
بسم الله الرحمن الرحيم

دانشکده مهندسی برق و رباتیک

پایان‌نامه کارشناسی ارشد مهندسی مخابرات سیستم

شناسایی فعالیت‌های انسانی به کمک روش‌های معنایی و غیر معنایی

نگارنده: زهراسادات شواکندی

استاد راهنما

دکتر علیرضا احمدی فرد

بهمن ۱۳۹۶

شماره: ۱۵۴/ت.ت
تاریخ: ۹۶/۱۱/۱۱

باسمه‌تعالی

دانشگاه صنعتی شاهرود

مدیریت تحصیلات تکمیلی

فرم شماره (۳) صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با نام و یاد خداوند متعال، ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای **زهرا شواکندی** با شماره دانشجویی **۹۴۱۰۵۴۴** رشته **مهندسی برق- مخابرات** گرایش **سیستم** تحت عنوان : **شناسایی فعالیت های انسانی به کمک روش های معنایی و غیر معنایی** که درتاریخ **۱۳۹۶/۱۱/۱۱** با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می گردد:

قبول (با درجه: خیلی خوب) ☑ مردود ☐

نوع تحقیق: نظری ■ عملی ☐

عضو هیأت داوران	نام ونام خانوادگی	مرتبهٔ علمی	امضاء
۱ـ استادراهنمای اول	عوض احمدی فر	دانشیار	
۲- استادراهنمای دوم	ـــــ	ـــــ	ـــــ
۳- استاد مشاور	ـــــ	ـــــ	ـــــ
۴- نماینده تحصیلات تکمیلی	ولی ا... بشیری	استادیار	
۵- استاد ممتحن اول	سیدمهدی سجادی امیری	دانشیار	
۶ـ استاد ممتحن دوم	حسین خسروی	استادیار	

نام و نام خانوادگی رئیس دانشکده:

تاریخ و امضاء و مهر دانشکده:

تبصره: در صورتی که کسی مردود شود حداکثر یکبار دیگر (در مدت مجاز تحصیل) می تواند از پایان نامه خود دفاع نماید (دفاع مجدد نباید زودتر از ۴ ماه برگزار شود).

تقدیم به

مادر عزیز، دلسوز و مهربانم...

تشکر و قدردانی

سپاس خدای را که سخنوران، در ستودن او بمانند و شمارندگان، شمردن نعمتهای او ندانند و کوشندگان، حق او را گزاردن نتوانند.

بدون شک جایگاه و منزلت معلم، بالاتر از آن است که در مقام قدردانی از زحمات آنها، با زبان قاصر و دست ناتوان چیزی بنگاریم؛ اما بر حسب وظیفه بر خود واجب میدانم از استاد شایسته و فرهیخته جناب آقای **دکتر علیرضا احمدی‌فرد** که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این رساله را بر عهده گرفتند کمال تشکر و قدردانی را بجا آورم. همچنین در انتها از جناب آقای مهندس **خشایار نوریان** و تمامی دوستانم که به نوعی در موفقیت و پیشرفت من نقش داشته‌اند کمال تشکر را دارم. باشد که این خردترین بخشی از زحمات آنان را سپاس گوید.

تعهد نامه

اینجانب **زهراسادات شواکندی** دانشجوی دوره کارشناسی ارشد رشته مهندسی برق/رباتیک دانشکده مهندسی برق و رباتیک دانشگاه صنعتی شاهرود نویسنده پایان نامه **فعالیت‌های انسانی به کمک روش‌های معنایی و غیر معنایی** تحت راهنمائی دکتر علیرضا احمدی فرد متعهد می شوم .

- 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 مورد استفاده 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 می باشد 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 بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- 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 .

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . 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 نمی باشد.

چکیده

در سال‌های اخیر با افزایش نگرانی‌های امنیتی و ضرورت مطالعه رفتار انسان‌ها نیاز به سامانه-ی نظارت هوشمندی که بتواند محتوای فعالیت‌های رخ داده شده را شناسایی کند، افزایش یافته است. در سیستم‌های نظارت غیرخودکار، بازبینی ویدئوها توسط یک ناظر انسانی انجام می‌شود که بازبینی حجم بالای محتوی، موجب خستگی و نظارت غیر موثر می‌گردد. با توجه به دلایل مطرح شده، زمینه تحقیقاتی برای شناسایی خودکار فعالیت در رشته‌های ویدئویی ایجاد شده‌است.

