s = \text{vec}(\log(P_s)) \\ = [b_{s(1,1)}, \sqrt{2}b_{s(1,1)}, \dots, \sqrt{2}b_{s(1,d+1)}, b_{s(2,2)}, \sqrt{2}b_{s(2,3)}, \dots, b_{s(d+1,d+1)}]^T \quad (4-3)$$

در این رابطه، عملگر $\log(\cdot)$ نشان‌دهنده عملگر لگاریتم ماتریس است. با توجه به اینکه حاصل $\log(P_s)$ یک ماتریس متقارن است، در این رابطه، تابع vec عناصر بالامثلثی ماتریس ورودی را به بردار تبدیل می‌کند.

در واقع با استفاده از رابطه (۴-۳)، هر تکه از تصویر، به برداری با m مؤلفه تبدیل می‌شود. در نتیجه هر کدام از نواحی تصویر با توجه به تعداد تکه‌های موجود در آن، شامل تعدادی بردار m مؤلفه‌ای خواهد بود.

از بردارهای مربوط به هر کدام از هفت ناحیه تصویر، برای محاسبه توزیع گوسین ناحیه مربوطه استفاده می‌شود. در این راستا از رابطه‌ای مشابه با رابطه (۲-۳) استفاده می‌شود. برای محاسبه میانگین و واریانس هر ناحیه از تصویر و جایگذاری آن در رابطه (۲-۳)، به ترتیب از روابط (۳-۵) و (۳-۶) استفاده می‌شود:

¹ Symmetric Positive Definite (SPD)

³ Tangent space

² Principle matrix log

$$\mu^G = \frac{1}{\sum_{s \in G} w_s} \sum_{s \in G} w_s g_s \quad (5-3)$$

$$\Sigma^G = \frac{1}{\sum_{s \in G} w_s} \sum_{s \in G} w_s (g_s - \mu^G)(g_s - \mu^G)^T \quad (6-3)$$

در این روابط، G نشان‌دهنده یکی از هفت ناحیه موجود در تصویر و μ^G و Σ^G به ترتیب نشان‌دهنده میانگین و واریانس ناحیه مربوطه است. همچنین ضریب w_s مورد استفاده در این روابط، برای وزندهی تکه‌های مختلف تصویر استفاده می‌شود.

با فرض اینکه فرد در مرکز تصویر قرار دارد، مقدار ضریب w_s برای هر تکه s ، با توجه به فاصله تکه مورد نظر تا مرکز تصویر، توسط رابطه (۷-۳) تعیین می‌شود:

$$w_s = \exp\left(\frac{-\frac{1}{2}(x_s - x_c)^2}{\sigma^2}\right) \quad (7-3)$$

در این رابطه $x_c = \frac{L}{2}$ ، $\sigma = \frac{L}{4}$ ، x_s مختصات محور افقی پیکسل مرکزی در تکه s و L عرض تصویر را نشان می‌دهد. در واقع طبق این رابطه هر چه تکه s ، از مرکز تصویر دورتر باشد مقدار ضریب w_s ، کوچکتر خواهد بود و بر عکس. با این کار در محاسبه میانگین و واریانس هر یک از هفت ناحیه تصویر، تاثیر تکه‌هایی که در پس‌زمینه تصویر قرار دارند، کاهش می‌یابد.

پس از محاسبه توزیع گوسین هر یک از هفت ناحیه موجود در تصویر، فضاهاى گوسین حاصل، توسط روابط (۳-۳) و (۴-۳) به فضای برداری با $1 + \frac{(m^2+3m)}{2}$ مؤلفه نگاشت می‌شود. در نهایت با فرض اینکه بردار توصیفگر ناحیه G را z_G بنامیم، با کنار هم قرار دادن بردارهای توصیفگر هر یک از هفت ناحیه تصویر، بردار ویژگی برای کل تصویر توسط رابطه (۸-۳) حاصل می‌شود [۴۶] و [۴۷]:

$$Z_{RGB} = [z_{G1}^T, z_{G2}^T, z_{G3}^T, z_{G4}^T, z_{G5}^T, z_{G6}^T, z_{G7}^T]^T \quad (8-3)$$

در این مراجع، به منظور افزایش دقت بردار ویژگی حاصل، برای ایجاد بردار ویژگی اولیه f_i که در رابطه (۱-۳) نشان داده شد، علاوه بر مدل رنگی RGB، از مدل‌های رنگی Lab، HSV و nRnG استفاده می‌شود. در نتیجه چهار بردار ویژگی متفاوت برای هر تصویر بدست می‌آید. فرض کنید بردارهای ویژگی استخراج شده با استفاده از مدل رنگی RGB، Lab، HSV و nRnG به ترتیب Z_{RGB} ، Z_{LAB} ، Z_{HSV} و Z_{nRGB} نامگذاری شوند. در این صورت با کنار هم قرار دادن این چهار بردار ویژگی، بردار ویژگی نهایی توسط رابطه (۹-۳) بدست می‌آید:

$$Z_{Fusion} = [Z_{RGB}^T, Z_{LAB}^T, Z_{HSV}^T, Z_{nRGB}^T]^T \quad (9-3)$$

در [۴۶] و [۴۷]، بردارهای ویژگی استخراج شده از تصاویر یعنی Z_{Fusion} ، برای آموزش معیار فاصله XQDA^۱ [۴۵] استفاده می‌شود. در مرجع [۴۵]، آموزش معیار فاصله XQDA به طور همزمان با کاهش بُعد ویژگی‌های استخراجی از تصاویر به زیر فضایی دیگر انجام می‌شود. خروجی معیار فاصله XQDA یک ماتریس به نام ماتریس نمره^۲ است. با فرض اینکه M تصویر در مجموعه آزمون و N تصویر در مجموعه گالری وجود داشته باشد، ابعاد این ماتریس برابر $M \times N$ است به طوری که هر عنصر ماتریس نمره، فاصله بین تصاویر مربوطه را در مجموعه آزمون و گالری نشان می‌دهد. بنابراین، با مرتب‌سازی هر سطر از ماتریس نمره، تصاویر گالری متناسب با فاصله‌شان تا تصویر آزمون مربوطه رتبه‌بندی می‌شوند.

۳-۲-۲- بهبود عملکرد سیستم بازناسایی

در این پژوهش برای بهبود عملکرد توصیفگرهای GoG و HGD و در نتیجه بهبود عملکرد سیستم‌های بازناسایی انسان، تلاش می‌شود با استفاده از نقشه‌های وزن پیشنهادی، تاثیر پیکسل‌های تصویر بر بردار ویژگی استخراجی توسط توصیفگرهای GoG و HGD، متناسب با تعلق پیکسل‌ها به پس‌زمینه و پیش‌زمینه و با توجه به میزان اهمیت آنها در مسئله بازناسایی تعیین شود. در ادامه این بخش، جزئیات راهکارهای پیشنهادی برای بهبود عملکرد سیستم بازناسایی انسان مورد بررسی قرار می‌گیرد.

¹ Cross-view Quadratic Discriminant Analysis (XQDA)

² Score matrix

۳-۲-۱- وزن‌دهی ویژگی‌ها با استفاده از توزیع گوسین و تابع چگالی احتمال^۱

در راهکار پیشنهادی اول، وزن‌دهی ویژگی‌های استخراجی به نحوی انجام می‌شود که به نواحی مرتبط با بدن فرد وزن بیشتری تعلق بگیرد. در این راهکار با استفاده از توزیع گوسین و تابع چگالی احتمال، روشی برای قطعه‌بندی معنایی تصاویر ارائه می‌شود که در آن پیکسل‌هایی با اهمیت یکسان در یک قطعه قرار می‌گیرند. لازم به ذکر است که روش قطعه‌بندی پیشنهادی، روشی بدون ناظر و بدون نیاز به مرحله آموزش و داده‌های برچسب‌گذاری شده است. در این رویکرد به ازای هر یک از تصاویر ورودی، یک نقشه وزن با ابعادی برابر با تصویر محاسبه می‌شود، به طوری که هر یک از پیکسل‌های نقشه وزن، نشان‌دهنده تاثیر پیکسل متناظر از تصویر ورودی بر ویژگی‌های نهایی استخراج شده است.

همان‌طور که اشاره شد، نقشه وزن را می‌توان به‌عنوان یک تصویر قطعه‌بندی شده در نظر گرفت که در آن پیکسل‌ها با اهمیت یکسان به یک قطعه اختصاص داده می‌شوند. یکی از مدل‌های رنگی رایج برای قطعه‌بندی تصاویر رنگی، مدل رنگی Lab است [۹۶]. این مدل یک فضای رنگی یکنواخت است [۹۷]، که در آن تفاوت بین مقادیر رنگ متناسب با تفاوت رنگ‌های ادراک شده توسط انسان می‌باشد. این ویژگی، مدل رنگی Lab را به یک فضای رنگ مناسب برای قطعه‌بندی تصاویر رنگی تبدیل می‌کند. بنابراین، در روش پیشنهادی از مدل رنگ Lab برای بدست آوردن نقشه وزن استفاده می‌شود.

در شرایط واقعی، معمولاً رنگ پوشش افراد با رنگ پس‌زمینه متفاوت است. همچنین می‌توان صحنه طبیعی موجود در تصاویر را با استفاده از توزیع‌های گوسین مدل‌سازی کرد [۹۸]. فرض کنید پیکسل‌های یک تصویر ورودی با استفاده از توزیع گوسین مدل‌سازی شود. واضح است که تعداد تکرار هر پیکسل با مقدار سطح خاکستری مشخص، در توزیع گوسین بدست آمده تاثیر می‌گذارد. به‌عنوان مثال، در تصویر ورودی با فرض اینکه پیکسل‌ها در بازه [۰ و ۲۵۵] باشند، اگر تعداد پیکسل‌ها با مقدار سطح خاکستری نزدیک به صفر بیشتر از تعداد پیکسل‌ها با مقدار سطح خاکستری نزدیک به ۲۵۵ باشد،

^۱ Probability Density Function (PDF)

میانگین توزیع گوسین به مقادیر نزدیک به صفر متمایل می‌شود. از این‌رو، احتمال اینکه پیکسل با مقدار سطح خاکستری نزدیک به صفر متعلق به این توزیع باشد زیاد است و برعکس احتمال اینکه پیکسل با مقدار سطح خاکستری نزدیک به ۲۵۵ متعلق به این توزیع باشد کم است. بنابراین، در رویکرد پیشنهادی با فرض تفاوت بین مقادیر پیکسل‌های پس‌زمینه و پوشش افراد، پیکسل‌های تصویر را با استفاده از توزیع‌های گوسین مدل می‌کنیم. در این راستا با فرض اینکه اندازه تصویر ورودی $m \times n \times 3$ باشد، روش وزن‌دهی پیشنهادی شامل سه مرحله به شرح زیر است:

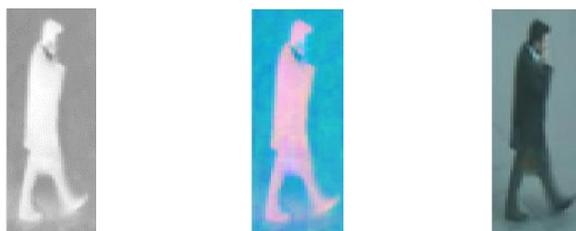
(۱) در مرحله اول، ابتدا تصویر RGB ورودی به مدل رنگی Lab انتقال می‌یابد. سپس، با در نظر گرفتن مؤلفه‌های L ، a و b این تصویر به صورت یک بردار $1 \times k$ ، توزیع گوسین تصویر را محاسبه می‌کنیم. توجه داشته باشید که k تعداد پیکسل‌های موجود در سه مؤلفه تصویر، یعنی $k = m \times n \times 3$ است.

(۲) در مرحله دوم، برای هر پیکسل تصویر، تابع چگالی احتمال با توجه به میانگین و واریانس توزیع گوسین بدست آمده از مرحله اول، محاسبه می‌شود. از این‌رو، برای هر پیکسل از مؤلفه‌های تصویر ورودی، یک مقدار بدست می‌آید که احتمال تعلق پیکسل مربوطه را به توزیع گوسین نشان می‌دهد. بنابراین، خروجی‌های مرحله دوم را می‌توان به صورت یک تصویر با اندازه تصویر ورودی (یعنی $m \times n \times 3$) در نظر گرفت. در این رساله این تصویر را W_{LAB} می‌نامیم. پیکسل‌های موجود در مؤلفه‌های تصویر W_{LAB} یعنی W_L ، W_A و W_B نشان‌دهنده احتمال تعلق پیکسل‌های موجود در هر یک از مؤلفه‌های L ، a یا b به توزیع گوسین تصویر ورودی است.

(۳) در مرحله سوم، توسط رابطه (۱۰-۳) تصویر W_{LAB} به یک تصویر خاکستری نرمال با ابعاد $m \times n$ تبدیل می‌شود. در این رابطه \widehat{W} نقشه وزن پیشنهادی است:

$$\widehat{W} = \frac{W_L + W_A + W_B}{\max(W_L + W_A + W_B)} \quad (10-3)$$

در شکل (۱-۳)، یک نمونه تصویر از پایگاه داده PRID450s [۲۵]، به همراه تصویر W_{LAB} و نقشه وزن پیشنهادی (\widehat{W}) مربوطه نشان داده شده است.



الف) نمونه تصویر W_{LAB} (ب) تصویر ورودی، تصویر W_{LAB} و نقشه وزن (\hat{W}) (ج) شکل (۱-۳): یک نمونه تصویر ورودی، تصویر W_{LAB} و نقشه وزن (\hat{W})

همان‌طور که در شکل (۱-۳) مشاهده می‌شود، در تصویر نقشه وزن پیشنهادی (\hat{W}) ، سطوح خاکستری با مقادیر بالا (پیکسل‌های روشن‌تر) به بدن فرد اختصاص داده شده است. این امر نشان می‌دهد که در فضای رنگی Lab، پیکسل‌های پس‌زمینه در مقایسه با پیکسل‌های مرتبط با بدن شخص، احتمال تعلق کمتری به توزیع گوسین بدست آمده دارند. در واقع در نقشه وزن پیشنهادی به پیکسل‌های بدن فرد نسبت به پیکسل‌های پس‌زمینه، وزن بیشتری تعلق می‌گیرد.

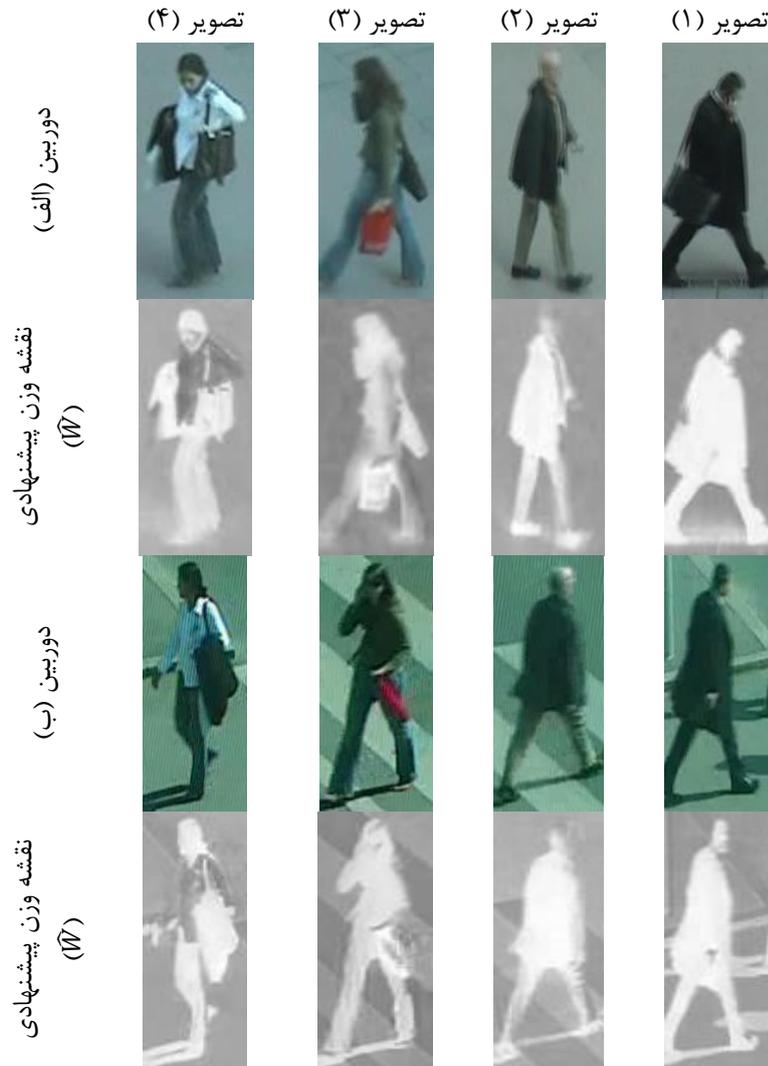
در این پژوهش، از نقشه وزن پیشنهادی (\hat{W}) ، برای وزن‌دهی ویژگی‌های اولیه استخراج شده (F_i) در رابطه (۱-۳) به صورت زیر استفاده می‌شود:

$$\hat{F}_i = \hat{W}_i \times F_i \quad (۱۱-۳)$$

در این رابطه \hat{W}_i وزن مربوط به پیکسل i ام تصویر ورودی و \hat{F}_i بردار ویژگی‌های استخراجی وزن‌دار هستند.

لازم به ذکر است که در نقشه وزن پیشنهادی، پیکسل‌های مربوط به اشیاء و نواحی منحصربفرد در مقایسه با سایر پیکسل‌ها، احتمال تعلق بیشتری به توزیع گوسین بدست آمده دارند. در واقع، مقادیر پیکسل‌های متعلق به اشیاء و نواحی منحصربفرد معمولاً با مقادیر پیکسل‌های مرتبط با پس‌زمینه و بدن فرد متفاوت است. از این‌رو، مقدار احتمال تعلق به توزیع گوسین بدست آمده، برای پیکسل‌های مربوط به اشیاء و نواحی منحصربفرد، متفاوت از سایر پیکسل‌های تصویر است. بنابراین، استخراج وزن‌دار ویژگی‌های اولیه از تصاویر با استفاده از نقشه وزن پیشنهادی، منجر به تمرکز بیشتر بر پیکسل‌های اشیاء و نواحی منحصربفرد می‌شود.

در شکل (۲-۳)، هشت نمونه تصویر از پایگاه داده PRID450s به همراه نقشه‌های وزن حاصل از روش پیشنهادی نشان داده شده است. لازم به ذکر است تصاویر نشان داده شده در هر ستون از این شکل، تصاویر مربوط به یک فرد مقابل دو دوربین متفاوت است.



شکل (۲-۳): نمونه تصاویر اخذ شده توسط دوربین (الف) و (ب) و نقشه‌های وزن حاصل از روش پیشنهادی (\hat{W})

همان‌طور که در شکل (۲-۳) مشاهده می‌شود، موقعیت بدن و پس‌زمینه یک فرد در زاویه دید دوربین‌های مختلف تغییر کرده است. این در حالی است که روش وزن‌دهی پیشنهادی تقریباً وزن‌های مشابهی برای قسمت‌های مختلف بدن یک فرد در دو زاویه دید متفاوت در نظر می‌گیرد.

لازم به ذکر است که علی‌رغم مزایای عنوان شده در مورد نقشه‌های وزن پیشنهادی، این نقشه‌ها ممکن است دقت کافی نداشته باشند. به‌عنوان مثال، در شکل (۳-۳)، یک نمونه تصویر از پایگاه داده VIPeR (نمونه تصویر (۱) [۱۵] و یک نمونه تصویر از پایگاه داده PRID450s (نمونه تصویر (۲))، به همراه نقشه‌های وزن پیشنهادی اول (\hat{W}) نشان داده شده است. در نقشه‌های وزن نشان داده شده در این شکل برخی از پیکسل‌های مربوط به بدن فرد مقادیر کمتری نسبت به پس‌زمینه دارند. از این‌رو، لازم است راهکار مورد استفاده برای بدست آوردن نقشه وزن بهبود یابد.