شناسایی فعالیت انسانی به دلایلی چون حرکت دوربین، تنوع پس‌زمینه، ظاهر مختلف افراد و از همه مهمتر تنوع در انجام یک فعالیت توسط افراد مختلف بسیار چالش‌برانگیز است. استفاده از ویژگی‌های سطح پایین برای شناسایی فعالیت‌های انسان از دیرباز مورد توجه قرار گرفته است ولی رسیدگی به چالش‌های اشاره شده نیازمند درک معنایی و دانش انسانی درباره فعالیت‌های رخ داده است. برخلاف ویژگی‌های سطح پایین، ویژگی‌های معنایی خصوصیات بارز فعالیت‌ها را توصیف می-کنند. در نتیجه، شناسایی معنایی فعالیت انسان، بخصوص زمانی که فعالیت‌های یکسان از نظر بصری متفاوت باشند، قابل اطمینان‌تر و کارآمدتر هستند.

در این پایان‌نامه یک روش برای شناسایی فعالیت انسان پیشنهاد شده است که در آن از ویژگی معنایی در کنار ویژگی‌های غیرمعنایی استفاده نموده‌ایم. در روش پیشنهادی از مسیر حرکت متراکم و ویژگی‌های HOG، HOF و MBH برای استخراج ویژگی غیرمعنایی استفاده شده و سپس مسیرهای حرکت شاخص تعیین می‌شوند. با بکارگیری الگوریتم خوشه‌بندی دو لایه بر روی مسیرهای حرکت شاخص به تولید ویژگی معنایی می‌پردازیم. این ویژگی‌ها با استفاده از کیف کلمات کد می-شوند، و سپس بوسیله طبقه‌بند نزدیک‌ترین همسایه شناسایی فعالیت‌ها انجام می‌شود. با اجرای این روش بر روی پایگاه داده UCFsports صحت ۹۲/۹٪ بدست آمده‌است. نتایج نشان می‌دهد ویژگی‌های معنایی در کنار ویژگی‌های غیرمعنایی شناسایی رفتار را بطور چشم گیری بهبود می‌بخشد.

کلمات کلیدی: شناسایی فعالیت انسانی، شناسایی معنایی، ویژگی معنایی، مسیرحرکت متراکم، کیف‌کلمات، پایگاه داده UCF sports

فهرست مطالب

فهرست شکل‌ها

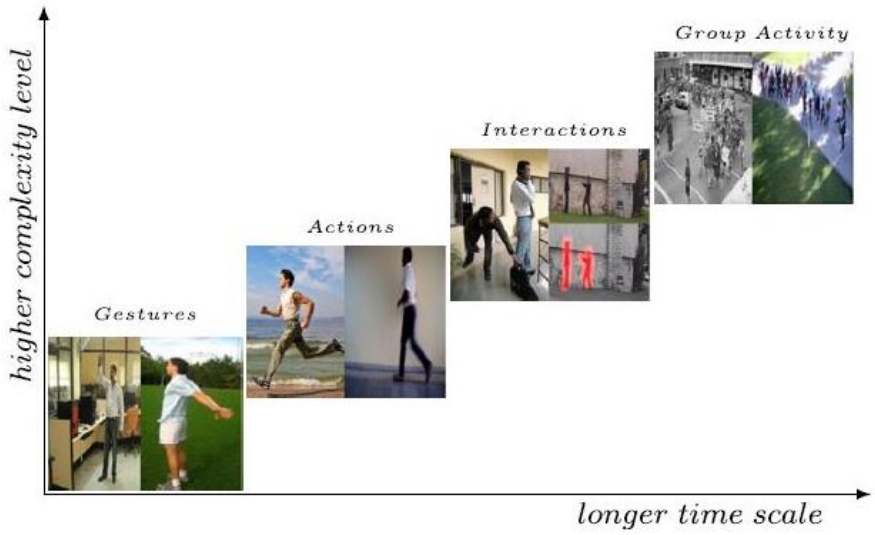
فهرست جدول‌ها

فصل ۱: مقدمه

۱-۱ تعریف مسئله

یکی از مطالعات در حال توسعه در حوزه بینایی ماشین شناسایی فعالیت است. امروزه هدف اصلی شناسایی فعالیت، تجزیه و تحلیل خودکار فعالیت‌های انجام شده یا در حال اجرا می‌باشد. انواع مختلفی از فعالیت‌های انسان وجود دارد، از جمله این فعالیت‌ها می‌توان به فعالیت‌های ورزشی، فعالیت‌های اجتماعی، فعالیت‌های روزمره و... اشاره کرد. براساس پیچیدگی فعالیت‌ها می‌توان آنها را در چهار سطح دسته بندی کرد: حرکات ابتدایی، تعاملات بین افراد، تعاملات بین افراد و اشیا، فعالیت‌های گروهی [۱].

حرکات انسانی از ساده ترین حرکت یک عضو تا حرکات پیچیده توام گروهی از اندام‌ها و بدن گسترده هستند. برای مثال، حرکت پا در یک بازی فوتبال یک حرکت ساده است در حالی‌که پریدن برای ضربه سر در فوتبال شامل مجموعه ایی از حرکات پا، بازو، سر و کل بدن است شکل (۱-۱).



شکل ۱-۱ انواع فعالیت‌های انسانی از نظر پیچیدگی و مدت زمان[۲].

اغلب روش‌های شناسایی فعالیت برای توصیف فعالیت‌ها از ویژگی‌های سطح پایین و متوسط استفاده می‌کنند و دانش انسان درباره فعالیت‌ها را در شناسایی در نظر نمی‌گیرند، این روش‌ها برای فعالیت‌های ساده مناسب هستند ولی در وضعیت های پیچیده شکست می‌خورند. نشان داده شده است که معانی نقش کلیدی در شناسایی و ادراک بصری انسان دارند. یکی از قابلیت‌های برجسته

روش‌های شناسایی استفاده کننده از ویژگی معنایی، زمانی که فعالیت‌های مشابه متفاوت به نظر برسند، قابل اطمینان تر هستند و همچنین می‌توانند به نرخ دقت بالایی در شناسایی فعالیت دست یابند. از طرفی روش‌های شناسایی فعالیت معنایی برای یادگیری معمولا به مجموعه داده آموزشی وابسته هستند. از این رو، این روش‌ها نسبت به تغییرات مجموعه داده، مقاوم نیستند.

۱-۲ ضرورت انجام مسئله

به دلیل افزایش نگرانی‌های امنیتی و هزینه‌های کم سخت افزاری تقاضای رو به رشدی برای سیستم‌های نظارت هوشمند بوجود آمده است. در سیستم‌های نظارتی غیر خودکار نیروی انسانی مورد نیاز برای نظارت، گران است. در نتیجه ویدئوی ضبط شده از این دوربین‌ها معمولاً کم نظارت شده، یا اصلاً نظارت نمی‌شود. آنها اغلب تنها به عنوان منبعی برای رجوع برای حادثه‌هایی که صورت گرفته، استفاده می‌شوند. از طرف دیگر ناظر، معمولاً جنبه‌های خاصی از فعالیت و رفتار انسان را به منظور پیش بینی حوادث خطرناک مد نظر قرار می‌دهد. بنابراین اگر این دوربین‌ها به جای تنها ضبط ویدئو، بتوانند برای شناسایی فعالیت‌ها و اقدام در زمان واقعی مورد استفاده قرار بگیرند، می‌توانند ابزاری به مراتب مفیدتر باشند.

۱-۳ کاربردهای شناسایی فعالیت

- شناسایی فعالیت‌های غیر عادی و مشکوک در مکان‌های عمومی مانند فرودگاه‌ها و ایستگاه‌های مترو و مراکز خرید بوسیله سیستم های نظارتی هوشمند
- قابلیت مراقبت و نظارت بر بیماران، کودکان و افراد سالمند در خانه‌های هوشمند
- شناسایی و تحلیل رفتار انسان‌ها: در برخی از مکان‌ها لازم است غیرطبیعی بودن رفتار انسان‌ها یا وسایل نقلیه در شرایط خاص بطور مثال جلوگیری از سرقت، بررسی شود.

- امکان تعامل با رایانه و بازی‌های رایانه‌ای
- نظارت متقابل با استفاده از چند دوربین: در محیط‌های اجتماعی بزرگ و وسیع از چند دوربین برای کنترل امنیتی استفاده می‌شود. تحلیل و نظارت غیرخودکار با استفاده از افراد در این حالت‌ها تقریباً غیرممکن و یا غیرموثر است.