الف) نمونه تصویر (۱) ب) نقشه وزن تصویر (۱) ج) نمونه تصویر (۲) د) نقشه وزن تصویر (۲)
شکل (۳-۳): دو نمونه تصویر به همراه نقشه‌های وزن پیشنهادی اول (\hat{W})

۳-۲-۲-۲-۳- وزن‌دهی ویژگی‌ها با توجه به قطعه‌بندی معنایی ۱

برای بهبود دقت نقشه وزن پیشنهادی اول، در نقشه وزن پیشنهادی دوم، وزن‌دهی ویژگی‌های استخراجی با توجه به قطعه‌بندی معنایی تصاویر انجام می‌شود. در این راستا، ابتدا با استفاده از روش قطعه‌بندی مبتنی بر شبکه عمیق DeepLabv3+ [۹۹]، تصاویر را به‌صورت معنایی به سه ناحیه بدن فرد، اشیاء و پس‌زمینه قطعه‌بندی می‌کنیم. سپس با اختصاص عددی ثابت در بازه [۱ و ۰] به‌عنوان وزن پیکسل‌های مرتبط با هر ناحیه، سعی می‌شود ویژگی‌های استخراجی از پیکسل‌های نواحی با توجه اهمیت هر ناحیه در مسئله بازشناسایی به‌صورت وزن‌دار استخراج شود.

در روش قطعه‌بندی مبتنی بر شبکه عمیق DeepLabv3+ از ساختار رمزگذار-رمزگشا^۲ برای تقسیم‌بندی معنایی تصویر استفاده می‌شود. در این روش، اطلاعات زمینه‌ای^۳ مناسب با استفاده از ماژول رمزگذار بدست می‌آید. همچنین، مرزهای اشیاء^۴ با استفاده از ماژول رمزگشا استخراج می‌شوند. با توجه به دقت

^۱ Semantic segmentation

^۲ Encoder-decoder structure

^۳ Contextual information

^۴ Object boundaries

قابل قبول شبکه عصبی DeepLabv3+ در قطعه‌بندی تصاویر، در راهکار دوم برای قطعه‌بندی معنایی تصاویر به سه ناحیه بدن، اشیاء در حال حمل و پس‌زمینه، از این شبکه عصبی استفاده می‌شود. برای استفاده از روش قطعه‌بندی مبتنی بر شبکه عمیق DeepLabv3+ در نقشه وزن پیشنهادی، ابتدا این شبکه به‌منظور قطعه‌بندی تصاویر به سه ناحیه بدن فرد، اشیاء و پس‌زمینه آموزش داده می‌شود. در این راستا تعدادی از تصاویر پایگاه داده VIPeR که به‌صورت دستی قطعه‌بندی شده‌اند، به همراه تصاویر قطعه‌بندی شده از پایگاه داده PRID450s برای آموزش شبکه عمیق DeepLabv3+ مورد استفاده قرار گرفته است.

نمونه‌هایی از تصاویر پایگاه‌های داده CUHK01 (نمونه‌های ۱ و ۲) [۲۲] و VIPeR (نمونه‌های ۳ و ۴) به همراه نقشه‌های قطعه‌بندی معنایی^۱ (SSM) حاصل از شبکه عمیق DeepLabv3+ آموزش‌دیده، در شکل (۳-۴) نشان داده شده است. در نقشه‌های قطعه‌بندی معنایی (SSM) نشان داده شده در این شکل، پس‌زمینه، فرد و مناطق مسدود شده توسط اشیاء به ترتیب با استفاده از رنگ‌های سیاه، سفید و خاکستری نشان داده شده است.



شکل (۳-۴): نمونه‌هایی از تصاویر پایگاه‌های داده CUHK01 و VIPeR به همراه نقشه‌های قطعه‌بندی معنایی حاصل از شبکه عمیق DeepLabv3+

^۱ Semantic Segmented Map (SSM)

در این پژوهش، پس از بدست آوردن SSM برای تصویر ورودی، یک نقشه وزن به نام \widehat{W}_S به اندازه تصویر ورودی پیشنهاد می‌شود. برای هر پیکسل از تصویر ورودی، \widehat{W}_S شامل یک مقدار عددی در بازه [۱ و ۰) است، به طوری که این مقدار عددی تاثیر پیکسل مربوطه را بر بردار ویژگی استخراج شده نشان می‌دهد. در نقشه وزن پیشنهادی دوم مقادیر \widehat{W}_S با در نظر گرفتن SSM و با توجه به نکات زیر تعیین می‌شود:

الف) از آنجایی که ظاهر فرد یک مشخصه مهم برای بازشناسایی است، حداکثر مقدار عددی در بازه [۱ و ۰)، یعنی عدد ۱ به پیکسل‌هایی از \widehat{W}_S که مربوط به بدن فرد است، اختصاص می‌یابد.

ب) ممکن است اشیاء در حال حمل برخی از قسمت‌های بدن فرد را مسدود کرده و ویژگی‌های ظاهری را تحت تاثیر قرار دهند. از طرفی، در برخی مواقع اشیاء در حال حمل در نماهایی از دوربین مشخص هستند. از این‌رو، می‌توان از اشیاء در حال حمل به‌عنوان یک ویژگی کمکی برای بازشناسایی افراد استفاده کرد. بنابراین در راهکار پیشنهادی، برای پیکسل‌هایی از \widehat{W}_S که مربوط به نواحی مسدود شده توسط اشیاء در حال حمل هستند، مقدار عددی ۰٫۹ در نظر گرفته می‌شود. از این‌رو، اشیاء در حال حمل در مقایسه با بدن فرد تاثیر کمتری در بازشناسایی افراد خواهد داشت.

ج) بدیهی است که پس‌زمینه در مقایسه با بدن و اشیاء در حال حمل، اهمیت کمتری در بازشناسایی افراد دارد. بنابراین، مقدار عددی ۰٫۵ به پیکسل‌هایی از \widehat{W}_S که مربوط به پس‌زمینه است، اختصاص داده می‌شود.

در این پژوهش، از نقشه وزن پیشنهادی (\widehat{W}_S)، برای وزن‌دهی ویژگی‌های اولیه استخراج شده (F_i) در رابطه (۱-۳) به‌صورت زیر استفاده می‌شود:

$$\widehat{F}_i = \widehat{W}_S^i \times F_i \quad (۱۲-۳)$$

در این رابطه، \widehat{W}_S^i وزن مربوط به پیکسل i ام تصویر ورودی و \widehat{F}_i بردار ویژگی‌های استخراجی وزن‌دار هستند.

همان‌طور که اشاره شد، در نقشه وزن پیشنهادی دوم، به هر یک از نواحی بدن فرد، اشیاء و پس‌زمینه عددی ثابت اختصاص می‌یابد. این در حالی است که بهتر است وزن پیکسل‌های هر ناحیه، به‌صورت خودکار و با توجه به اندازه ناحیه و فاصله پیکسل‌ها تا مرکز تصویر تعیین شود.

۳-۲-۲-۳- وزن‌دهی ویژگی‌ها با توجه به اندازه نواحی و فاصله پیکسل‌ها تا مرکز تصویر

برای جلوگیری از وزن‌دهی دستی نواحی مختلف تصویر، در ادامه، نقشه وزنی پیشنهاد می‌شود که در آن ویژگی‌های استخراجی از پیکسل‌های مرتبط با اشیاء در حال حمل و پس‌زمینه با توجه به اندازه هر یک از نواحی و فاصله پیکسل‌های مربوطه تا مرکز تصویر، به‌صورت خودکار تنظیم می‌شود.

اشیاء بزرگ در مقایسه با اشیاء کوچک پیکسل‌های بیشتری را در بر گرفته و در نتیجه قسمت‌های وسیع‌تری از بدن فرد را پوشش می‌دهند. از این‌رو، اشیاء بزرگ نسبت به اشیاء کوچک اثرات مخرب بیشتری بر ویژگی‌های ظاهری فرد دارند. همچنین تعداد پیکسل‌های پس‌زمینه معمولاً بیشتر از تعداد پیکسل‌های مرتبط با اشیاء و بدن فرد است. از این‌رو، پیکسل‌های پس‌زمینه به خصوص پس‌زمینه‌های شلوغ، اثرات مخرب قابل‌توجهی بر ویژگی‌های استخراجی از تصاویر دارند.

با توجه به این موارد، در نقشه وزن پیشنهادی سوم برای هر تصویر ورودی یک نقشه وزن بر اساس اندازه اشیاء در حال حمل و پس‌زمینه، با عنوان W_s محاسبه می‌شود. مشابه با نقشه‌های وزن پیشنهادی اول و دوم، اندازه نقشه وزن W_s به اندازه تصویر ورودی است. همچنین مقادیر این نقشه در بازه عددی [۱ و ۰] است به‌طوری که مقادیر عددی بزرگتر، نشان‌دهنده اهمیت بیشتر پیکسل‌های مربوطه در بازشناسایی افراد است.

در رویکرد پیشنهادی، ابتدا تصویر ورودی به‌صورت دستی به سه ناحیه بدن فرد، پس‌زمینه و اشیاء در حال حمل تقسیم می‌شود. سپس، وزن پیکسل مربوط به هر ناحیه به‌صورت زیر تعیین می‌شود:

الف) با توجه به اهمیت پیکسل‌های مربوط به بدن در بازشناسایی افراد، مقدار عددی ۱ را به پیکسل‌هایی از W_s که مربوط به بدن است، اختصاص می‌دهیم.

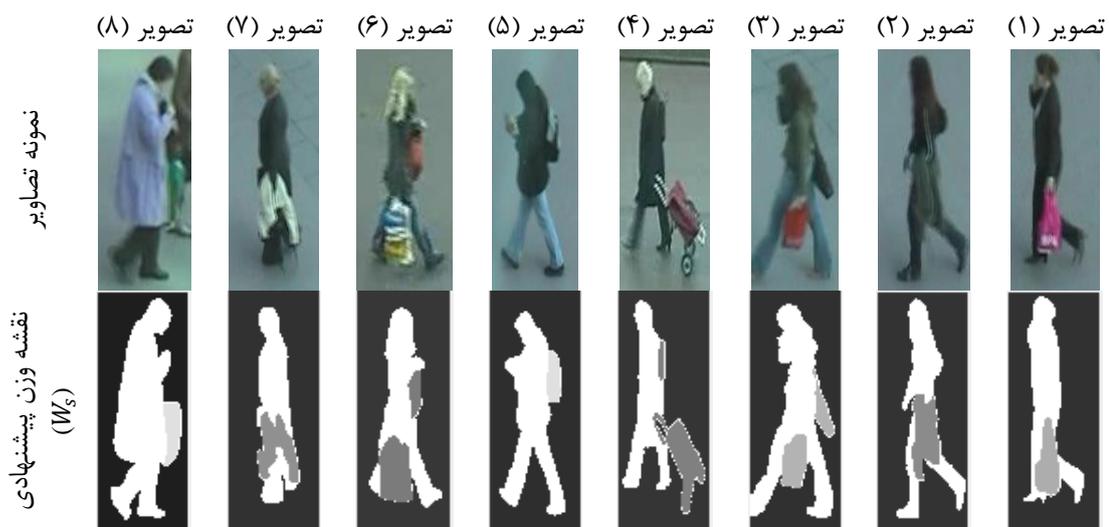
ب) با فرض بزرگتر بودن اندازه بدن فرد از پس‌زمینه و اشیاء در حال حمل، مقادیر عددی $W_{objects}$ و $W_{background}$ بدست آمده از رابطه‌های (۱۳-۳) و (۱۴-۳)، به ترتیب به پیکسل‌هایی از W_S که مربوط به اشیاء و نواحی پس‌زمینه هستند، نسبت داده می‌شود:

$$W_{objects} = \left(1 - \frac{N_O}{N_I}\right) \times \left(1 - \frac{N_O}{N_P}\right) \quad (13-3)$$

$$W_{background} = \left(1 - \frac{N_B}{N_I}\right) \times \left(1 - \frac{N_B}{N_P}\right) \quad (14-3)$$

در رابطه‌های (۱۳-۳) و (۱۴-۳)، N_O تعداد پیکسل‌های موجود در اشیاء، N_P تعداد پیکسل‌های موجود در بدن فرد، N_B تعداد پیکسل‌های موجود در پس‌زمینه و N_I تعداد کل پیکسل‌های تصویر است. توجه به این رابطه‌ها، تاثیر نواحی مربوط به اشیاء و پس‌زمینه بر ویژگی‌های ظاهری استخراجی، متناسب با اندازه این نواحی کاهش می‌یابد.

در شکل (۵-۳)، تعدادی نمونه تصویر از پایگاه داده PRID450s، به همراه نقشه‌های وزن مربوطه (W_S) نشان داده شده است. همان‌طور که در شکل (۵-۳) نشان داده شده است، در نقشه وزن پیشنهادی، به پیکسل‌های اشیاء بزرگتر نسبت به اشیاء کوچکتر مقادیر کمتری اختصاص داده می‌شود. همچنین پیکسل‌های پس‌زمینه کمترین وزن را نسبت به پیکسل‌های بدن فرد و اشیاء در حال حمل دارند.



شکل (۵-۳): تعدادی نمونه تصویر به همراه نقشه‌های وزن مربوطه (W_S)

همچنین در نقشه وزن پیشنهادی سوم، با فرض اینکه بدن فرد در مرکز تصویر ورودی قرار دارد، نقشه وزنی را براساس فاصله پس‌زمینه و اشیاء در حال حمل از مرکز تصویر پیشنهاد می‌کنیم. توجه داشته باشید که اشیاء در حال حملی که در مرکز تصویر قرار دارند، معمولاً قسمتی از بدن فرد را می‌پوشانند. از این‌رو، این اشیاء در مقایسه با اشیائی که در گوشه‌های تصویر قرار دارند، تأثیر مخرب بیشتری بر ویژگی‌های ظاهری فرد دارند. از طرفی، پیکسل‌هایی که نزدیک به مرکز تصویر (یعنی بدن فرد) هستند، مهم‌تر از پیکسل‌هایی هستند که از مرکز تصویر فاصله دارند.

با توجه به نکات ذکر شده، در راهکار سوم، ابتدا یک ماتریس فاصله (D) با اندازه تصویر ورودی محاسبه می‌شود که شامل مقدار فاصله پیکسل‌های متناظر در تصویر ورودی از مرکز تصویر است. برای بدست آوردن این ماتریس، فاصله اقلیدسی هر پیکسل تا مرکز تصویر با استفاده از روابط (۳-۱۵) تا (۳-۱۷) محاسبه می‌شود:

$$\hat{y}_k = \frac{|y_{center} - y_k|}{y_{center}} \quad (۳-۱۵)$$

$$\hat{x}_k = \frac{|x_{center} - x_k|}{x_{center}} \quad (۳-۱۶)$$

$$D_k = \sqrt{\hat{y}_k^2 + \hat{x}_k^2} \quad (۳-۱۷)$$

در رابطه‌های (۳-۱۶) و (۳-۱۷)، x_{center} و y_{center} به ترتیب مختصات x و y پیکسل مرکزی در تصویر ورودی هستند. همچنین y_k و x_k به ترتیب مختصات y و x پیکسل k تصویر و D_k فاصله اقلیدسی پیکسل مربوطه (پیکسل k) را نشان می‌دهند.

همچنین مقادیر ماتریس D با استفاده از رابطه (۳-۱۸) در بازه عددی $[۱$ و $۰)$ نرمال می‌شوند که در آن (M) بیشینه مقدار عددی موجود در ماتریس D است:

$$\hat{D} = \frac{D}{M} \quad (۳-۱۸)$$

با استفاده از \hat{D} و نقشه قطعه‌بندی شده دستی، نقشه وزن دیگری با عنوان W_d را در بازه عددی $[۱$ و $۰)$ و با اندازه یکسان با تصویر ورودی به صورت زیر پیشنهاد می‌شود:

الف) مشابه با W_s ، مقدار عددی ۱ را به پیکسل‌هایی از W_d که مربوط به بدن فرد است، اختصاص می‌دهیم.

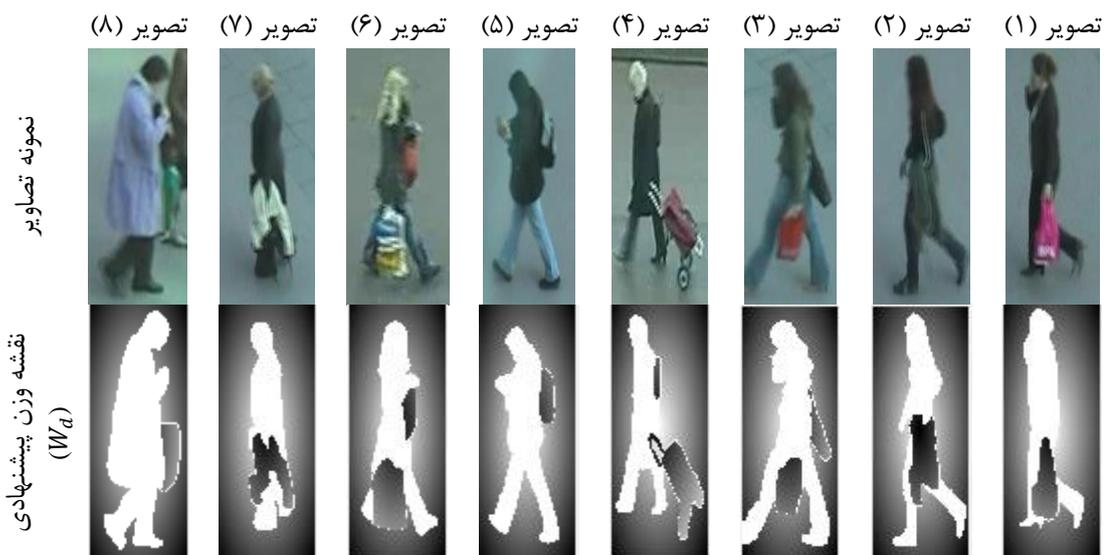
ب) مقادیر $\widehat{W}_{background}^j$ و $\widehat{W}_{objects}^i$ بدست آمده از روابط (۱۹-۳) و (۲۰-۳)، به ترتیب به پیکسل‌هایی از W_d که مربوط به پیکسل i ام شیء در حال حمل و پیکسل j ام پس‌زمینه است، اختصاص می‌یابد:

$$\widehat{W}_{objects}^i = \widehat{D}_i \quad (۱۹-۳)$$

$$\widehat{W}_{background}^j = 1 - \widehat{D}_j \quad (۲۰-۳)$$

با استفاده از رابطه (۱۹-۳)، تاثیر اشیاء در حال حمل بر ویژگی‌های ظاهری استخراجی، متناسب با فاصله پیکسل‌های آن از مرکز تصویر (بدن فرد) کاهش می‌یابد. همچنین، در رابطه (۲۰-۳)، اثر پیکسل‌های پس‌زمینه که از مرکز تصویر فاصله دارند، کمتر از سایر پیکسل‌های پس‌زمینه است.

در شکل (۳-۶)، تعدادی نمونه تصویر از پایگاه داده PRID450s به همراه نقشه‌های وزن مربوطه (W_d) نشان داده شده است. همان‌طور که در این شکل نشان داده شده است، در نقشه وزن پیشنهادی به پیکسل‌های پس‌زمینه که از مرکز تصویر فاصله دارند، مقادیر کمتری اختصاص می‌یابد. همچنین پیکسل‌هایی از اشیاء که نزدیک به مرکز تصویر هستند، مقادیر کمتری در نقشه وزن پیشنهادی دارند.



شکل (۳-۶): تعدادی نمونه تصویر به همراه نقشه‌های وزن مربوطه (W_d)

در این پژوهش، از نقشه‌های وزن پیشنهادی (W_s) و (W_d)، برای وزن‌دهی ویژگی‌های اولیه استخراج شده (F_i) در رابطه (۳-۱) به صورت زیر استفاده می‌شود:

$$\hat{F}_i = W_s^i \times W_d^i \times F_i \quad (۳-۲۱)$$

در این رابطه W_s^i و W_d^i وزن‌های مربوط به پیکسل i ام تصویر هستند که به ترتیب بر اساس اندازه نواحی پس‌زمینه و اشیاء و فاصله پیکسل‌های این نواحی از مرکز تصویر ورودی بدست می‌آیند. همچنین \hat{F}_i نشان‌دهنده بردار ویژگی‌های استخراجی وزن‌دار است.

علاوه بر نقشه وزن‌های پیشنهادی، برای بهبود بازشناسایی و کاهش اثرات مخرب اشیاء در حال حمل بر ویژگی‌های ظاهری افراد، بهتر است ابتدا نواحی مسدود بدن را با توجه به مقادیر پیکسل‌های همسایه بازیابی کنیم. سپس، روشی برای وزن‌دهی ویژگی‌های استخراجی از تصاویر پردازش شده ارائه شود.

۳-۲-۲-۴- وزن‌دهی ویژگی‌ها با توجه به نواحی بازیابی شده

طبق بررسی‌های انجام شده، برای کاهش اثرات چشمگیر اشیاء در حال حمل بر ویژگی‌های ظاهری، هیچکدام از سیستم‌های بازشناسایی موجود، سعی در بازیابی نواحی مسدود بدن ندارند. از این‌رو، در این پژوهش، ابتدا تصاویر به صورت دستی به سه ناحیه بدن فرد، اشیاء و پس‌زمینه قطعه‌بندی می‌شوند. سپس مقادیر پیکسل‌هایی از بدن فرد که توسط اشیاء در حال حمل، مسدود شده‌اند، با توجه به پیکسل‌های همسایه (نواحی غیر مسدود) بازیابی می‌شوند. در این رساله راهکار پیشنهادی برای بازیابی مقادیر پیکسل‌ها، فرآیند یکسان‌سازی^۱ نامیده می‌شود. پس از انجام عمل یکسان‌سازی روی تصویر ورودی، مشابه با راهکار اول تا سوم، نقشه وزنی پیشنهاد می‌شود که در آن بیشترین وزن برای ویژگی‌های استخراجی از پیکسل‌های مرتبط با بدن فرد در نظر گرفته می‌شود. همچنین وزن پیکسل‌های مرتبط با اشیاء حمل شده، نواحی یکسان‌سازی شده و پس‌زمینه با توجه به فاصله پیکسل‌های مربوطه تا مرکز تصویر، تنظیم می‌شود.

^۱ Unification process

همان‌طور که اشاره شد، در مرحله پیش‌پردازش، پیکسل‌های بدن فرد که توسط اشیاء در حال حمل، مسدود شده است، با توجه به رنگ پیکسل‌های همسایه بازیابی می‌شوند. فرض کنید ماتریس تصویر ورودی (I) و تصویر قطعه‌بندی شده آن یعنی (I_s) شامل m سطر و n ستون هستند. همچنین (I_s) شامل مقادیر عددی ۰، ۰٫۵ و ۱ است که به ترتیب نشان‌دهنده ناحیه مربوط به پس‌زمینه، اشیاء و بدن فرد است. در این حالت برای بدست آوردن خروجی فرآیند یکسان‌سازی، یعنی تصویر بازیابی شده (I_u)، مراحل زیر انجام می‌شود:

$$\text{مرحله (۱): } I_u = I; \text{ و } r = 1;$$

مرحله (۲): سطر r ام از ماتریس‌های (I) و (I_s) مربوطه را به ترتیب در بردارهای X و Y قرار دهید:

$$X(j) = I(r, j), \quad j = 1, 2, 3, \dots, n;$$

$$Y(j) = I_s(r, j), \quad j = 1, 2, 3, \dots, n;$$

مرحله (۳): اگر بردار Y شامل برچسب شیء (یعنی ۰٫۵) و برچسب بدن فرد (یعنی ۱) باشد، آنگاه مراحل (۱-۳) تا (۳-۳) را انجام دهید. در غیراین صورت به مرحله (۴) بروید:

مرحله (۳-۱): مقادیری از بردار X را که مربوط به بدن فرد است در بردار U با اندازه

$1 \times T$ قرار دهید، به طوری که T تعداد پیکسل‌های بدن در بردار X

است:

$$Y(j) = 1 \rightarrow U(i) = X(j), \quad j = 1, 2, 3, \dots, n \text{ and } i = 1, 2, 3, \dots, T;$$

مرحله (۳-۲): در بردار X تعداد پیکسل‌های شیء را محاسبه کنید:

$$\text{counter} = 0;$$

$$Y(j) = 0.5 \rightarrow \text{counter} = \text{counter} + 1, \quad j = 1, 2, 3, \dots, n;$$

مرحله (۳-۳): اگر $counter > T$. آنگاه:

$$\hat{U} = \text{upsample}(U),$$

به طوری که اندازه بردار \hat{U} برابر با $1 \times counter$ و $T = counter$

است.

$$\hat{U} = U; \text{ در غیراین صورت}$$

مرحله (۴-۳): در سطر r ام ماتریس I_u ، مقادیر پیکسل‌های شیء را با پیکسل‌ها در

بردار \hat{U} جایگزین کنید:

$$Y(j) = 0.5 \rightarrow I_u(r, j) = \hat{U}(i), \quad j = 1, 2, 3, \dots, n \text{ and } i = 1, 2, 3, \dots, T;$$

مرحله (۴): $r = r + 1$; اگر $r \leq m$ آنگاه به مرحله (۲) بروید؛

مرحله (۵): پایان.

لازم به ذکر است که در فرآیند یکسان‌سازی، اگر شیء در سطر r ام ماتریس (I) همسایه افقی نداشته

باشد، مراحل فوق در راستای ستون‌های تصویر (I) و تصویر قطعه‌بندی شده (I_S) اعمال می‌شود.

تعدادی نمونه تصویر از پایگاه داده VIPeR به همراه تصویر قطعه‌بندی شده دستی (I_S) و نتایج فرآیند

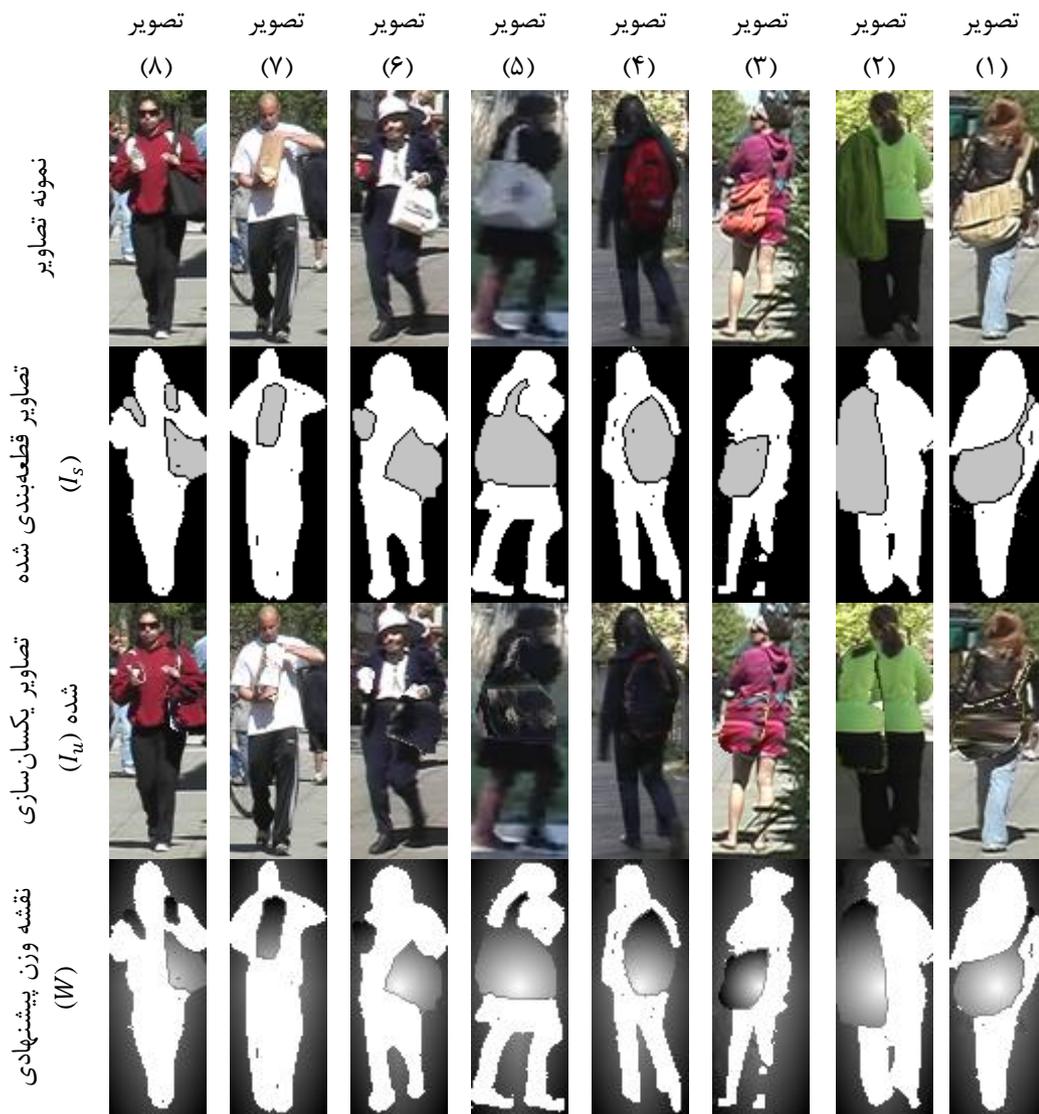
یکسان‌سازی (I_u) به ترتیب در سطرهای (۱) تا (۳) از شکل (۷-۳) نشان داده شده است. در این شکل،

هر ستون به یک نمونه تصویر (یک فرد) اختصاص دارد. همان‌طور که در شکل (۷-۳) نشان داده شده

است، با استفاده از فرآیند یکسان‌سازی، اثر اشیاء بر ویژگی‌های ظاهری کاهش می‌یابد. در واقع، در

فرآیند یکسان‌سازی، پیکسل‌های مسدود بدن با جایگزینی پیکسل‌های همسایه خود به درستی بازیابی

شده‌اند.



شکل (۳-۷): تعدادی نمونه تصویر به همراه تصویر قطعه‌بندی شده دستی (I_s) ، تصاویر حاصل از فرآیند یکسان‌سازی (I_u) و نقشه وزن پیشنهادی (W)

در ادامه برای وزن‌دهی ویژگی‌های استخراجی از تصاویر پیش‌پردازش شده توسط فرآیند یکسان‌سازی، نقشه وزنی با عنوان W به اندازه تصویر ورودی و با توجه به تصویر قطعه‌بندی شده دستی (I_s) ارائه می‌شود. مشابه با نقشه‌های وزن پیشنهادی اول تا سوم، مقادیر پیکسل‌های نقشه وزن در بازه $[۱ و ۰)$ قرار دارند که نشان‌دهنده اهمیت پیکسل مربوطه در تصویر ورودی است. بنابراین وزن پیکسل مربوط به هر ناحیه به صورت زیر تعیین می‌شود:

الف) با توجه به اینکه ظاهر فرد شامل مهمترین ویژگی‌ها برای بازشناسایی است، حداکثر مقدار عددی در بازه [۱ و ۰) یعنی عدد ۱ به پیکسل‌هایی از W که به بدن فرد مربوط می‌شوند، اختصاص می‌یابد.

ب) لازم به ذکر است که مناطق بازیابی شده ممکن است دقیق نباشند؛ زیرا فرآیند یکسان‌سازی، قسمت‌های مسدود بدن را با استفاده از نواحی همسایه به صورت تقریبی شبیه‌سازی می‌کند. از این‌رو، در بازشناسایی فرد، نواحی بازیابی شده نباید اهمیتی مشابه سایر نواحی بدن داشته باشند. از طرفی، می‌توان از اشیاء در حال حمل که بدن را مسدود نمی‌کنند، به‌عنوان یک ویژگی کمکی برای بازشناسایی افراد استفاده کرد. از طرف دیگر، باید اثرات چشمگیر عدم وجود اشیاء حمل شده قبلی را بر ویژگی‌های ظاهری در نظر گرفت. ضمناً پیکسل‌های پس‌زمینه نزدیک به بدن فرد مهمتر از سایر پیکسل‌های پس‌زمینه هستند.

با توجه به نکات فوق، با فرض اینکه بدن فرد در مرکز تصویر قرار دارد، در نقشه وزن پیشنهادی، برای نواحی بازیابی شده، اشیاء در حال حمل و پس‌زمینه، نسبت به بدن فرد وزن کمتری در نظر گرفته می‌شود. از این‌رو، در این راهکار، سعی می‌شود به پیکسل‌های نزدیک به مرکز تصویر نسبت به سایر پیکسل‌ها، وزن بیشتری اختصاص یابد. برای رسیدن به این هدف، مشابه با نقشه وزن پیشنهادی سوم، ابتدا یک ماتریس فاصله (D) با اندازه تصویر ورودی محاسبه می‌شود که شامل مقدار فاصله اقلیدسی پیکسل‌های متناظر در تصویر ورودی از مرکز تصویر است. با فرض اینکه ناحیه S یکی از نواحی بازیابی شده، شیء یا پس‌زمینه تصویر و شامل k پیکسل باشد، ماتریس فاصله (D) توسط روابط (۳-۲۲) تا (۳-۲۴) به صورت زیر محاسبه می‌شود:

$$\hat{y}_k = |y_{center} - y_k| \quad (۳-۲۲)$$

$$\hat{x}_k = |x_{center} - x_k| \quad (۳-۲۳)$$

$$D_k = \sqrt{\hat{y}_k^2 + \hat{x}_k^2} \quad (۳-۲۴)$$

در این روابط، x_{center} و y_{center} به ترتیب مختصات x و y پیکسل مرکزی در تصویر ورودی هستند. همچنین x_k و y_k به ترتیب مختصات x و y پیکسل k تصویر و D_k فاصله اقلیدسی پیکسل مربوطه (پیکسل k) را نشان می‌دهند.

همچنین مقادیر ماتریس D با استفاده از رابطه (۳-۲۵) در بازه عددی [۱ و ۰) نرمال می‌شوند که در آن (M) بیشینه مقدار عددی موجود در ماتریس D است:

$$\hat{D} = \frac{D}{M} \quad (۳-۲۵)$$

با استفاده از \hat{D} و نقشه تقسیم‌بندی شده دستی، مقادیری از $(1 - \hat{D})$ را به پیکسل‌هایی از W که مربوط به ناحیه S هستند، اختصاص می‌دهیم.

در سطر چهارم شکل (۳-۷)، نقشه‌های وزن پیشنهادی (W) مربوط به نمونه تصاویر پایگاه داده VIPeR نشان داده شده است. همان‌طور که در این شکل مشاهده می‌شود، نقشه وزن پیشنهادی مقادیر وزن کمتری را به پیکسل‌های دور از مرکز تصویر اختصاص می‌دهد.

در این پژوهش، از نقشه وزن پیشنهادی (W)، برای وزن‌دهی ویژگی‌های اولیه استخراج شده (F_i) در رابطه (۳-۱) به صورت زیر استفاده می‌شود:

$$\hat{F}_i = W_i \times F_i \quad (۳-۲۶)$$

در این رابطه W_i وزن مربوط به پیکسل i ام تصویر پیش پردازش شده و \hat{F}_i نشان دهنده بردار ویژگی‌های استخراجی وزن‌دار است.

همان‌طور که در بخش (۳-۲-۱) اشاره شد، در توصیفگرهای GoG و HGD به منظور افزایش دقت بردار ویژگی استخراجی، برای ایجاد بردار ویژگی اولیه f_i که در رابطه (۳-۱) نشان داده شد، علاوه بر مدل رنگی RGB، از مدل‌های رنگی Lab، HSV و nRnG استفاده می‌شود. از این‌رو، با استفاده از این توصیفگرها، تاثیر تغییرات روشنایی محیط در ویژگی‌های استخراج شده از تصویر کاهش می‌یابد.

از طرفی طبق [۱۰۲-۱۰۰]، تغییرات روشنایی محیط را می‌توان با استفاده از یک تبدیل خطی با عنوان

مدل قطری^۱ به صورت رابطه (۲۷-۳) مدل کرد:

$$\begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{bmatrix} \times \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix} \quad (27-3)$$

در این رابطه، R_2 ، G_2 و B_2 مقادیر RGB تصویر در شرایط روشنایی شماره ۲ هستند، که با استفاده از ضرب ماتریس مورب در مقادیر RGB تصویر در شرایط روشنایی شماره ۱، یعنی R_1 ، G_1 و B_1 محاسبه می‌شوند. در واقع، مقادیر x ، y و z موجود در ماتریس مورب به سادگی تغییرات روشنایی تصاویر اخذ شده در دو شرایط محیطی متفاوت را مدل می‌کنند.

با توجه به مدل مورب برای تغییرات روشنایی محیط، یکی از مدل‌های رنگی مناسب برای مقاوم‌سازی ویژگی‌های استخراجی در برابر تغییرات روشنایی، مدل رنگی Log-Chromaticity است. مؤلفه‌های این مدل رنگی با استفاده از مؤلفه‌های مدل رنگی RGB به صورت رابطه‌های (۲۸-۳) و (۲۹-۳) بدست می‌آید [۱۰۳]:

$$L_r = \ln \left(\frac{R}{G} \right) \quad (28-3)$$

$$L_b = \ln \left(\frac{B}{G} \right) \quad (29-3)$$

در این روابط، $\ln(\cdot)$ نشان‌دهنده عمل لگاریتم و L_r و L_b مؤلفه‌های مدل رنگی Log-Chromaticity هستند.

از طرفی، طبق روابط (۳۰-۳) و (۳۱-۳)، نگاشت تغییرات روشنایی با استفاده از مدل مورب، تنها منجر به جابجایی^۲ در مختصات پیکسل‌ها می‌شود. از این‌رو، مقادیر پیکسل‌ها در مدل رنگی Log-Chromaticity مستقل از تغییرات روشنایی محیط هستند [۱۰۲ و ۱۰۳]:

¹ Diagonal model

² Shift

$$L_r^2 = \ln\left(\frac{xR}{yG}\right) = L_r^1 + \ln\left(\frac{x}{y}\right) \quad (30-3)$$

$$L_b^2 = \ln\left(\frac{zB}{yG}\right) = L_b^1 + \ln\left(\frac{z}{y}\right) \quad (31-3)$$

در این روابط، L_r^1 ، L_b^1 ، L_r^2 و L_b^2 به ترتیب نشان‌دهنده مؤلفه‌های مدل رنگی Log-Chromaticity در شرایط روشنایی شماره ۱ و ۲ هستند.

با توجه به نکات ذکر شده در مورد مدل رنگی Log-Chromaticity، برای بهبود عملکرد توصیفگرهای GoG و HGD، علاوه بر مدل‌های رنگی مورد استفاده در این توصیفگرها، از مدل رنگی Log-Chromaticity هم استفاده می‌کنیم. از این‌رو، مؤلفه‌های مدل رنگی RGB در رابطه (۳-۱)، با مؤلفه‌های مدل رنگی Log-Chromaticity جایگزین می‌شود. فرض کنید بردار ویژگی استخراج شده با استفاده از مدل رنگی RGB، Lab، HSV و nRnG و Log-Chromaticity، به ترتیب Z_{nRGB} ، Z_{HSV} ، Z_{LAB} ، Z_{RGB} و Z_{LC} نامگذاری شوند. در این صورت با کنار هم قرار دادن این پنج بردار ویژگی، بردار ویژگی نهایی توسط رابطه (۳۲-۳) بدست می‌آید:

$$Z_{Fusion} = [Z_{RGB}^T, Z_{LAB}^T, Z_{HSV}^T, Z_{nRGB}^T, Z_{LC}^T]^T \quad (32-3)$$

۳-۳- جمع‌بندی

در این فصل با توجه به مسائل و مشکلات سیستم‌های بازشناسایی موجود، و با توجه به فرض‌های مورد نظر، جزئیات سیستم بازشناسایی پیشنهادی مورد بررسی قرار گرفت.