۱-۴ چالش های موجود در شناسایی فعالیت

با وجود اینکه پیشرفت‌های مهمی در شناسایی فعالیت صورت گرفته است، بسیاری از الگوریتم‌های پیشرفته هنوز فعالیت‌های ویدئویی را اشتباه طبقه‌بندی می‌کنند. شناسایی فعالیت انسان به دلایلی چون تغییرات در عملکرد حرکت و تنظیمات[1] مختلف تصویر برداری و تفاوت‌های بین انسانی، بسیار چالش برانگیز است. مواردی از این چالش‌ها را در زیر مرور می‌کنیم [۳].

- **تنوع درون کلاسی و شباهت بین کلاسی**

همانطور که می‌دانیم، افراد یک فعالیت مشخص را متفاوت انجام می‌دهند. برای یک فعالیت داده شده، برای مثال، "دویدن"، یک فرد می‌تواند سریع، آهسته و یا حتی دویدن با پرش داشته باشد. می‌توان گفت، یک کلاس فعالیت ممکن است شامل چندین شکل متفاوت از حرکت انسان باشد. علاوه براین، ویدئوهای فعالیت‌های مشابه می‌تواند در نقاط دید مختلف ضبط شده باشد. آن‌ها می-توانند از روبروی فرد، در کنار و یا حتی از بالای سر فرد گرفته شوند. از این گذشته، افراد مختلف ممکن است شکل بدن مختلفی برای انجام دادن فعالیت یکسان نشان دهند. تمام این موارد تنوع درون کلاسی را بوجود می‌آورند، که بسیاری از الگوریتم های شناسایی فعالیت را دچار اشتباه می‌کند. شکل (۱-۲) نمونه‌ای از تنوع درون‌کلاسی فعالیت شوت کردن را نشان می‌دهد.

[1] setting

علاوه براین، شباهت‌هایی بین فعالیت‌های مختلف وجود دارد. برای مثال، "دویدن" و "راه رفتن" شامل الگوهای حرکتی مشابه در افراد است. این شباهت‌ها چالشی برای طبقه‌بندی فعالیت‌ها هستند.

شکل ۱-۲ تنوع درون کلاسی فعالیت شوت کردن

- **تغییرات محیطی**

تعدادی از الگوریتم‌های شناسایی فعالیت انسانی در محیط‌های داخلی عملکرد بسیار خوبی دارند اما در محیط‌های بیرون با شکست روبرو می‌شوند. این امر عمدتا بدلیل نویز پس‌زمینه است. چرا که در هنگام استخراج ویژگی، نویز پس زمینه ویژگی استخراج شده را خراب می‌کند که باعث کاهش عملکرد شناسایی می‌شود.

حرکت دوربین عامل دیگری است که باید در برنامه‌های کاربردی دنیای واقعی در نظر گرفته شود. به دلیل اهمیت حرکت دوربین، ویژگی‌های فعالیت نمی‌توانند دقیق استخراج شوند. برای استخراج بهتر ویژگی‌های فعالیت، حرکت دوربین باید مدل شده و جبران شود.

موضوعات وابسته به محیط دیگری، مانند وضعیت‌های روشنایی، تغییرات نقاط دید و پس-زمینه دارای حرکت، عواملی هستند که باعث به وجود آمدن چالش‌هایی برای استفاده ازالگوریتم‌های شناسایی فعالیت در دنیای واقعی می‌شوند.

- **داده ناکافی**

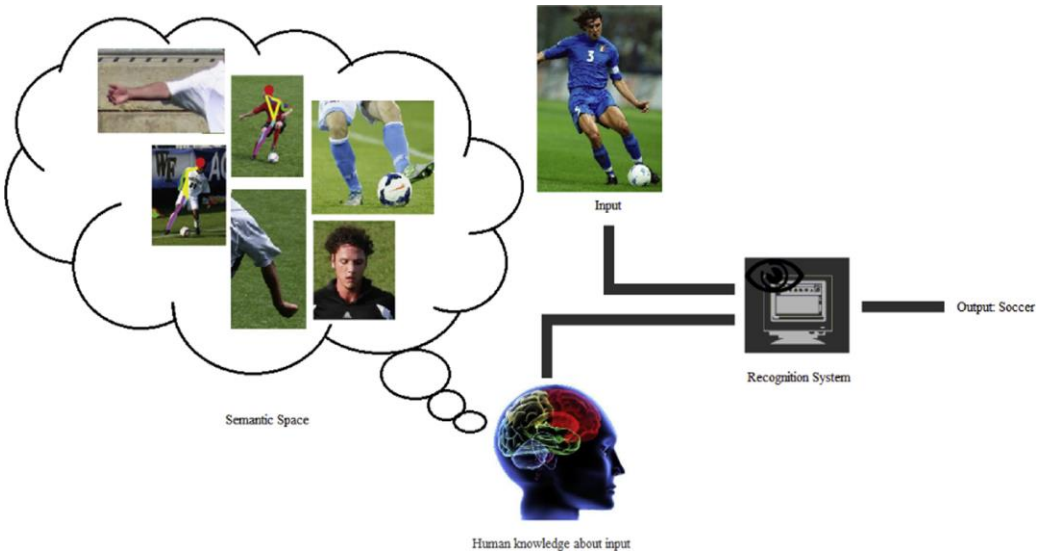
یک سیستم هوشمند مقاوم و کارآمد معمولا نیازمند تعداد زیادی داده آموزشی است. به دلیل در دسترس نبودن پایگاه داده‌ای بزرگ برای شناسایی فعالیت، چالشی بزرگ در این امر به وجود می-