در بررسی جزئیات سیستم پیشنهادی اشاره شد که حرکت افراد مقابل یک دوربین یا بین دوربین‌های مختلف، منجر به تغییر در زاویه افراد نسبت به دوربین می‌شود. از طرفی تغییر زاویه افراد نسبت به دوربین ممکن است منجر به تغییر در ویژگی‌های ظاهری افراد شود. از این‌رو، در سیستم بازشناسایی پیشنهادی سعی شد از توصیفگرهایی استفاده شود که ویژگی‌هایی مستقل از تغییرات ظاهری افراد را استخراج کند. در سیستم بازشناسایی پیشنهادی علاوه بر توجه به ویژگی‌های ظاهری، مسئله انسداد

ناشی از حمل اشیاء در افراد و همچنین مسئله پس‌زمینه شلوغ هم مورد توجه قرار گرفت. در این راستا، برای بهبود عملکرد سیستم‌های بازشناسایی انسان، چهار راهکار ارائه شد، به طوری که در هر راهکار تلاش شد نواقص موجود در راهکارهای قبلی برطرف شود. در تمامی این راهکارها تاثیر پیکسل‌های تصویر بر بردارهای ویژگی استخراجی، متناسب با تعلق پیکسل‌ها به پس‌زمینه و پیش‌زمینه و با توجه به میزان اهمیت آنها در مسئله بازشناسایی تعیین شد. در فصل بعدی نتایج حاصل از راهکارهای پیشنهادی، مورد بررسی قرار می‌گیرند.

فصل چهارم

ارزیابی راهکارهای پیشنهادی و بررسی نتایج

۴-۱- مقدمه

در فصل قبل برای بهبود سیستم‌های بازشناسایی انسان، راهکارهایی پیشنهاد دادیم. در تمامی راهکارهای پیشنهادی تلاش شد تاثیر پیکسل‌های تصویر بر بردارهای ویژگی استخراجی، با توجه به میزان اهمیت پیکسل‌ها در مسئله بازشناسایی تعیین شود. در این فصل به ارزیابی عملکرد راهکارهای پیشنهادی و مقایسه نتایج حاصل از آنها با برخی از سیستم‌های بازشناسایی موجود می‌پردازیم. در ابتدای این فصل به معرفی معیار ارزیابی مورد استفاده در این پژوهش می‌پردازیم. در ادامه این فصل مجدداً پایگاه‌های داده مورد استفاده در این رساله به صورت اجمالی معرفی می‌شوند. در نهایت با استفاده از معیار ارزیابی مورد نظر، عملکرد راهکارهای پیشنهادی و سیستم‌های بازشناسایی موجود مورد مقایسه و بررسی قرار می‌گیرند.

۴-۲- معیار ارزیابی مورد استفاده

یکی از معیارهای رایج مورد استفاده به منظور ارزیابی عملکرد سیستم‌های بازشناسایی، معیار ارزیابی Rank-k است. فرض کنید تعدادی نمونه تصویر از افراد مجزا در مجموعه گالری موجود باشد. در این حالت با استفاده از معیار ارزیابی Rank-k، ابتدا تصاویر مجموعه گالری متناسب با میزان شباهتی که به تصویر آزمون دارند، به صورت نزولی رتبه‌بندی می‌شوند، به طوری که تصویر با بیشترین شباهت در رتبه اول و تصویر با کمترین شباهت در رتبه آخر قرار می‌گیرد. در ادامه با توجه به مقدار عددی k، تصاویر مربوط به رتبه اول تا رتبه k، انتخاب می‌شود. در نهایت با توجه به برچسب تصویر آزمون، در k تصویر انتخابی، وجود تصویری با برچسبی برابر با برچسب تصویر آزمون بررسی می‌شود. در حالت ایده‌آل، مطلوب است برچسب تصویر با رتبه اول، برابر با برچسب تصویر آزمون باشد. بنابراین، معیار Rank-k به ازای مقدار عددی $k = 1$ ، بیشترین دقت و به ازای مقدار عددی $k > 1$ ، امکان وجود مقداری خطا را می‌دهد [۱۰۴، ۱۰۵]. در این پژوهش، ارزیابی عملکرد راهکارهای پیشنهادی و مقایسه نتایج حاصل از

آنها با سیستم‌های بازشناسایی توسط معیار ارزیابی Rank-k و با در نظر گرفتن مقادیر عددی ۱، ۵، ۱۰ و ۲۰ برای پارامتر k ، انجام می‌شود.

۳-۴- پایگاه‌های داده مورد استفاده

در این پژوهش، برای ارزیابی و مقایسه عملکرد راهکارهای پیشنهادی از پایگاه‌های داده VIPeR [۱۵]، PRID450s [۲۵]، CUHK01 [۲۲] و CUHK03 [۲۴] استفاده شده است. همان‌طور که در فصل اول اشاره شد، پایگاه‌های داده VIPeR و PRID450s به ترتیب شامل ۱۲۶۴ تصویر از ۶۳۲ فرد مختلف و ۹۰۰ تصویر از ۴۵۰ فرد مختلف هستند که در دو نمای مختلف گرفته شده‌اند. از آنجایی که این پایگاه‌های داده شامل یک تصویر از هر فرد در هر نمای دوربین هستند، عملکرد راهکارهای پیشنهادی روی این دو پایگاه داده، با تطبیق بین تک تصاویر^۱ ارزیابی می‌شود. همچنین در فصل اول اشاره شد که پایگاه داده CUHK01 شامل ۳۸۸۴ تصویر از ۹۷۱ فرد مختلف است. در این پایگاه داده در هر نمای دوربین دو تصویر از هر فرد گرفته شده است. از این رو، در این پژوهش عملکرد راهکارهای پیشنهادی روی پایگاه داده CUHK01 با تطبیق بین تک تصویر ($M=1$) و تطبیق بین چند تصویر^۲ ($M=2$) گزارش می‌شود. همان‌طور که در فصل اول اشاره شد، پایگاه داده CUHK03 شامل ۱۳۱۶۴ تصویر از ۱۳۶۰ فرد مختلف است که به‌طور میانگین ۴/۸ تصویر از هر فرد در هر نمای دوربین گرفته شده است. لازم به ذکر است که تعیین افراد موجود در تصاویر این پایگاه داده به دو صورت دستی (Labeled) و خودکار (Detected) انجام شده است. در این پژوهش، راهکارهای پیشنهادی با استفاده از مجموعه تصاویر Labeled موجود در پایگاه داده CUHK03 و با تطبیق بین چند تصویر ارزیابی می‌شوند.

۴-۴- نتایج حاصل از راهکارهای پیشنهادی

در فصل سوم اشاره شد که در سیستم بازشناسایی پیشنهادی، راهکارهایی برای بهبود عملکرد توصیفگرهای GoG و HGD و در نتیجه بهبود عملکرد سیستم‌های بازشناسایی انسان، ارائه می‌شود.

¹ Single shot matching

² Multi shot matching

همچنین اشاره شد که در تمامی راهکارهای پیشنهادی تلاش می‌شود تا تاثیر پیکسل‌های تصویر بر بردار ویژگی استخراجی توسط توصیفگرهای GoG و HGD، متناسب با تعلق پیکسل‌های تصویر به پس‌زمینه و پیش‌زمینه و با توجه به میزان اهمیت آنها در مسئله بازناسایی تعیین شود.

سیستم بازناسایی پیشنهادی در این پژوهش، پس از استخراج ویژگی‌های تصویر توسط توصیفگرهای بهبود یافته GoG و HGD، مشابه با [۴۶] و [۴۷]، از ویژگی‌های استخراجی، برای آموزش معیار فاصله XQDA [۴۵] استفاده می‌کند. در فصل سوم اشاره شد که خروجی این معیار فاصله، ماتریسی است به نام ماتریس نمره به‌طوری‌که هر عنصر از این ماتریس، فاصله بین تصاویر مربوطه را در مجموعه آزمون و گالری نشان می‌دهد. بنابراین، با مرتب‌سازی هر سطر از ماتریس نمره، تصاویر گالری متناسب با فاصله‌شان تا تصویر آزمون مربوطه رتبه‌بندی می‌شوند. لازم به ذکر است که در توصیفگرهای GoG و HGD، برای آموزش و آزمون معیار فاصله XQDA به ازای تصاویر هر پایگاه داده، به ترتیب از ۱۰ مجموعه تصویر آموزش و آزمون مختلف استفاده می‌شود. از این‌رو در این بخش، نتایج گزارش شده روی هر پایگاه داده، میانگینی از نتایج حاصل از سیستم بازناسایی پیشنهادی روی ۱۰ مجموعه تصویر آزمون مختلف است.

با توجه به نکات مطرح شده، در این بخش نتایج حاصل از هر یک از راهکارهای پیشنهادی با برخی از سیستم‌های بازناسایی موجود، مورد ارزیابی و مقایسه قرار می‌گیرد.

۴-۴-۱- نتایج حاصل از راهکار پیشنهادی اول و دوم

همان‌طور که در فصل سوم توضیح داده شد، در راهکار پیشنهادی اول، با استفاده از توزیع گوسین و تابع چگالی احتمال، روشی برای قطعه‌بندی معنایی تصاویر ارائه شد که در آن پیکسل‌هایی با ارزش یکسان در یک قطعه قرار می‌گیرند. در این رویکرد به ازای هر یک از تصاویر ورودی، یک نقشه وزن با ابعادی برابر با تصویر محاسبه می‌شود، به‌طوری‌که هر یک از پیکسل‌های نقشه وزن، نشان‌دهنده تاثیر پیکسل متناظر از تصویر ورودی بر ویژگی‌های نهایی استخراج شده است. در واقع در نقشه وزن پیشنهادی به پیکسل‌های بدن فرد و پیکسل‌های مربوط به اشیاء و نواحی منحصربفرد در مقایسه با

پیکسل‌های پس‌زمینه، وزن بیشتری تعلق می‌گیرد. در این پژوهش به‌منظور بهبود عملکرد توصیفگرهای GoG و HGD، از نقشه وزن پیشنهادی برای وزن‌دهی ویژگی‌های اولیه استخراج شده توسط این توصیفگرها استفاده می‌شود.

همان‌طور که در فصل سوم اشاره شد، نقشه وزن پیشنهادی اول ممکن است دقیق نباشد، به‌طوری‌که در برخی از نقشه‌های وزن بدست آمده، تعدادی از پیکسل‌های مربوط به بدن فرد مقادیر کمتری نسبت به پس‌زمینه داشته باشند. از این‌رو، در ادامه این پژوهش، سعی شد برای هر یک از تصاویر ورودی، نقشه وزن بهتری ایجاد شود. در این راستا، در راهکار دوم ابتدا تصاویر به‌صورت معنایی به سه ناحیه بدن فرد، اشیاء در حال حمل و پس‌زمینه قطعه‌بندی شدند. سپس در نقشه وزن پیشنهادی، بیشترین وزن (مقدار عددی ۱) به ویژگی‌های استخراجی از پیکسل‌های مرتبط با بدن تعلق گرفت. همچنین در نقشه وزن پیشنهادی با اختصاص مقدار عددی ۰/۹ به‌عنوان وزن ویژگی‌های استخراجی از پیکسل‌های مرتبط با اشیاء در حال حمل، وزن کمتری به پیکسل‌های این ناحیه تعلق گرفت. در نهایت در نقشه وزن پیشنهادی، کمترین وزن (مقدار عددی ۰/۵) به ویژگی‌های استخراجی از پیکسل‌های مرتبط با پس‌زمینه، اختصاص یافت تا پیکسل‌های مربوط به این ناحیه نسبت به بدن و اشیاء در حال حمل، دارای کمترین وزن باشند. در ادامه، برای بهبود عملکرد توصیفگرهای GoG و HGD، از نقشه وزن پیشنهادی دوم برای وزن‌دهی ویژگی‌های اولیه استخراج شده توسط این توصیفگرها استفاده می‌شود.

در جدول (۴-۱)، نتایج حاصل از توصیفگرهای GoG و HGD قبل از اعمال نقشه وزن پیشنهادی (یعنی Classic GoG و Classic HGD) و نتایج بدست آمده پس از اعمال نقشه وزن پیشنهادی اول و دوم روی این توصیفگرها (یعنی $IGoG^1$ و $IHGD^2$) گزارش شده است. در این جدول برای هر یک از پایگاه‌های داده CUHK01، VIPeR، PRID450s و CUHK03، معیار ارزیابی Rank-k به ازای هر یک از مقادیر عددی ۱، ۵، ۱۰ و ۲۰ برای پارامتر k گزارش شده است. نتایج پررنگ‌شده در هر رتبه‌بندی

¹ Improved GoG (IGoG)

² Improved HGD (IHGD)

در این جدول، برای هر کدام از راهکارهای اول و دوم، نشان‌دهنده دقیق‌ترین نتایج بدست آمده روی هر یک از پایگاه‌های داده مورد نظر است.

جدول (۴-۱): نتایج حاصل از بازنمایی انسان توسط توصیفگرهای GoG و HGD قبل و بعد از اعمال نقشه وزن پیشنهادی در راهکارهای اول و دوم

CUHK01 (M=1)	CUHK01 (M=2)	PRID 450s	VIPER		
۵۸٫۷	۶۸٫۸	۷۰٫۳	۵۰٫۶	(راهکار اول) IGoG	رتبه ۱
۶۲٫۵	۷۲٫۱	۷۸٫۷	۵۹٫۴	(راهکار دوم) IGoG	
۵۷٫۹	۶۷٫۳	۶۸٫۴	۴۹٫۷	Classic GoG, (2016)	
۶۱٫۴	۷۲٫۷	۷۲٫۸	۵۱٫۹	(راهکار اول) IHGD	رتبه ۲
۶۳٫۳	۷۴٫۰	۸۰٫۵	۶۰٫۹	(راهکار دوم) IHGD	
۵۹٫۰	۷۰٫۳	۷۰٫۴	۵۰٫۰	Classic HGD, (2019)	
۷۹٫۹	۸۸٫۰	۸۹٫۴	۷۹٫۷	(راهکار اول) IGoG	رتبه ۳
۸۲٫۴	۸۹٫۵	۹۳٫۵	۸۵٫۰	(راهکار دوم) IGoG	
۷۹٫۲	۸۶٫۹	۸۸٫۸	۷۹٫۷	Classic GoG, (2016)	
۸۲٫۳	۹۰٫۱	۹۱٫۵	۸۰٫۵	(راهکار اول) IHGD	رتبه ۴
۸۳٫۵	۹۰٫۰	۹۴٫۷	۸۵٫۹	(راهکار دوم) IHGD	
۷۹٫۷	۸۷٫۹	۹۱٫۲	۷۹٫۵	Classic HGD, (2019)	
۸۷٫۱	۹۲٫۸	۹۳٫۸	۸۸٫۳	(راهکار اول) IGoG	رتبه ۵
۸۸٫۹	۹۴٫۱	۹۷٫۱	۹۱٫۷	(راهکار دوم) IGoG	
۸۶٫۲	۹۱٫۸	۹۴٫۵	۸۸٫۷	Classic GoG, (2016)	
۸۸٫۷	۹۴٫۰	۹۵٫۰	۸۹٫۱	(راهکار اول) IHGD	رتبه ۶
۸۹٫۸	۹۴٫۱	۹۷٫۷	۹۲٫۲	(راهکار دوم) IHGD	
۸۶٫۲	۹۲٫۲	۹۴٫۸	۸۸٫۹	Classic HGD, (2019)	
۹۲٫۶	۹۶٫۶	۹۶٫۸	۹۵٫۲	(راهکار اول) IGoG	رتبه ۷
۹۴٫۰	۹۷٫۳	۹۸٫۸	۹۵٫۹	(راهکار دوم) IGoG	
۹۲٫۱	۹۵٫۹	۹۷٫۸	۹۴٫۵	Classic GoG, (2016)	
۹۳٫۶	۹۶٫۹	۹۷٫۶	۹۵٫۴	(راهکار اول) IHGD	رتبه ۸
۹۴٫۳	۹۷٫۵	۹۹٫۲	۹۶٫۷	(راهکار دوم) IHGD	
۹۲٫۰	۹۵٫۸	۹۷٫۶	۹۴٫۶	Classic HGD, (2019)	

همان طور که در جدول (۴-۱) نشان داده شده است، اعمال نقشه وزن پیشنهادی اول، روی توصیفگرهای GoG و HGD، منجر به بهبود عملکرد این سیستم‌های بازشناسایی در همه رتبه‌ها روی پایگاه‌های داده VIPeR و CUHK01، و در رتبه‌های ۱، ۵ و ۱۰ روی پایگاه داده PRID450s شده است. همچنین طبق نتایجی که در جدول (۴-۱) نشان داده شده است، اعمال نقشه وزن پیشنهادی دوم، روی توصیفگرهای GoG و HGD، منجر به بهبود عملکرد این سیستم‌های بازشناسایی در همه رتبه‌ها روی همه پایگاه‌های داده مورد نظر شده است.

در جدول‌های (۴-۲) تا (۴-۵)، عملکرد سیستم بازشناسایی پیشنهادی اول و دوم، با برخی از سیستم‌های بازشناسایی موجود، به ترتیب با استفاده از پایگاه‌های داده CUHK03، VIPeR، CUHK01 و PRID450s مقایسه شده است. لازم به ذکر است که بیشتر رویکردهای بازشناسایی موجود، دقت خود را روی پایگاه داده CUHK03، با استفاده از معیار Rank-k و فقط به ازای مقدار عددی $k = 1$ گزارش کرده اند. از این رو، در جدول (۴-۲)، مقایسه عملکرد سیستم بازشناسایی پیشنهادی با سیستم‌های بازشناسایی موجود روی این پایگاه داده، با استفاده از این معیار ارزیابی Rank-k و فقط به ازای مقدار عددی $k = 1$ انجام می‌شود.

همان طور که در جدول‌های (۴-۲) تا (۴-۵) نشان داده شده است، سیستم بازشناسایی پیشنهادی اول، نسبت به سیستم‌های بازشناسایی مورد مقایسه در رتبه ۱، روی پایگاه داده CUHK03، در رتبه‌های ۱، ۵ و ۲۰ روی پایگاه‌های داده VIPeR، در همه رتبه‌ها روی پایگاه داده CUHK01، و همچنین در رتبه‌های ۱ و ۲۰ در روی پایگاه داده PRID450s، عملکرد بهتری دارد. در واقع با توجه به نتایج ارائه شده در این جدول‌ها، می‌توان این طور استنباط کرد که در سیستم بازشناسایی پیشنهادی، تعیین میزان تاثیر هر یک از پیکسل‌های تصویر بر ویژگی‌های استخراجی، با توجه به تعلق پیکسل‌ها به پس‌زمینه و پیش‌زمینه تصویر، منجر به بهبود عملکرد سیستم‌های بازشناسایی می‌شود.

همچنین، طبق نتایجی که در جدول‌های (۴-۲) تا (۴-۵) نشان داده شده است، سیستم بازشناسایی پیشنهادی دوم، نسبت به سیستم‌های بازشناسایی مورد مقایسه در همه رتبه‌ها، روی همه پایگاه‌های

داده مورد نظر عملکرد بهتری دارد. در واقع با توجه به نتایج ارائه شده در این جدول‌ها، می‌توان این‌طور استنباط کرد که در سیستم بازشناسایی پیشنهادی، تعیین میزان تاثیر هر یک از پیکسل‌های تصویر بر ویژگی‌های استخراجی، با توجه به تعلق پیکسل‌ها به بدن، اشیاء در حال حمل و پس‌زمینه تصویر، منجر به بهبود عملکرد سیستم‌های بازشناسایی می‌شود.

جدول (۲-۴): مقایسه عملکرد سیستم بازشناسایی پیشنهادی با برخی از سیستم‌های بازشناسایی موجود، با استفاده از پایگاه داده CUHK03(Labeled)

رتبه ۱ (%)	سیستم‌های بازشناسایی
۴۰٫۹	Sun et al., [38], (2017)
۳۶٫۹	Zheng et al, [39], (2018)
۵۳٫۹	Yu et al, [40], (2020)
۶۷٫۳	Classic GoG, (2016)
۶۸٫۹	Classic HGD, (2019)
۷۰٫۱	IGoG (راهکار اول)
۷۱٫۲	IHGD (راهکار اول)
۶۹٫۶	IGoG (راهکار دوم)
۶۸٫۵	IHGD (راهکار دوم)

جدول (۳-۴): مقایسه عملکرد سیستم‌های بازشناسایی پیشنهادی اول و دوم با برخی از سیستم‌های بازشناسایی موجود، با استفاده از پایگاه داده VIPeR

رتبه‌بندی (%)				سیستم‌های بازشناسایی
۱	۵	۱۰	۲۰	
۱۸٫۸	۴۰٫۹	۵۴٫۹	-	Layne et al., [60], (2012)
۲۶٫۷	۵۰٫۷	۶۲٫۴	۷۶٫۴	Zhao et al., [61], (2013)
۳۳٫۰	-	۷۵٫۶	۸۶٫۹	Martinel et al., [63], (2014)
۴۰٫۰	-	۸۰٫۵	۹۱٫۱	Liao et al., [45] (2015)
۴۷٫۵	-	۸۷٫۹	۹۳٫۷	Vishwakarma et al., [48], (2018)
۵۱٫۶	۸۰٫۵	۸۹٫۵	۹۵٫۲	Prates et al. [51], (2019)
۵۱٫۲	۷۹٫۹	۸۹٫۹	-	Prates et al. [52] (2019)
۴۴٫۸	۷۲٫۳	۷۹٫۳	۸۶٫۱	Jia et al., [56], (2020)
۵۰٫۶	۷۹٫۷	۸۸٫۳	۹۵٫۲	IGoG (راهکار اول)
۵۱٫۹	۸۰٫۵	۸۹٫۱	۹۵٫۴	IHGD (راهکار اول)
۵۹٫۴	۸۵٫۰	۹۱٫۷	۹۵٫۹	IGoG (راهکار دوم)
۶۰٫۹	۸۵٫۹	۹۲٫۲	۹۶٫۷	IHGD (راهکار دوم)

جدول (۴-۴): مقایسه عملکرد سیستم‌های بازشناسایی پیشنهادی اول و دوم با برخی از سیستم‌های بازشناسایی موجود، با استفاده از پایگاه داده CUHK01(M=2)

رتبه‌بندی (%)				سیستم‌های بازشناسایی
۱	۵	۱۰	۲۰	
۶۳٫۲	-	۹۰٫۸	۹۴٫۹	Liao et al., [45] (2015)
۵۴٫۵	-	۸۳٫۵	۹۰٫۵	Vishwakarma et al., [48], (2018)
۶۳٫۱	۸۲٫۷	۸۹٫۰	۹۴٫۶	Prates et al., [51], (2019)
۶۸٫۴	۸۶٫۳	۹۳٫۶	۹۶٫۸	Zhao et al., [53], (2020)
۶۸٫۸	۸۸٫۰	۹۲٫۸	۹۶٫۶	IGoG (راهکار اول)
۷۲٫۷	۹۰٫۱	۹۴٫۰	۹۶٫۹	IHGD (راهکار اول)
۷۲٫۱	۸۹٫۵	۹۴٫۱	۹۷٫۳	IGoG (راهکار دوم)
۷۴٫۰	۹۰٫۰	۹۴٫۱	۹۷٫۵	IHGD (راهکار دوم)

جدول (۴-۵): مقایسه عملکرد سیستم‌های بازشناسایی پیشنهادی اول و دوم با برخی از سیستم‌های بازشناسایی موجود، با استفاده از پایگاه داده PRID450s

رتبه‌بندی (%)				سیستم‌های بازشناسایی
۱	۵	۱۰	۲۰	
۶۲٫۶	۸۵٫۶	۹۲٫۰	۹۶٫۶	Liao et al., [45] (2015)
۶۲٫۴	-	۹۳٫۵	۹۶٫۹	Vishwakarma et al., [48], (2018)
۵۸٫۴	۷۷٫۶	۸۴٫۳	۸۹٫۸	Zhou et al., [42], (2018)
۷۰٫۹	۸۹٫۱	۹۳٫۵	۹۶٫۵	Zhou et al., [43], (2019)
۷۱٫۳	۹۱٫۷	۹۶٫۰	۹۸٫۱	Prates et al. [51], (2019)
۶۸٫۱	۹۰٫۷	۹۵٫۰	-	Prates et al. [52] (2019)
۶۸٫۲	۹۰٫۲	۹۴٫۹	۹۸٫۰	Jia et al., [56], (2020)
۷۲٫۱	-	۹۴٫۶	-	Zhao et al., [53], (2020)
۷۰٫۳	۸۹٫۴	۹۳٫۸	۹۶٫۸	IGoG (راهکار اول)
۷۲٫۸	۹۱٫۵	۹۵٫۰	۹۷٫۶	IHGD (راهکار اول)
۷۸٫۷	۹۳٫۵	۹۷٫۱	۹۸٫۸	IGoG (راهکار دوم)
۸۰٫۵	۹۴٫۷	۹۷٫۷	۹۹٫۲	IHGD (راهکار دوم)

لازم به ذکر است که عملیات استفاده شده در مراحل (۱) تا (۳) از راهکار اول که در فصل سوم بخش (۱-۲-۳) ارائه شد، سرباری با مقدار ثابت، به صورت $O(1)$ دارد. از طرفی در راهکار دوم، برای قطعه-بندی تصاویر به سه قطعه بدن فرد، اشیاء در حال حمل و پس‌زمینه، ابتدا شبکه DeepLabv3+ با استفاده از ماسک‌های قطعه‌بندی شده دستی پایگاه‌های داده VIPeR و PRID450s آموزش داده

می‌شود. سپس، شبکه آموزش دیده برای قطعه‌بندی تصاویر سایر پایگاه‌های داده (یعنی CUHK01 و CUHK03) استفاده می‌شود. از این‌رو، با توجه به اینکه پارامترهای شبکه یکبار در طول آموزش تنظیم می‌شوند، می‌توان سربار آموزش DeepLapv3+ را نادیده گرفت. از این‌رو، استفاده از نقشه‌های وزن پیشنهادی در توصیفگرها، سربار ثابت $O(1)$ را به توصیفگرها تحمیل می‌کند.

همانطور که در فصل دوم اشاره شد، سیستم‌های بازشناسایی ارائه شده در [۳۸-۴۰] از شبکه‌های عصبی عمیق برای یادگیری و استخراج ویژگی‌های ظاهری افراد استفاده می‌کنند. در این سیستم‌ها، تنظیم پارامترهای شبکه‌های عصبی عمیق برای یادگیری ویژگی‌های مناسب، به زمان و مکان زیادی نیاز دارد. علاوه بر سیستم‌های بازشناسایی مبتنی بر یادگیری عمیق، برخی از سیستم‌های بازشناسایی مبتنی بر توصیفگرهای دستی شامل مرحله آموزش در استخراج ویژگی هستند. در [۴۲، ۴۳]، به‌منظور بهبود عملکرد سیستم‌های بازشناسایی در برابر تغییرات موقعیت و حالت بدن فرد نسبت به دوربین، روشی مبتنی بر انتقال تناظر گراف ارائه شده است. در این روش الگوهای تناظر مبتنی بر تکه، با استفاده از یک فرآیند تطبیق گراف مبتنی بر تکه، از زوج تصاویر مربوط به یک فرد یاد گرفته می‌شوند. سپس برای هر یک از زوج تصاویر آزمون، تعدادی از شبیه‌ترین جفت‌های آموزشی از نظر حالت بدن، به‌عنوان تصاویر مرجع انتخاب می‌شوند. در ادامه تناظر بین تصاویر مرجع، به تصاویر آزمون نگاشت داده می‌شود تا شباهت بین تصاویر مشخص شود. در مرحله آموزشی ذکر شده، استفاده از چندین زوج تصاویر مربوط به یک فرد، با حالت‌های بدنی مختلف، پیچیدگی محاسباتی این سیستم‌ها را افزایش می‌دهد. در مرجع [۵۳] یک روش یادگیری شباهت ارائه شده است که در آن سعی می‌شود سیستم‌های بازشناسایی در برابر مسئله ناهماهنگی توزیع داده‌ها در نمای دوربین‌ها مقاوم شوند. همان‌طور که در فصل دوم اشاره شد، در این رویکرد، مسئله بازشناسایی افراد به‌عنوان یک مسئله بهینه‌سازی یادگیری انتقالی مشترک با چند دید تکراری در نظر گرفته می‌شود. همچنین، برای توجه بیشتر به ویژگی‌های ظاهری متمایز، در مرجع [۶۰] از برخی ویژگی‌های متمایز سطح میانی مانند اشیاء در حال حمل، عینک آفتابی و آرم لباس استفاده شده است. در این مرجع به ازای هر یک از ویژگی‌های سطح میانی مورد نظر، یک دسته‌بند

توسط ویژگی‌های سطح پایین موجود در نمونه‌های آموزشی، آموزش داده می‌شود. اما، آموزش چندین دسته‌بند برای ویژگی‌های سطح میانی در مقایسه با استفاده مستقیم از ویژگی‌های سطح پایین، زمان‌بر خواهد بود. علاوه‌براین، در [۶۱]، با مقایسه رنگ و بافت تکه‌های تصاویر با تکه‌های تصاویر مرجع، یک نقشه برجستگی مبتنی بر تکه، برای هر تصویر محاسبه می‌شود. این مقایسه‌ها بار محاسباتی زیادی روی رویکرد بازشناسایی ارائه شده در [۶۱] تحمیل می‌کند. لازم به ذکر است که توصیفگرهای GoG و HGD شامل فرآیند یادگیری نیستند. از این‌رو، پیچیدگی محاسباتی GoG و HGD بهبودیافته توسط راهکارهای اول و دوم پیشنهادی، کمتر از سیستم‌های بازشناسایی ذکر شده است.

همچنین به‌منظور تمرکز بیشتر بر ویژگی‌های ظاهری متمایز، در مرجع [۶۳] به ازای هر تصویر، یک نقشه وزن با استفاده از زنجیره مارکوف محاسبه می‌شود. لازم به ذکر است که در نقشه‌های وزن پیشنهادی راهکار اول و دوم، به‌ترتیب، عملیات استفاده شده در مراحل (۱) تا (۳) از راهکار اول که در فصل سوم، بخش (۳-۲-۱) ارائه شد، و همچنین قطعه‌بندی معنایی تصویر توسط شبکه از پیش آموزش‌دیده و در نهایت اختصاص یک عدد ثابت به هر قطعه در راهکار دوم، نیاز به محاسبات کمتری نسبت به استفاده از زنجیره مارکوف دارد. در همین حال، توصیفگر LOMO ارائه شده در [۴۵] تنها براساس دو ویژگی دستی (یعنی HSV و SILTP) عمل می‌کند، بنابراین پیچیدگی محاسباتی کمتری نسبت به توصیفگرهای GoG و HGD دارد. اما، طبق جدول‌های (۴-۳) تا (۴-۵)، دقت GoG و HGD بهبودیافته توسط راهکارهای اول و دوم پیشنهادی، به‌طور قابل توجهی بیشتر از دقت LOMO است.

همچنین رویکرد استخراج ویژگی مورد استفاده در مرجع [۴۸] مشابه با توصیفگرهای GoG و HGD است. تنها تفاوت سیستم بازشناسایی ارائه شده در این مرجع استفاده از ویژگی‌های سطح پایین متفاوت با توصیفگرهای GoG و HGD است. در نتیجه، پیچیدگی محاسباتی رویکرد معرفی شده در [۴۸]، مشابه GoG و HGD است، این در حالی است که، طبق جدول‌های (۴-۳) تا (۴-۵)، عملکرد GoG و HGD بهبودیافته توسط راهکار اول و دوم پیشنهادی بهتر از این رویکرد است. همچنین، سیستم‌های بازشناسایی ارائه شده در [۵۱]، [۵۲] و [۵۶] براساس توصیفگر GoG عمل می‌کنند. از این‌رو، پیچیدگی

محاسباتی این سیستم‌ها در استخراج ویژگی، مشابه GoG و HGD بهبودیافته است. از طرفی این سیستم‌ها نیز مانند GoG و HGD، در برابر برخی مسائل مانند تغییرات در حالت بدن، تغییرات در زاویه فرد نسبت به دوربین و تغییرات روشنایی محیط، مقاوم هستند.

علاوه بر این، طبق جدول (۴-۲)، علیرغم استفاده از ویژگی‌های دستی در GoG و HGD بهبودیافته، عملکرد این سیستم‌های بازناسایی روی پایگاه داده CUHK03، بهتر از سیستم‌های بازناسایی مبتنی بر یادگیری عمیق است. از طرفی، همان‌طور که در جدول‌های (۴-۳) تا (۴-۵) نشان داده شده است، سیستم‌های بازناسایی پیشنهادی در راهکارهای اول و دوم، از روش‌های مورد مقایسه روی پایگاه‌های داده VIPeR، CUHK01 و PRID450s بهتر عمل می‌کنند. در نتیجه، اعمال نقشه‌های وزن پیشنهادی در راهکارهای اول و دوم، روی توصیفگرهای مبتنی بر ظاهر می‌تواند دقت بازناسایی را بدون تحمیل هزینه‌های پردازشی اضافه، بهبود دهد.

۴-۴-۲- نتایج حاصل از راهکار پیشنهادی سوم و چهارم

همان‌طور که در فصل سوم اشاره شد، بهتر است وزن پیکسل‌های هر ناحیه از تصویر، به صورت خودکار و با توجه به اندازه ناحیه و فاصله پیکسل‌های آن تا مرکز تصویر تعیین شود. در راهکار سوم مشابه با راهکارهای اول و دوم، ویژگی‌های استخراجی از پیکسل‌های مرتبط با بدن فرد دارای بیشترین وزن (مقدار عددی ۱) هستند. این در حالی است که در نقشه وزن پیشنهادی راهکار سوم، وزن پیکسل‌های مرتبط با اشیاء در حال حمل و پس‌زمینه با توجه به اندازه هر یک از نواحی و فاصله پیکسل‌های مربوطه تا مرکز تصویر، تنظیم می‌شود. مشابه با راهکارهای اول و دوم، به منظور بهبود عملکرد توصیفگرهای GoG و HGD، از نقشه وزن پیشنهادی برای وزن‌دهی ویژگی‌های اولیه استخراج شده توسط این توصیفگرها استفاده می‌شود.

لازم به ذکر است نقشه وزن پیشنهادی سوم نیاز به قطعه‌بندی تصاویر با دقت بالا دارد. از این رو، در ارزیابی و مقایسه راهکار پیشنهادی سوم، از تصاویر پایگاه‌های داده VIPeR و PRID450s که به صورت دستی قطعه‌بندی شده‌اند، استفاده می‌شود.

در جدول (۴-۶)، نتایج حاصل از توصیفگرهای GoG و HGD قبل از اعمال نقشه وزن پیشنهادی سوم (یعنی Classic GoG و Classic HGD) و نتایج بدست آمده پس از اعمال نقشه وزن پیشنهادی سوم روی این توصیفگرها (یعنی IGoG و IHGD) گزارش شده است. لازم به ذکر است در این جدول، نتایج حاصل از سیستم بازشناسایی پیشنهادی سوم، به ترتیب پس از اعمال نقشه وزن پیشنهادی براساس موارد زیر بدست آمده است:

الف) نقشه وزن پیشنهادی براساس اندازه هر یک از نواحی مربوط به اشیاء در حال حمل و پس‌زمینه (W_s)

ب) نقشه وزن پیشنهادی براساس فاصله پیکسل‌های هر یک از نواحی مربوط به اشیاء در حال حمل و پس‌زمینه تا مرکز تصویر (W_d)

ج) نقشه وزن پیشنهادی براساس اندازه هر یک از نواحی مربوط به اشیاء در حال حمل و پس‌زمینه، و فاصله پیکسل‌های مربوطه تا مرکز تصویر ($W_s \times W_d$)

در جدول (۴-۶) برای هر یک از پایگاه‌های داده VIPeR و PRID450s، معیار ارزیابی Rank-k به ازای هر یک از مقادیر عددی ۱، ۵، ۱۰ و ۲۰ برای پارامتر k گزارش شده است. در این جدول، نتایج پررنگ‌شده در هر رتبه‌بندی، نشان‌دهنده دقیق‌ترین نتایج بدست آمده روی هر یک از پایگاه‌های داده مورد نظر است.

همان‌طور که در جدول (۴-۶) نشان داده شده است، اعمال نقشه وزن پیشنهادی سوم روی توصیفگرهای GoG و HGD، منجر به بهبود عملکرد این سیستم‌های بازشناسایی در همه رتبه‌ها روی هر دو پایگاه داده مورد نظر شده است.

جدول (۴-۶): نتایج حاصل از توصیفگرهای GoG و HGD قبل و بعد از اعمال نقشه وزن پیشنهادی سوم

سیستم‌های بازشناسایی									
Classic GoG	IGoG با استفاده از			Classic HGD	IHGD با استفاده از			پایگاه داده	رتبه
	W_s	W_d	$W_s \times W_d$		W_s	W_d	$W_s \times W_d$		
۴۹٫۷	۵۹٫۹	۵۹٫۲	۶۰٫۸	۵۰٫۰	۶۰٫۴	۶۰٫۱	۶۰٫۴	VIPER	۱ (٪)
۶۸٫۰	۸۲٫۰	۸۰٫۰	۸۲٫۴	۷۰٫۴	۸۳٫۷	۸۲٫۵	۸۲٫۳	PRID450s	
۷۹٫۷	۸۴٫۰	۸۵٫۴	۸۴٫۹	۷۹٫۵	۸۴٫۵	۸۶٫۷	۸۶٫۱	VIPER	۵ (٪)
۸۸٫۷	۹۵٫۴	۹۵٫۳	۹۶٫۰	۹۱٫۲	۹۵٫۶	۹۵٫۵	۹۵٫۶	PRID450s	
۸۸٫۷	۹۰٫۵	۹۲٫۴	۹۱٫۰	۸۸٫۹	۹۱٫۰	۹۲٫۵	۹۱٫۳	VIPER	۱۰ (٪)
۹۴٫۴	۹۷٫۹	۹۷٫۴	۹۸٫۰	۹۴٫۸	۹۸٫۰	۹۷٫۶	۹۷٫۵	PRID450s	
۹۴٫۵	۹۵٫۳	۹۶٫۶	۹۵٫۷	۹۴٫۶	۹۶٫۱	۹۶٫۶	۹۵٫۹	VIPER	۲۰ (٪)
۹۷٫۶	۹۸٫۸	۹۸٫۶	۹۹٫۳	۹۷٫۶	۹۹٫۲	۹۸٫۸	۹۹٫۱	PRID450s	

همان‌طور که در فصل سوم اشاره شد، برای بهبود عملکرد سیستم‌های بازشناسایی و کاهش اثرات مخرب اشیاء در حال حمل بر ویژگی‌های ظاهری افراد، در ادامه این پژوهش، ابتدا توسط فرآیند یکسان‌سازی، نواحی مسدود بدن، با توجه به مقادیر پیکسل‌های همسایه بازمی‌شوند. سپس، نقشه وزنی پیشنهاد می‌شود که در آن بیشترین وزن (مقدار عددی ۱) برای ویژگی‌های استخراجی از پیکسل‌های مرتبط با بدن فرد در نظر گرفته می‌شود. همچنین وزن پیکسل‌های مرتبط با اشیاء حمل شده، نواحی یکسان‌سازی شده و پس‌زمینه با توجه به فاصله پیکسل‌های مربوطه تا مرکز تصویر، تنظیم می‌شود. مشابه با راهکارهای اول تا سوم، به‌منظور بهبود عملکرد توصیفگرهای GoG و HGD، از نقشه وزن پیشنهادی برای وزن‌دهی ویژگی‌های اولیه استخراج شده توسط این توصیفگرها استفاده می‌شود. همچنین در فصل سوم اشاره شد، در توصیفگرهای GoG و HGD به‌منظور افزایش دقت بردار ویژگی استخراجی، علاوه بر مدل رنگی RGB، از مدل‌های رنگی Lab، HSV و nRnG نیز استفاده می‌شود. این‌رو، با استفاده از این توصیفگرها، تاثیر تغییرات روشنایی محیط در ویژگی‌های استخراج شده از تصویر کاهش می‌یابد. در راهکار پیشنهادی چهارم برای مقاوم‌سازی ویژگی‌های استخراجی توصیفگرهای

GoG و HGD در برابر تغییرات روشنایی، علاوه بر مدل‌های رنگی RGB، Lab، HSV و nRnG از مدل رنگی Log-Chromaticity هم استفاده می‌شود.

لازم به ذکر است فرآیند یکسان‌سازی و نقشه وزن پیشنهادی در راهکار چهارم نیاز به قطعه‌بندی تصاویر با دقت بالا دارد. از این‌رو، در ارزیابی و مقایسه راهکار پیشنهادی چهارم، مشابه با راهکار سوم، از تصاویر پایگاه‌های داده VIPeR و PRID450s که به صورت دستی قطعه‌بندی شده‌اند، استفاده می‌شود. در جدول (۷-۴)، نتایج حاصل از توصیفگرهای GoG و HGD قبل از اعمال راهکار پیشنهادی چهارم (یعنی Classic GoG و Classic HGD) و نتایج بدست آمده پس از اعمال راهکار پیشنهادی چهارم روی این توصیفگرها (یعنی IGoG و IHGD) گزارش شده است. لازم به ذکر است در این جدول، نتایج حاصل از سیستم بازنمایی پیشنهادی چهارم، به ترتیب براساس موارد زیر بدست آمده است:

الف) اعمال فرآیند یکسان‌سازی روی تصاویر (UI)؛

ب) اعمال فرآیند یکسان‌سازی روی تصاویر (UI) و نقشه وزن پیشنهادی براساس فاصله پیکسل‌های هر یک از نواحی مربوط به اشیاء در حال حمل و پس‌زمینه تا مرکز تصویر (W)، یعنی ستون مربوط به ($UI + W$) در جدول (۷-۴)؛

ج) اعمال فرآیند یکسان‌سازی روی تصاویر (UI)، نقشه وزن پیشنهادی براساس فاصله پیکسل‌های هر یک از نواحی مربوط به اشیاء در حال حمل و پس‌زمینه تا مرکز تصویر (W) و مدل رنگی Log-Chromaticity (log_ch)، یعنی ستون مربوط به ($UI + W + log_ch$) در جدول (۷-۴).

در جدول (۷-۴) برای هر یک از پایگاه‌های داده VIPeR و PRID450s، معیار ارزیابی Rank-k به ازای هر یک از مقادیر عددی ۱، ۵، ۱۰ و ۲۰ برای پارامتر k گزارش شده است. در این جدول، نتایج پررنگ‌شده در هر رتبه‌بندی، نشان‌دهنده دقیق‌ترین نتایج بدست آمده روی هر یک از پایگاه‌های داده مورد نظر است.

جدول (۷-۴): نتایج حاصل از توصیفگرهای GoG و HGD قبل و بعد از اعمال راهکار پیشنهادی چهارم

سیستم‌های بازشناسایی									
Classic GoG	IGoG با استفاده از			Classic HGD	IHGD با استفاده از			پایگاه داده	رتبه
	UI	UI + W	UI + W + log_ch		UI	UI + W	UI + W + log_ch		
۴۹,۷	۵۳,۴	۵۹,۴	۶۱,۹	۵۰,۰	۵۴,۸	۶۱,۵	۶۱,۸	VIPeR	۱
۶۸,۰	۶۸,۳	۸۰,۴	۸۰,۹	۷۰,۴	۷۰,۷	۸۲,۹	۸۳,۴	PRID450s	(/.)
۷۹,۷	۸۲,۷	۸۶,۰	۸۶,۲	۷۹,۵	۸۳,۲	۸۶,۷	۸۶,۷	VIPeR	۵
۸۸,۷	۸۸,۶	۹۴,۸	۹۴,۹	۹۱,۲	۹۰,۱	۹۵,۶	۹۵,۲	PRID450s	(/.)
۸۸,۷	۹۰,۳	۹۲,۰	۹۲,۷	۸۸,۹	۹۰,۴	۹۲,۵	۹۲,۸	VIPeR	۱۰
۹۴,۴	۹۴,۱	۹۷,۶	۹۷,۶	۹۴,۸	۹۵,۱	۹۷,۷	۹۷,۸	PRID450s	(/.)
۹۴,۵	۹۵,۵	۹۶,۵	۹۷,۴	۹۴,۶	۹۵,۹	۹۶,۶	۹۶,۸	VIPeR	۲۰
۹۷,۶	۹۷,۲	۹۹,۱	۹۹,۳	۹۷,۶	۹۷,۷	۹۸,۸	۹۹,۰	PRID450s	(/.)

همان‌طور که در جدول (۷-۴) نشان داده شده است، استفاده از فرآیند یکسان‌سازی، نقشه وزن پیشنهادی چهارم و مدل رنگی Log-Chromaticity در توصیفگرهای GoG و HGD، منجر به بهبود عملکرد این سیستم‌های بازشناسایی در همه رتبه‌ها روی هر دو پایگاه داده مورد نظر شده است.

لازم به ذکر است، اعمال فرآیند یکسان‌سازی روی تصاویر پایگاه داده VIPeR در مقایسه با پایگاه داده PRID450s منجر به نتایج بهتری می‌شود. علت این امر این است که تصاویر موجود در پایگاه داده PRID450s نسبت به تصاویر پایگاه داده VIPeR دارای انسداد کمتری هستند. همچنین، تصاویر پایگاه داده PRID450s بیشتر شامل اشیاء در حال حملی هستند که بدن افراد را مسدود نکرده‌اند. از طرفی ویژگی‌های ظاهری افراد در این پایگاه داده بسیار شبیه به هم است. در این شرایط می‌توان از اشیاء در حال حمل به‌عنوان یک ویژگی کمکی برای بازشناسایی افراد استفاده کرد. از این‌رو، استفاده از نقشه وزن پیشنهادی چهارم روی تصاویر پایگاه داده PRID450s در مقایسه با پایگاه داده VIPeR، منجر به بهبود بیشتری در عملکرد سیستم بازشناسایی می‌شود.

در شکل (۱-۴) با استفاده از تعدادی نمونه تصویر از پایگاه داده PRID450s، نکات عنوان شده در مورد این پایگاه داده نشان داده شده است. در تصاویر نشان داده شده در این شکل، ویژگی‌های ظاهری افراد به هم شبیه است. از طرفی اشیاء در حال حمل، بدن افراد را مسدود نکرده‌اند.



شکل (۴-۱): تعدادی نمونه تصویر از پایگاه داده PRID450s

علاوه بر این، تغییرات روشنایی در تصاویر پایگاه داده VIPeR نسبت به پایگاه داده PRID450s بیشتر است. از این رو، استفاده از مدل رنگی Log-Chromaticity برای مقاوم سازی توصیفگرهای GoG و HGD در برابر تغییرات روشنایی، بیشتر منجر به بهبود نتایج سیستم بازنمایی پیشنهادی چهارم روی تصاویر پایگاه داده VIPeR شده است.

همچنین، در جدول (۴-۸) عملکرد سیستم‌های بازنمایی پیشنهادی سوم و چهارم، با برخی از سیستم‌های بازنمایی موجود، با استفاده از پایگاه‌های داده VIPeR و PRID450s مقایسه شده است. همان‌طور که در این جدول نشان داده شده است، سیستم بازنمایی پیشنهادی سوم، نسبت به سیستم‌های بازنمایی مورد مقایسه در رتبه‌های ۱، ۵ و ۱۰ روی پایگاه داده VIPeR، و در همه رتبه‌ها روی پایگاه داده PRID450s عملکرد بهتری دارد. در واقع با توجه به نتایج ارائه شده در جدول (۴-۸)، می‌توان این‌طور استنباط کرد که در سیستم بازنمایی پیشنهادی سوم، تعیین میزان تاثیر هر یک از پیکسل‌های تصویر بر ویژگی‌های استخراجی، با توجه به تعلق پیکسل‌ها به بدن فرد، اشیاء در حال حمل و پس‌زمینه تصویر، و همچنین با در نظر گرفتن اندازه هر ناحیه و فاصله پیکسل‌های هر ناحیه تا مرکز تصویر، منجر به بهبود عملکرد سیستم‌های بازنمایی می‌شود.

همچنین، همان‌طور که در جدول (۴-۸) نشان داده شده است، سیستم بازنمایی پیشنهادی چهارم نسبت به سیستم‌های بازنمایی مورد مقایسه در همه رتبه‌ها روی هر دو پایگاه داده مورد نظر عملکرد بهتری دارد. در واقع با توجه به نتایج ارائه شده در جدول (۴-۸)، می‌توان این‌طور استنباط کرد که در سیستم بازنمایی پیشنهادی چهارم، بازیابی نواحی مسدود شده توسط فرآیند یکسان‌سازی، تعیین

میزان تاثیر هر یک از پیکسل‌های تصویر بر ویژگی‌های استخراجی، با توجه به تعلق پیکسل‌ها به بدن فرد، اشیاء در حال حمل، نواحی بازتابی شده و پس‌زمینه تصویر، و همچنین استفاده از مدل رنگی Log-Chromaticity منجر به بهبود عملکرد سیستم‌های بازشناسایی می‌شود.

جدول (۴-۸): مقایسه عملکرد سیستم‌های بازشناسایی پیشنهادی سوم و چهارم با برخی از سیستم‌های بازشناسایی موجود، با استفاده از پایگاه‌های داده VIPeR و PRID450s

رتبه‌بندی روی VIPeR				رتبه‌بندی روی PRID450s				سیستم‌های بازشناسایی
۱ (%)	۵ (%)	۱۰ (%)	۲۰ (%)	۱ (%)	۵ (%)	۱۰ (%)	۲۰ (%)	
۳۲٫۹	۶۰٫۳	۷۳٫۰	-	۳۸٫۲	۶۷٫۱	۷۶٫۰	-	Leng [55], (2018)
۴۷٫۵	-	۸۷٫۹	۹۳٫۷	۶۲٫۴	-	۹۳٫۵	۹۶٫۹	Vishwakarma et al. [48], (2018)
۵۱٫۹	۷۴٫۴	۸۴٫۸	۹۰٫۲	-	-	-	-	Tian et al. [87], (2018)
۴۹٫۴	۷۷٫۶	۸۷٫۲	۹۴٫۰	۵۸٫۴	۷۷٫۶	۸۴٫۳	۸۹٫۸	Zhou et al. [42], (2018)
۵۳٫۵	۸۱٫۳	۸۹٫۱	۹۴٫۷	۷۰٫۹	۸۹٫۱	۹۳٫۵	۹۶٫۵	Zhou et al. [43], (2019)
۴۲٫۱	۶۴٫۰	۷۳٫۴	-	۶۰٫۶	۸۲٫۸	۹۰٫۸	-	Ren et al. [49], (2019)
۴۹٫۰	۷۴٫۱	۸۴٫۴	۹۳٫۱	-	-	-	-	Chu et al. [44], (2019)
۵۱٫۶	۸۰٫۵	۸۹٫۵	۹۵٫۲	۷۱٫۳	۹۱٫۷	۹۶٫۰	۹۸٫۱	Prates et al. [51], (2019)
۵۶٫۶	۸۲٫۶	۸۹٫۹	-	-	-	-	-	Yao et al. [41], (2019)
۵۱٫۲	۷۹٫۹	۸۹٫۹	-	۶۸٫۱	۹۰٫۷	۹۵٫۰	-	Prates et al. [52] (2019)
۵۵٫۳	-	۹۰٫۵	۹۵٫۳	-	-	-	-	Zhu et al. [54], (2020)
۴۳٫۹	۷۲٫۲	۸۰٫۹	۸۷٫۸	۶۸٫۲	۹۰٫۲	۹۴٫۹	۹۸٫۰	Jia et al. [56] (2020)
۵۵٫۹	-	۹۱٫۹	۹۶٫۹	۷۵٫۳	-	۹۷٫۴	۹۹٫۰	Guo et al. [50] (2020)
۵۶٫۸	-	۹۲٫۰	۹۷٫۳	۷۲٫۵	-	۹۶٫۴	۹۸٫۷	Liu et al. [88] (2020)
۵۶٫۳	۸۳٫۰	۹۰٫۰	۹۵٫۸	۷۲٫۱	-	۹۴٫۶	-	Zhao et al. [53] (2020)
۶۰٫۸	۸۵٫۴	۹۲٫۴	۹۶٫۶	۸۲٫۴	۹۶٫۰	۹۸٫۰	۹۹٫۳	IGoG (راهکار سوم)
۶۰٫۴	۸۶٫۷	۹۲٫۵	۹۶٫۶	۸۳٫۷	۹۵٫۶	۹۸٫۰	۹۹٫۲	IHGD (راهکار سوم)
۶۱٫۹	۸۶٫۲	۹۲٫۷	۹۷٫۴	۸۰٫۹	۹۴٫۹	۹۷٫۶	۹۹٫۳	IGoG (راهکار چهارم)
۶۱٫۸	۸۶٫۷	۹۲٫۸	۹۶٫۸	۸۳٫۴	۹۵٫۶	۹۷٫۸	۹۹٫۰	IHGD (راهکار چهارم)

همان‌طور که در تحلیل سیستم‌های بازشناسایی پیشنهادی اول و دوم اشاره شد، در مرجع‌های [۴۲]، [۴۳]، روشی مبتنی بر انتقال تناظر گراف، در راستای بهبود عملکرد سیستم‌های بازشناسایی در برابر تغییرات موقعیت و حالت بدن فرد نسبت به دوربین، ارائه شده است. در این روش برای آموزش الگوهای تناظر مبتنی بر تکه از زوج تصاویر مربوط به یک فرد، از یک فرآیند تطبیق گراف مبتنی بر تکه استفاده می‌شود. سپس به ازای هر یک از زوج تصاویر آزمون، تعدادی از شبیه‌ترین زوج‌های آموزشی از نظر حالت بدن، به‌عنوان تصاویر مرجع انتخاب می‌شوند. در ادامه برای تعیین میزان شباهت بین تصاویر، تناظر بین تصاویر مرجع، به تصاویر آزمون نگاشت داده می‌شود. لازم به ذکر است که در مرحله آموزشی ذکر شده، استفاده از چندین زوج تصاویر مربوط به یک فرد، با حالت‌های بدنی مختلف، پیچیدگی محاسباتی این سیستم‌ها را افزایش می‌دهد. همچنین در مرجع [۵۳] یک روش یادگیری شباهت ارائه شده است که در آن سعی می‌شود سیستم‌های بازشناسایی در برابر مسئله ناهماهنگی توزیع داده‌ها در نمای دوربین‌ها مقاوم شوند. همان‌طور که در فصل دوم اشاره شد، در این مرجع، برای مقاوم‌سازی سیستم‌های بازشناسایی در برابر مسئله ناهماهنگی توزیع داده‌ها بین نماهای دوربین، بازشناسایی افراد به‌عنوان یک مسئله بهینه‌سازی یادگیری انتقالی مشترک با چند دید تکراری در نظر گرفته می‌شود. همچنین مسئله عدم وجود نمونه‌های آموزشی برجسب‌دار در مرجع‌های [۵۴-۵۶] مورد توجه قرار گرفته است. در این راستا همان‌طور که در فصل دوم اشاره شد، در مرجع [۵۴]، بین ماژول‌های یادگیری ویژگی‌های عمیق و یادگیری شباهت، ماژولی شامل دو لایه غیرپارامتری با عنوان‌های لایه یکپارچگی میانگین تقارن بدن و لایه نرمال‌سازی محلی تعبیه شده است. از طرفی، در [۵۵] یک ماتریس فاصله شبیه به فاصله ماهالانویس، با استفاده از چند نمونه آموزشی یاد گرفته می‌شود. در این مرجع، ابتدا ویژگی‌هایی مانند اسامی رنگ‌های منحصربفرد موجود در ظاهر افراد و ویژگی‌های استخراج شده از شبکه عصبی پیچشی سیامز برای نمایش تصاویر در یک نما استفاده می‌شود. همچنین در مرجع [۵۶] سعی می‌شود با استفاده از داده‌های کم آموزشی، زیر فضاهایی برای نگاشت داده‌های اخذ شده توسط هر یک از دوربین‌ها بدست آید. از طرفی، لازم به ذکر است که در سیستم‌های بازشناسایی ارائه شده در

[۴۱، ۵۴-۵۶]، از شبکه‌های عصبی پیچشی برای استخراج ویژگی‌های ظاهری متمایز استفاده می‌شود. نیاز به آموزش در مرحله استخراج ویژگی این سیستم‌های بازنمایی، منجر به تحمیل بار محاسباتی زیادی به خصوص در سیستم‌های مبتنی بر یادگیری عمیق می‌شود. این در حالی است که توصیفگرهای دستی GoG و HGD، برای استخراج ویژگی، نیازی به آموزش ندارند.

برای کاهش اثرات مخرب انسداد و تغییرات حالت بدن افراد، در [۴۴] از ویژگی‌های HSV و SILTP به شکل محلی استفاده شده است. علیرغم اینکه استفاده از دو ویژگی دستی در این مرجع، بار محاسباتی کمتری نسبت به توصیفگرهای GoG و HGD دارد، اما طبق نتایج نشان داده شده در جدول (۴-۸)، دقت سیستم‌های بازنمایی پیشنهادی در این پژوهش به طور قابل ملاحظه‌ای از دقت سیستم بازنمایی ارائه شده در [۴۴] بیشتر است. همچنین، برخی از سیستم‌های بازنمایی مورد مقایسه در جدول (۴-۸)، براساس توصیفگرهای GoG و HGD عمل می‌کنند [۴۸، ۵۰-۵۲]. از این رو، مشابه GoG و HGD، این سیستم‌ها در برابر برخی مسائل مانند تغییرات در حالت بدن، تغییرات در زاویه فرد نسبت به دوربین و تغییرات روشنایی محیط، مقاوم هستند. علاوه بر این، پیچیدگی محاسباتی استخراج ویژگی در این سیستم‌ها مشابه با GoG و HGD است.

همچنین در [۸۷، ۸۸] به موضوع پس‌زمینه‌های شلوغ پرداخته شده است. در مرجع [۸۷]، ابتدا از یک رویکرد قسمت‌بندی بدن برای جداسازی نواحی پس‌زمینه و پیش‌زمینه تصویر استفاده شده است. در ادامه، برای استخراج ویژگی‌های متمایز از پیش‌زمینه تصاویر (بدن افراد)، یک شبکه عصبی عمیق ارائه می‌شود که بر نواحی از تصویر با اطلاعات سودمند، متمرکز می‌شود. علاوه بر این، برای حذف تکه‌های مربوط به پس‌زمینه، در [۸۸] ابتدا تصاویر با استفاده از شبکه تجزیه عمیق به صورت مفهومی قسمت‌بندی می‌شوند. سپس، تکه‌های پس‌زمینه با استفاده از پنجره لغزان و رویکردهای تطبیق رنگ از تصاویر قسمت‌بندی شده، حذف می‌شوند. لازم به ذکر است که استفاده از شبکه‌های عصبی عمیق برای کاهش اثرات مخرب پس‌زمینه‌های شلوغ، نسبت به سیستم‌های بازنمایی پیشنهادی در این پژوهش، بار محاسباتی بیشتری دارند. از طرفی، علیرغم تلاش‌های انجام شده برای بهبود عملکرد سیستم‌های

بازشناسایی، هیچکدام از سیستم‌های بازشناسایی مورد بررسی، به‌طور همزمان مسئله انسداد ناشی از حمل اشیاء و پس‌زمینه‌های شلوغ را در نظر نگرفته‌اند.

لازم به ذکر است که در سیستم‌های بازشناسایی پیشنهادی در این پژوهش، از توصیفگرهای قوی و متمایز GoG و HGD به شکل وزن‌دهی شده برای استخراج ویژگی‌های تصاویر استفاده می‌شود. در واقع، در سیستم‌های بازشناسایی پیشنهادی، توصیفگرهای GoG و HGD در برابر مسائل مربوط به انسداد، پس‌زمینه شلوغ و تغییرات روشنائی، مقاوم‌سازی می‌شوند. از این‌رو، دقت رویکردهای بازشناسایی مبتنی بر GoG و HGD [۴۸، ۵۰-۵۳]، می‌تواند با استفاده از توصیفگرهای GoG و HGD بهبودیافته در این پژوهش، افزایش یابد.

همچنین، طبق نتایج ارائه شده در جدول (۴-۸)، علیرغم استفاده از ویژگی‌های ظاهری دستی در سیستم‌های بازشناسایی پیشنهادی، عملکرد این سیستم‌ها نسبت به سیستم‌های بازشناسایی مورد مقایسه از جمله سیستم‌های بازشناسایی مبتنی بر یادگیری عمیق [۴۱، ۴۹، ۵۴، ۵۵، ۸۷]، بهتر است. از این‌رو، می‌توان این‌طور استنباط کرد که استفاده از نقشه‌های وزن پیشنهادی در سیستم‌های بازشناسایی مبتنی بر یادگیری عمیق، می‌تواند منجر به بهبود عملکرد این سیستم‌ها می‌شود.

۴-۵- بررسی سربار محاسباتی راهکارهای پیشنهادی بر سیستم‌های بازشناسایی موجود

فرض کنید که تصویر ورودی دارای A پیکسل و ناحیه S یکی از مناطق بازبایی شده، اشیاء در حال حمل و پس‌زمینه تصویر باشد به‌طوری‌که این ناحیه شامل a پیکسل است. از طرفی، عملیات استفاده شده در مراحل (۱) تا (۳) از راهکار پیشنهادی اول، مقداردهی عددی به نقشه وزن در راهکار دوم، روابط (۳-۱۳) تا (۳-۲۰) مورد استفاده در راهکار پیشنهادی سوم و روابط راهکار چهارم یعنی (۳-۲۲) تا (۳-۲۵)، سرباری به‌صورت $O(1)$ مقداری ثابت دارند. از این‌رو، در سیستم‌های بازشناسایی موجود، استفاده از نقشه وزن پیشنهادی اول و دوم منجر به سربار $O(A)$ و استفاده از نقشه وزن پیشنهادی سوم و چهارم منجر به سربار $O(a)$ می‌شود.

علاوه بر این، سربار عملیات استفاده شده در مراحل (۱) تا (۵) فرآیند یکسان سازی به صورت $O(1)$ مقداری ثابت باشد، سربار استفاده از فرآیند یکسان سازی پیشنهادی در سیستم‌های بازشناسایی، متناسب با تعداد پیکسل‌های تصویر به صورت $O(A)$ است. همچنین سربار استفاده از مدل رنگی Log-Chromaticity در سیستم‌های بازشناسایی، به طور مستقیم به الگوریتم مورد استفاده در هر سیستم بستگی دارد. در نتیجه، استفاده از راهکارهای وزن دهی ویژگی‌های استخراجی و همچنین اعمال فرآیند یکسان سازی پیشنهادی، در سیستم‌های بازشناسایی موجود، هزینه کمی را بر این سیستم‌ها تحمیل می‌کند.

۴-۶- جمع بندی

در این فصل، عملکرد راهکارهای پیشنهادی برای بهبود سیستم‌های بازشناسایی انسان، مورد ارزیابی و مقایسه قرار گرفت. در این راستا از معیار ارزیابی Rank-k و پایگاه‌های داده مختلف VIPeR، PRID450s، CUHK01 و CUHK03 استفاده شده است. در راهکارهای پیشنهادی تاثیر انسداد ناشی از حمل اشیاء و پس‌زمینه‌های شلوغ بر ویژگی‌های ظاهری افراد مورد توجه قرار گرفت. از این رو، در تمامی راهکارهای پیشنهادی، تاثیر پیکسل‌های تصویر بر بردارهای ویژگی استخراجی، با توجه به میزان اهمیت پیکسل‌ها و تعلق آنها به بدن فرد، اشیاء در حال حمل و پس‌زمینه، تعیین شد. نتایج حاصل از سیستم‌های بازشناسایی پیشنهادی، برتری دقت این سیستم‌ها را نسبت به سیستم‌های بازشناسایی مورد مقایسه نشان می‌دهد.

فصل پنجم

جمع‌بندی و راهکارهای پیشنهادی برای ادامه کار

۵-۱- جمع‌بندی

سیستم‌های بازشناسایی بخش مهمی از سیستم‌های نظارتی هستند، به طوری که عملکرد سیستم‌های بازشناسایی تأثیری مستقیم بر عملکرد سیستم‌های نظارتی دارد. بیشتر سیستم‌های بازشناسایی فرض می‌کنند ویژگی‌های ظاهری افراد حین فرآیند بازشناسایی تغییری نمی‌کند. از طرفی، حرکت افراد مقابل دوربین‌ها، تغییر در روشنایی محیط، انسداد تمام یا بخشی از بدن افراد به دلیل حمل اشیاء و پس‌زمینه‌های شلوغ، ممکن است منجر به ایجاد برخی تغییرات در ویژگی‌های ظاهری افراد شود. از این‌رو، فرض عدم تغییر ویژگی‌های ظاهری در سیستم‌های بازشناسایی منجر به کاهش دقت این سیستم‌ها می‌شود. همچنین مسئله بازشناسایی افراد به دو موضوع مهم قطعه‌بندی تصاویر و اثر انسداد ناشی از حمل اشیاء بسیار وابسته است. در این پژوهش سعی شد با در نظر گرفتن مسئله انسداد ناشی از حمل اشیاء و پس‌زمینه‌های شلوغ، عملکرد سیستم بازشناسایی انسان بهبود یابد. در این راستا، راهکارهایی برای بهبود عملکرد توصیفگرهای GoG و HGD و در نتیجه بهبود عملکرد سیستم‌های بازشناسایی انسان، ارائه شد. در راهکار اول و دوم، اهمیت موضوع قطعه‌بندی تصاویر مشخص شد. همچنین در راهکار سوم و چهارم اهمیت موضوع انسداد ناشی از حمل اشیاء و چگونگی برخورد با آن مورد بررسی قرار گرفت. در تمامی راهکارهای پیشنهادی تأثیر پیکسل‌های تصویر بر بردار ویژگی استخراجی توسط توصیفگرهای GoG و HGD، متناسب با تعلق پیکسل‌ها به پس‌زمینه و پیش‌زمینه و با توجه به میزان اهمیت آنها در مسئله بازشناسایی به صورت عددی در بازه [۱ و ۰] تعیین می‌شود.

در راهکار اول، روشی برای قطعه‌بندی معنایی تصویر به نواحی پس‌زمینه و پیش‌زمینه بر مبنای توزیع گوسین و تابع چگالی احتمال، ارائه شد. سپس از تصویر قطعه‌بندی شده به‌عنوان نقشه وزن برای وزن‌دهی ویژگی‌های استخراجی از پیکسل‌های تصویر استفاده شد.

به منظور کاهش خطاهای قطعه‌بندی معنایی ارائه شده در راهکار اول، برای قطعه‌بندی معنایی تصاویر به سه ناحیه بدن، اشیاء در حال حمل، و پس‌زمینه، در راهکار دوم از روش قطعه‌بندی مبتنی بر شبکه

عمیق DeepLabv3+ استفاده می‌شود. سپس با توجه به ارزش پیکسل‌های هر ناحیه در مسئله بازشناسایی، به هر پیکسل عددی ثابت در بازه [۱ و ۰) اختصاص می‌یابد.

در راهکار سوم، بیشترین وزن به ویژگی‌های استخراجی از پیکسل‌های مرتبط با بدن فرد تعلق گرفت. همچنین، به‌منظور اجتناب از وزن‌دهی دستی پیکسل‌های تصویر، وزن پیکسل‌های مرتبط با اشیاء در حال حمل و پس‌زمینه با توجه به اندازه هر یک از نواحی و فاصله پیکسل‌های مربوطه تا مرکز تصویر، تعیین می‌شود.

در راهکار چهارم، به‌منظور بازیابی پیکسل‌های ناحیه مسدود بدن با توجه به پیکسل‌های همسایه، فرآیندی به نام فرآیند یکسان‌سازی ارائه شد. پس از بازیابی نواحی مسدود، بیشترین وزن برای ویژگی‌های استخراجی از پیکسل‌های مرتبط با بدن فرد در نظر گرفته می‌شود. همچنین وزن پیکسل‌های مرتبط با اشیاء در حال حمل، نواحی یکسان‌سازی شده و پس‌زمینه به‌صورت خودکار و با توجه به فاصله پیکسل‌های مربوطه تا مرکز تصویر، تعیین می‌شود.

عملکرد سیستم‌های بازشناسایی پیشنهادی روی پایگاه‌های داده VIPeR، PRID450s، CUHK01 و CUHK03، مورد ارزیابی و مقایسه قرار گرفت. نتایج حاصل از ارزیابی و مقایسه سیستم‌های بازشناسایی پیشنهادی، برتری دقت این سیستم‌ها را نسبت به سیستم‌های بازشناسایی مورد مقایسه نشان می‌دهد.

۲-۵- راهکارهای پیشنهادی برای ادامه کار

در این پژوهش، به ازای هر یک از تصاویر ورودی، از یک نقشه وزن برای تعیین میزان اهمیت پیکسل‌های تصویر در مسئله بازشناسایی، استفاده شده است. هر یک از روش‌های پیشنهادی دارای مزایا و معایبی هستند که به تفصیل در فصل سوم به آنها اشاره شد. در ادامه برای بهبود عملکرد این روش‌ها، راهکارهای زیر پیشنهاد می‌شود:

● ارائه شبکه‌های عمیق با دقت بالا برای قطعه‌بندی معنایی تصاویر

عمکرد سیستم‌های بازشناسایی پیشنهادی وابسته به دقت روش‌های قطعه‌بندی مورد استفاده است. از طرفی، وضوح تصاویر مورد استفاده در سیستم‌های بازشناسایی پایین است. از این‌رو، روش‌های قطعه‌بندی موجود نمی‌توانند این تصاویر را با دقت مناسب قطعه‌بندی کنند. این مسئله بیشتر در سیستم‌های بازشناسایی پیشنهادی سوم و چهارم تاثیرگذار است تا جایی که تصاویر مورد استفاده در این دو روش به‌صورت دستی قطعه‌بندی شدند. بنابراین، برای کاربردی‌تر شدن سیستم‌های بازشناسایی پیشنهادی و ارزیابی آنها روی سایر پایگاه‌های داده، لازم است روشی برای قطعه‌بندی معنایی تصاویر با وضوح کم ارائه شود به‌طوری‌که این روش توانایی قطعه‌بندی معنایی تصاویر را با دقت بالا داشته باشد.

● استفاده از نقشه‌های وزن پیشنهادی در سایر توصیفگرها و شبکه‌های عمیق

در سیستم‌های بازشناسایی پیشنهادی از توصیفگرهای GoG و HGD استفاده شد که در آنها از ویژگی‌های سطح پایین مانند رنگ و بافت تصاویر استفاده می‌شود. این در حالی است که راهکارهای پیشنهادی در مقایسه با بسیاری از روش‌های بازشناسایی که از ویژگی‌های شبکه‌های عمیق در آنها استفاده شده است، بهتر عمل می‌کند. از این‌رو، می‌توان از نقشه‌های وزن پیشنهادی در شبکه‌های عمیق هم استفاده کرد به‌طوری‌که ویژگی‌های خروجی هر لایه از این شبکه‌ها به‌صورت وزن‌دار و با توجه به اهمیت و تعلق ویژگی‌ها به هر یک از نواحی تصویر ایجاد شوند.

● استفاده از شبکه‌های عمیق برای بدست آوردن نقشه وزن

اساس این پژوهش توجه به اهمیت موضوع استفاده از نقشه وزن پیکسل‌ها در بازشناسایی افراد است. وجود خطا در نقشه وزن ارائه شده در راهکار اول و تعیین دستی وزن پیکسل‌ها در راهکار دوم از جمله معایب این دو راهکار هستند. از طرفی در راهکارهای سوم و چهارم، نقشه‌های وزن به‌صورت خودکار تعیین می‌شدند ولی دقت این دو راهکار وابسته به دقت قطعه‌بندی تصاویر است.

با توجه به نکات مطرح شده، می‌توان از قابلیت‌های شبکه‌های عمیق در ایجاد نقشه وزن تصویر استفاده کرد. به‌عنوان مثال با توجه به ویژگی‌های شبکه مولد متقابل شرطی^۱ [۱۰۶] و با در نظر گرفتن توابع هدف مناسب، می‌توان حین تولید یک تصویر، نقشه وزن مربوط به آن را هم تولید کرده و از نقشه وزن تولید شده برای تصاویر مربوطه استفاده کرد.

- استفاده از شبکه‌های عمیق برای بازیابی نواحی مسدود

در این پژوهش برای بازیابی نواحی مسدود بدن، از فرآیندی پیشنهادی به نام فرآیند یکسان‌سازی استفاده شد. لازم به ذکر است که دقت فرآیند یکسان‌سازی، وابسته به دقت قطعه‌بندی معنایی تصاویر است. در راستای بهبود فرآیند یکسان‌سازی پیشنهادی، با توجه به قابلیت‌های موجود در شبکه‌های عمیق، می‌توان برای بازیابی نواحی مسدود بدن و ایجاد تصاویر بدون انسداد، از شبکه‌های عمیق مانند شبکه مولد متقابل شرطی استفاده کرد؛ به‌طوری‌که یکسان‌سازی نواحی مسدود در این شبکه، حین تولید تصاویر جدید و با توجه به تابع‌های هدف مناسب انجام شود.

¹ Conditional Generative Adversarial Network (CGAN)

- [1] Alairaji, R.A.M., Aljazaery, I.A. and ALRikabi, H.T., (2022). "Abnormal Behavior Detection of Students in the Examination Hall from Surveillance Videos", In *Advanced Computational Paradigms and Hybrid Intelligent Computing (ICACCP)*, pp. 113-125. Springer, Singapore.
- [2] Nagrath, P., Dwivedi, S., Negi, R. and Singh, N., (2022). "Real-Time Anomaly Detection Surveillance System", In *Proceedings of Data Analytics and Management (ICDAM)*, pp. 665-678. Springer, Singapore.
- [3] Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L. and Hoi, S.C., (2021). "Deep Learning for Person Re-Identification: A Survey and Outlook", *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, doi: 10.1109/TPAMI.2021.3054775.
- [4] Sharma, A., Anand, S. and Kaul, S.K., (2022). "Intelligent Camera Selection Decisions for Target Tracking in a Camera Network", In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3388-3397. Waikoloa, Hawaii.
- [5] Tu, P.H., Doretto, G., Krahnstoeber, N.O., Perera, A.A., Wheeler, F.W., Liu, X., Rittscher, J., Sebastian, T.B., Yu, T. and Harding, K.G., (2007). "An Intelligent Video Framework for Homeland Protection", In *International Society for Optics and Photonics, Unattended Ground, Sea, and Air Sensor Technologies and Applications IX*, 6562, p. 65620C. Orlando, Florida, USA.
- [6] Specker, A., Moritz, L., Cormier, M. and Beyerer, J., (2022). "Fast and Lightweight Online Person Search for Large-Scale Surveillance Systems", In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 570-580. Waikoloa, Hawaii.
- [7] بیدی م، منصوری زاده م، (۱۴۰۰)، "بازشناسی فعالیت های انسانی در خانه های هوشمند با اینترنت هوشمند اشیاء"، پنجمین دوره کنفرانس بین المللی اینترنت اشیاء و کاربردها، اصفهان،
<https://civilica.com/doc/1238639>
- [8] شبانی س، امینیان مدرس ا.ف، (۱۳۹۶)، "تاثیر کاهش ابعاد ویژگی ها در دقت بازشناسی فعالیت انسان"، چهارمین کنفرانس بین المللی مطالعات نوین در علوم کامپیوتر و فناوری اطلاعات، مشهد،
<https://civilica.com/doc/779219>
- [9] حقیقی م، عطارزاده ا، (۱۳۹۹)، "تحلیل روش های بازشناسی اشیاء در تصاویر با استفاده از تکنیک های داده کاوی و یادگیری ماشین"، چهارمین کنفرانس بین المللی تحقیقات بین رشته-ای در مهندسی برق، کامپیوتر، مکانیک و مکاترونیک در ایران و جهان اسلام، تهران،
<https://civilica.com/doc/1184137>

[۱۰] سبطی ع، (۱۳۹۶)، رساله دکتری: "بازشناسی انسان با استفاده از ویژگی‌های مدل سه‌بعدی در سیستم‌های نظارت ویدیویی"، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شاهرود.

[11] Filipi Gonçalves dos Santos, C., Oliveira, D.D.S., A. Passos, L., Gonçalves Pires, R., Felipe Silva Santos, D., Pascotti Valem, L., P. Moreira, T., Cleison S. Santana, M., Roder, M., Paulo Papa, J. and Colombo, D., (2022). "Gait Recognition Based on Deep Learning: A Survey", *ACM Computing Surveys (CSUR)*, 55(2), pp.1-34.

[۱۲] صالحی ف، معین م.ش.، (۱۴۰۰)، "استفاده از الگوریتم‌های فرامکاشف‌های برای بهبود نسبت فشرده‌سازی و نرخ بازشناسی تصاویر در یک سامانه بازشناسی چهره"، نشریه ماشین بینایی و پردازش تصویر، سال هشتم، شماره ۱، صفحه ۴۵-۵۸.

[13] Rajasekar, V., Predić, B., Saracevic, M., Elhoseny, M., Karabasevic, D., Stanujkic, D. and Jayapaul, P., (2022). "Enhanced Multimodal Biometric Recognition Approach for Smart Cities Based on an Optimized Fuzzy Genetic Algorithm", *Scientific Reports (SR)*, 12(1), pp.1-11.

[14] Wu, J., Huang, Y., Wu, Q., Gao, Z., Zhao, J. and Huang, L., (2022). "Dual-Stream Guided-Learning via a Priori Optimization for Person Re-Identification", *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17(4), pp.1-22.

[15] Gray D, Tao H. (2008). "Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features", In *European Conference on Computer Vision (ECCV)*, pp. 262-275. Springer, Berlin, Heidelberg.

[۱۶] سبطی ع، حسن پور ح، (۱۳۹۸)، "بازشناسی انسان در سیستم‌های نظارت ویدیویی با بهره‌گیری از اطلاعات زاویه‌ای"، مجله رایانش نرم و فناوری اطلاعات، سال هشتم شماره چهارم، صفحه ۳۰-۴۳.

[17] Loy, C.C., Liu, C. and Gong, S., (2013). "Person Re-Identification by Manifold Ranking", In *20th IEEE International Conference on Image Processing (ICIP)*, pp. 3567-3571. Melbourne, Australia.

[۱۸] قنبری سرخی ع، حسن پور ح، (۱۳۹۵)، "ردیابی و شناسایی مجدد افراد در شبکه‌ای از دوربین‌ها با دید مجزا بر مبنای سیستم فازی در محیط‌های بسته"، مجله مهندسی برق و الکترونیک ایران، سال سیزدهم، صفحه ۷۵.

[19] Baltieri, D., Vezzani, R. and Cucchiara, R., (2011). "3DPeS: 3D People Dataset for Surveillance and Forensics", In *Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding (J-HGBU)*, pp. 59-64. Scottsdale, Arizona, USA.

- [20] Zhao, T., Xiao, J.S., Wang, W., Sui, H.G., Ma, J.J. and Xu, C., (2022). “Robust Detection and Tracking Algorithm for Pedestrian Targets with Posture Change”, In *Proceedings of the 7th China High Resolution Earth Observation Conference (CHREOC 2020)*, pp. 173-183. Springer, Singapore.
- [21] Martinel, N. and Micheloni, C., (2012). “Re-Identify People in Wide Area Camera Network”, In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 31-36. Rhode Island, USA.
- [22] Li, W., Zhao, R. and Wang, X., (2012). “Human Reidentification with Transferred Metric Learning”, In *Asian Conference on Computer Vision (ACCV)*, pp. 31-44. Springer, Berlin, Heidelberg.
- [23] Li, W. and Wang, X., (2013). “Locally Aligned Feature Transforms Across Views”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3594-3601. Portland, Oregon, USA.
- [24] Li, W., Zhao, R., Xiao, T. and Wang, X., (2014). “Deepreid: Deep Filter Pairing Neural Network for Person Re-Identification”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 152-159. Columbus, Ohio, USA.
- [25] Roth, P.M., Hirzer, M., Köstinger, M., Beleznai, C., Bischof H., (2014). “Mahalanobis Distance Learning for Person Re-Identification”, In *Advances in Computer Vision and Pattern Recognition (ACVPR)*, pp. 247-267. Springer, London.
- [26] Das, A., Chakraborty, A. and Roy-Chowdhury, A.K., (2014). “Consistent Re-Identification in a Camera Network”, In *European Conference on Computer Vision (ECCV)*, pp. 330-345. Zurich, Switzerland.
- [27] Figueira, D., Taiana, M., Nambiar, A., Nascimento, J. and Bernardino, A., (2014). “The HDA+ Data Set for Research on Fully Automated Re-Identification Systems”, In *European Conference on Computer Vision (ECCV)*, pp. 241-255. Zurich, Switzerland.
- [28] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J. and Tian, Q., (2015). “Scalable Person Re-Identification: A Benchmark”, In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1116-1124. Santiago, Chile.
- [29] Zheng, L., Zhang, H., Sun, S., Chandraker, M., Yang, Y. and Tian, Q., (2017). “Person Re-Identification in the Wild”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1367-1376. Honolulu, HI, USA.
- [30] Xiao, T., Li, S., Wang, B., Lin, L. and Wang, X., (2016). “End-to-End Deep Learning for Person Search”, *arXiv preprint arXiv:1604.01850*, 1(2), p. 4.
- [31] Zheng, L., Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S. and Tian, Q., (2016). “Mars: A Video Benchmark for Large-Scale Person Re-Identification”, In *European Conference on Computer Vision (ECCV)*, pp. 868-884. Amsterdam, The Netherlands.

- [32] Ristani, E., Solera, F., Zou, R., Cucchiara, R. and Tomasi, C., (2016). “Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking”, In *European Conference on Computer Vision (ECCV)*, pp. 17-35. Amsterdam, The Netherlands.
- [33] Zheng, Z., Zheng, L. and Yang, Y., (2017). “Unlabeled Samples Generated by Gan Improve the Person Re-Identification Baseline in Vitro”, In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 3754-3762. Venice, Italy.
- [34] Gou, M., Karanam, S., Liu, W., Camps, O. and Radke, R.J., (2017). “DukeMTMC4ReID: A Large-Scale Multi-Camera Person Re-Identification Dataset”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 10-19. Honolulu, HI, USA.
- [35] Gou, M., Wu, Z., Rates-Borras, A., Camps, O. and Radke, R.J., (2018). “A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 41(3), pp.523-536.
- [36] Wei, L., Zhang, S., Gao, W. and Tian, Q., (2018). “Person Transfer Gan to Bridge Domain Gap for Person Re-Identification”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 79-88. Salt Lake City, Utah, USA.
- [37] Zheng, M., Karanam, S. and Radke, R.J., (2018). “RPIfield: A New Dataset for Temporally Evaluating Person Re-Identification”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1893-1895. Salt Lake City, Utah, USA.
- [38] Sun, Y., Zheng, L., Deng, W. and Wang, S., (2017). “Svdnet for Pedestrian Retrieval”, In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 3800-3808. Venice, Italy.
- [39] Zheng, Z., Zheng, L. and Yang, Y., (2019). “Pedestrian Alignment Network for Large-scale Person Re-Identification”, In *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 29(10), pp. 3037-3045.
- [40] Yu, Y., Liang, C., Ruan, W. and Jiang, L., (2020). “Crowdsourcing-Based Ranking Aggregation for Person Re-Identification”, In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1933-1937. Barcelona, Spain.
- [41] Yao, H., Zhang, S., Hong, R., Zhang, Y., Xu, C. and Tian, Q., (2019). “Deep Representation Learning with Part Loss for Person Re-Identification”, *IEEE Transactions on Image Processing (TIP)*, 28(6), pp. 2860-2871.
- [42] Zhou, Q., Fan, H., Zheng, S., Su, H., Li, X., Wu, S. and Ling, H., (2018). “Graph Correspondence Transfer for Person Re-Identification”, In *Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pp. 7599-7606. New Orleans, Louisiana, USA.

- [43] Zhou, Q., Fan, H., Yang, H., Su, H., Zheng, S., Wu, S. and Ling, H., (2021). “Robust and Efficient Graph Correspondence Transfer for Person Re-Identification”, *IEEE Transactions on Image Processing (TIP)*, 30, pp. 1623-1638.
- [44] Chu, H., Qi, M., Liu, H. and Jiang, J., (2019). “Local Region Partition for Person Re-Identification”, *Multimedia Tools and Applications (MTAP)*, 78(19), pp. 27067–27083.
- [45] Liao, S., Hu, Y., Zhu, X. and Li, S.Z., (2015). “Person Re-Identification by Local Maximal Occurrence Representation and Metric Learning”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2197-2206. Boston, Massachusetts, USA.
- [46] Matsukawa, T., Okabe, T., Suzuki, E. and Sato, Y., (2016). “Hierarchical Gaussian Descriptor for Person Re-Identification”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1363-1372. Las Vegas, USA.
- [47] Matsukawa, T., Okabe, T., Suzuki, E. and Sato, Y., (2019). “Hierarchical Gaussian Descriptors with Application to Person Re-Identification”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 42(9), pp. 2179-2194.
- [48] Vishwakarma, D.K. and Upadhyay, S., (2018). “A Deep Structure of Person Re-Identification Using Multi-Level Gaussian Models”, *IEEE Transactions on Multi-Scale Computing Systems (TMSCS)*, 4(4), pp. 513-521.
- [49] Ren, Q.Q., Tian, W.D. and Zhao, Z.Q., (2019). “Person Re-Identification Based on Feature Fusion”, In *International Conference on Intelligent Computing (ICIC)*, pp. 65-73. Springer, Nanchang, China.
- [50] Guo, R., Li, C.G., Li, Y., Lin, J. and Guo, J., (2020). “Density-Adaptive Kernel Based Efficient Re-Ranking Approaches for Person Re-Identification” *Neurocomputing*, 411, pp. 91-111.
- [51] Prates, R. and Schwartz, W.R., (2019). “Kernel Cross-View Collaborative Representation Based Classification for Person Re-Identification”, *Journal of Visual Communication and Image Representation (JVCIR)*, 58, pp. 304-315.
- [52] Prates, R. and Schwartz, W.R., (2019). “Matching People Across Surveillance Cameras”, In *Anais Estendidos da XXXII Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 84-90. Rio de Janeiro, Brazil.
- [53] Zhao, C., Wang, X., Zuo, W., Shen, F., Shao, L. and Miao, D., (2020). “Similarity Learning with Joint Transfer Constraints for Person Re-Identification”, *Pattern Recognition*, 97, p. 107014.
- [54] Zhu, J., Zeng, H., Huang, J., Zhu, X., Lei, Z., Cai, C. and Zheng, L., (2020). “Body Symmetry and Part-Locality-Guided Direct Nonparametric Deep Feature Enhancement for Person Re-Identification”, *IEEE Internet of Things Journal (IoT-J)*, 7(3), pp. 2053-65.

- [55] Leng, Q., (2018). “Co-Metric Learning for Person Re-Identification”, *Advances in Multimedia*, 2018, pp. 1-9.
- [56] Jia, J., Ruan, Q., Jin, Y., An, G. and Ge, S., (2020). “View-Specific Subspace Learning and Re-Ranking for Semi-Supervised Person Re-Identification”, *Pattern Recognition*, 108, p. 107568.
- [57] Garcia, J., Martinel, N., Micheloni, C. and Gardel, A., (2015). “Person Re-Identification Ranking Optimisation by Discriminant Context Information Analysis”, In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1305-1313. Santiago, Chile.
- [58] Li, H., Zhu, J. and Tao, D., (2018). “Asymmetric Projection and Dictionary Learning with Listwise and Identity Consistency Constraints for Person Re-Identification” *IEEE Access*, 6, pp.37977-37990.
- [59] Liu, C., Gong, S., Loy, C.C. and Lin, X., (2012). “Person Re-Identification: What Features Are Important?”, In *European Conference on Computer Vision (ECCV)*, pp. 391-401. Heidelberg, Berlin.
- [60] Layne, R., Hospedales, T.M. and Gong, S., (2012). “Towards Person Identification and Re-Identification with Attributes”, In *European Conference on Computer Vision (ECCV)*, pp. 402-412. Heidelberg, Berlin.
- [61] Zhao, R., Ouyang, W. and Wang, X., (2013). “Unsupervised Saliency Learning for Person Re-Identification”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3586-3593. Portland, Oregon, USA.
- [62] Zhao, R., Ouyang, W. and Wang, X., (2017). “Person Re-Identification by Saliency Learning”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 39(2), pp.356-370.
- [63] Martinel, N., Micheloni, C. and Foresti, G.L., (2014). “Saliency Weighted Features for Person Re-Identification”, In *European Conference on Computer Vision (ECCV)*, pp. 191-208. Zurich, Switzerland.
- [64] Martinel, N., Micheloni, C. and Foresti, G.L., (2015). “Kernelized Saliency-Based Person Re-Identification Through Multiple Metric Learning”, *IEEE Transactions on Image Processing (TIP)*, 24(12), pp.5645-5658.
- [65] An, L., Kafai, M., Yang, S. and Bhanu, B., (2013). “Reference-Based Person Re-Identification”, In *10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 244-249. Krakow, Poland.
- [66] An, L., Kafai, M., Yang, S. and Bhanu, B., (2016). “Person Re-Identification with Reference Descriptor”, *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 26(4), pp.776-787.
- [67] Liu, H., Feng, J., Qi, M., Jiang, J. and Yan, S., (2017). “End-to-End Comparative Attention Networks for Person Re-Identification” *IEEE Transactions on Image Processing (TIP)*, 26(7), pp.3492-3506.

- [68] Li, T., Sun, L., Han, C. and Guo, J., (2018). “Salient Region-Based Least-Squares Log-Density Gradient Clustering for Image-To-Video Person Re-Identification”, *IEEE Access*, 6, pp.8638-8648.
- [69] Cao, Z., Simon, T., Wei, S.E. and Sheikh, Y., (2017). “Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7291-7299. Honolulu, Hawaii, USA.
- [70] Liao, S., Zhao, G., Kellokumpu, V., Pietikäinen, M. and Li, S.Z., (2010). “Modeling Pixel Process with Scale Invariant Local Patterns for Background Subtraction in Complex Scenes”, In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1301-1306. San Francisco, California, USA.
- [71] Xu, Y., Fang, X., Wu, J., Li, X. and Zhang, D., (2015). “Discriminative Transfer Subspace Learning via Low-Rank and Sparse Representation”, *IEEE Transactions on Image Processing (TIP)*, 25(2), pp. 850-863.
- [72] Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D. and Li, SZ., (2014). “Salient Color Names for Person Re-Identification”, In *European Conference on Computer Vision (ECCV)*, pp. 536-551. Springer, Zurich, Switzerland.
- [73] Bromley, J., Bentz, J.W., Bottou, L., Guyon, I., LeCun, Y., Moore, C., Sackinger, E. and Shah, R., (1993). “Signature Verification Using a “Siamese” Time Delay Neural Network”, *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, 7(4), pp. 669-688.
- [74] Susan, S. and Kumar, A., (2016). “Auto-Segmentation Using Mean-Shift and Entropy Analysis”, In *3rd IEEE International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 292-296. New Delhi, India.
- [75] Ashizawa, M., Sasaki, H., Sakai, T. and Sugiyama, M., (2017). “Least-Squares Log-Density Gradient Clustering for Riemannian Manifolds”, In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 537-546. Fort Lauderdale, Florida, USA.
- [76] Cheng, M.M., Mitra, N.J., Huang, X., Torr, P.H. and Hu, S.M., (2015). “Global Contrast Based Salient Region Detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 37(3), pp.569-582
- [77] Huang, H., Li, D., Zhang, Z., Chen, X. and Huang, K., (2018). “Adversarially Occluded Samples for Person Re-Identification”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5098-5107. Salt Lake City, Utah, USA.
- [78] Zhuo, J., Chen, Z., Lai, J. and Wang, G., (2018). “Occluded Person Re-Identification”, In *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1-6. San Diego, California, USA.
- [79] Miao, J., Wu, Y., Liu, P., Ding, Y. and Yang, Y., (2019). “Pose-Guided Feature Alignment for Occluded Person Re-Identification”, In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 542-551. Seoul, South Korea.

- [80] Wang, G.A., Yang, S., Liu, H., Wang, Z., Yang, Y., Wang, S., Yu, G., Zhou, E. and Sun, J., (2020). “High-Order Information Matters: Learning Relation and Topology for Occluded Person Re-Identification”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6449-6458. Seattle, Washington, USA.
- [81] Zhang, X., Yan, Y., Xue, J.H., Hua, Y. and Wang H., (2020). “Semantic-Aware Occlusion-Robust Network for Occluded Person Re-Identification”, *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 31(7), pp. 2764-2778.
- [82] Yan, C., Pang, G., Jiao, J., Bai, X., Feng, X. and Shen, C., (2021). “Occluded Person Re-Identification with Single-Scale Global Representations”, In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11875-11884. Montreal, Canada.
- [83] Wang, H., Chen, X. and Liu, C., (2021). “Pose-Guided Part Matching Network via Shrinking and Reweighting for Occluded Person Re-Identification”, *Image and Vision Computing (IVC)*, 111, p.104186.
- [84] Xie, G., Wen, X., Yuan, L., Wang, J., Guo, C., Jia, Y. and Li, M., (2021). “Pose-Guided Feature Region-Based Fusion Network for Occluded Person Re-Identification”, *Multimedia Systems*, pp.1-13.
- [85] Huang, W., Liu, S., Luo, R., Si, T. and Zhang, Z., (2021). “Dynamically Occluded Samples via Adversarial Learning for Person Re-Identification in Sensor Networks”, *Ad Hoc Networks*, 110, p.102316.
- [86] Liu, Q., Chen, D., Chu, Q., Yuan, L., Liu, B., Zhang, L. and Yu, N., (2022). “Online Multi-Object Tracking with Unsupervised Re-Identification Learning and Occlusion Estimation”, *Neurocomputing*, <https://doi.org/10.1016/j.neucom.2022.01.008>.
- [87] Tian, M., Yi, S., Li, H., Li, S., Zhang, X., Shi, J., Yan, J. and Wang, X., (2018). “Eliminating Background-Bias for Robust Person Re-Identification”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5794-5803. Salt Lake City, Utah, USA.
- [88] Liu, Y., Zhang, Y., Coleman, S., Bhanu, B. and Liu, S., (2020). “A New Patch Selection Method Based on Parsing and Saliency Detection for Person Re-Identification”, *Neurocomputing*, 374, pp. 86-99.
- [89] Zhang, Z., Lan, C., Zeng, W., Jin, X. and Chen Z., (2020). “Relation-Aware Global Attention for Person Re-Identification”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3186-3195. Seattle, Washington, USA.
- [90] Gong, Y., Wang, R., Yang, J., Xue, L. and Hu, M., (2021). “Person Re-Identification with Global-Local Background_Bias Net”, *Journal of Visual Communication and Image Representation (JVCIR)*, 74, p.102961.

- [91] He, K., Zhang, X., Ren, S. and Sun, J., (2016). “Deep Residual Learning for Image Recognition” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778. Las Vegas, Nevada, USA.
- [92] Gour, N. and Khanna, P., (2020). “Automated Glaucoma Detection Using GIST and Pyramid Histogram of Oriented Gradients (PHOG) Descriptors”, *Pattern Recognition Letters (PRL)*, 137, pp.3-11.
- [93] You, K. and Park, H.J., (2021). “Re-Visiting Riemannian Geometry of Symmetric Positive Definite Matrices for the Analysis of Functional Connectivity”, *Neuroimage*, 225, p.117464.
- [94] Son, N.T., Absil, P.A., Gao, B. and Stykel, T., (2021). “Computing Symplectic Eigenpairs of Symmetric Positive-Definite Matrices via Trace Minimization and Riemannian Optimization”, *SIAM Journal on Matrix Analysis and Applications (SIMAX)*, 42(4), pp.1732-1757.
- [95] Lovrić, M., Min-Oo, M. and Ruh, E.A., (2000). “Multivariate Normal Distributions Parametrized as a Riemannian Symmetric Space”, *Journal of Multivariate Analysis (JMVA)*, 74(1) pp. 36-48.
- [96] Cai, X., Chan, R., Nikolova, M. and Zeng, T., (2017). “A Three-Stage Approach for Segmenting Degraded Color Images: Smoothing, Lifting and Thresholding (SLaT)”, *Journal of Scientific Computing (JSC)*, 72(3), pp. 1313-1332.
- [97] Luong, Q.T., (1993). “Color in Computer Vision”, In *Handbook of Pattern Recognition and Computer Vision*, World Scientific Publishing Co., Inc., PP. 311-368.
- [98] Le, K.N., (2011). “A Mathematical Approach to Edge Detection in Hyperbolic-Distributed and Gaussian-Distributed Pixel-Intensity Images Using Hyperbolic and Gaussian Masks”, *Digital Signal Processing (DSP)*, 21(1), pp. 162-181,
- [99] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., (2018). “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”, In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 801-818. Munich, Germany.
- [100] Finlayson, G., Drew, M. and Funt, B., (1993). “Diagonal Transforms Suffice for Color Constancy”, *4th International Conference on Computer Vision (ICCV)*, pp. 164-171. Berlin, Germany.
- [101] Finlayson, G., Drew, M. and Funt, B., (1994). “Spectral Sharpening: Sensor Transformations for Improved Color Constancy”, *Journal of the Optical Society of America A (JOSA A)*, 11(5), pp. 1553-1563.
- [102] Kviatkovsky, I., Adam, A. and Rivlin, E., (2012). “Color Invariants for Person Re-Identification”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 35(7), pp. 1622-1634.
- [103] Berwick, D. and Lee, S., (1998). “A Chromaticity Space for Specularity, Illumination Color and Illumination Pose-Invariant 3-D Object Recognition”, *6th International Conference on Computer Vision (ICCV)*, pp. 165-170. Bombay, India.

- [104] Swearingen, T. and Ross, A., (2021). "Lookalike Disambiguation: Improving Face Identification Performance at Top Ranks", In *2020 25th IEEE International Conference on Pattern Recognition (ICPR)*, pp. 10508-10515. Cancun, Mexico.
- [105] Mortezaie, Z. and Hassanpour, H., (2019). "A Survey on Age-Invariant Face Recognition Methods", *Jordanian Journal of Computers and Information Technology (JJCIT)*, 5(2), pp. 87-96.
- [106] Chen, X., Sun, Y., Shu, X. and Li, Q., (2021). "Attention-Aware Conditional Generative Adversarial Networks for Facial Age Synthesis", *Neurocomputing*, 451, pp.167-180.

Abstract

Human behavior analysis and visual anomaly detection are important applications of video surveillance systems, in fields such as security systems, intelligent houses, and elderly care. People re-identification is one of the main steps in a surveillance system that directly affects system performance. This step is used to determine labels of people in images considering their visual appearance. It is a challenging task as the appearance may change across the camera's network. Body occlusion, crowded background, and variations in scene illumination and pose are some challenging issues in people re-identification. Previous re-identification approaches faced limitations while considering appearance changes in their tracking task. In this research, considering body occlusion caused by carried objects and crowded background, four weighing mechanisms are proposed to improve the performance of re-identification approaches. In each proposed weighing mechanism, a weight map is proposed to tune the effect of background and foreground regions on the extracted appearance characteristics proportional to the region's importance in re-identification.

In the first approach, using the Gaussian distribution and the probability density function, an unsupervised semantic segmentation approach is proposed, where it assigns pixels with the same importance to the same segment. The segmented image is then used as a weight map to extract appearance features from the image's pixels in the weighted form. Due to some errors in the first approach, in the second weighing mechanism, using DeepLabv3+, the image is semantically segmented into three regions as: person body, possible carried objects, and background. Then, a constant number is assigned to the pixels associated with each region depending to its importance in people re-identification. To avoid weighing regions manually, in the third mechanism, a weight map is computed representing the importance of each region in the re-identification, depending to its size and location. Besides, considering the importance of retrieving the occluded parts of the body in re-identification, in the fourth approach, a pre-processing step namely unification process is proposed to retrieve the occluded parts of the body using their neighboring pixels. Then, a weight map is introduced in order to tune the effect of each pixel of the processed image on the extracted features considering its distance from the center of the image.

Our proposed weighing mechanisms along with robust descriptors, such as the Gaussian of Gaussian (GoG) and Hierarchical Gaussian Descriptors (HGD), can enhance existing methods in dealing with the challenging issues in re-identification. Experimental results on VIPeR, PRID450s, CUHK01, and CUHK03 datasets demonstrate effectiveness of the proposed techniques in improving performance of existing re-identification methods.

Keywords: Surveillance System; Re-identification System; Weight Map; Occlusion; Carried Object; Crowded Background.



Shahrood University of Technology

Faculty of Computer Engineering

PhD Dissertation in Artificial Intelligence Engineering

**People Re-identification in Video Surveillance Systems
Considering Occlusion Due to Carrying Objects**

By: Zahra Mortezaie

Supervisor:

Prof. H. Hassanpour

Adviser:

Prof. A. Beghdadi

Feb 2022