آید. بسیاری از پایگاه داده‌های فعالیت موجود تنها شامل چندین ویدئو است که برای آموزش یک الگوریتم شناسایی فعالیت پیچیده کافی نیست.

۱-۵ فضای معنایی

برخلاف ویژگی‌های سطح پایین، ویژگی‌های معنایی، خصوصیات فعالیت‌ها را توصیف می‌کند. ویژگی‌های معنایی برای رفع مشکل تنوع درون‌کلاسی و چگونگی ایجاد تمایز بین فعالیت‌های مشابه مفید هستند. توانایی افراد برای شناسایی فعالیت‌ها تنها به تحلیل و بررسی بصری حالت بدن انسان تکیه نمی‌کند، همچنین نیازمند منابع اضافی از اطلاعات مانند زمینه و یا صحنه، دانش مرتبط با فعالیت‌ها، و یا دانش در مورد خصوصیات بصری آنها است.

ضیایی‌فرد و همکارانش در مرجع [۴]، یک فضای ویژگی که "فضای معنایی" نامیده می‌شود، را معرفی می‌کنند، که شامل دانش انسانی درباره فعالیت‌ها مانند ویژگی‌های pose ، poselet ، شی ، صحنه و مشخصه[1] (صفات) فعالیت است. فضای معنایی بصورت شماتیک در شکل (۱-۳) معرفی شده‌است، که در آن شی مرتبط "توپ فوتبال" ، poselet "دست دراز شده"، pose "شکل خاص بدن در بازی فوتبال" و مشخصه فعالیت "نگاه رو به جلو سر" است.



شکل ۱-۳ مفهوم فضای معنایی[۴].

[1] attribute

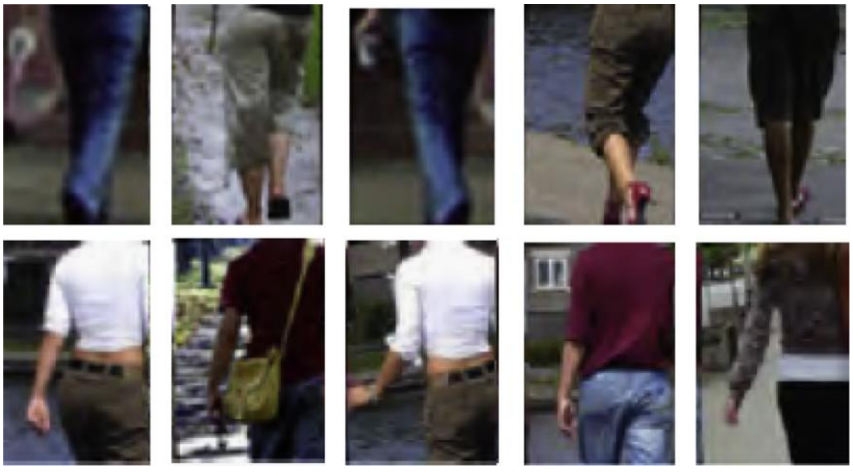
اجزای فضای معنایی معرفی شده درمرجع [۴]، در بخش‌های زیر شرح داده شده‌اند.

Pose

مطالعات اخیر نشان داده است که ویژگی‌های مبتنی بر شکل بدن در هنگام فعالیت، بهتر از ویژگی‌های سطح پایین است خصوصاً در مواردی که بدلیل نویز استخراج ویژگی‌های محلی نامعتبر است. اطلاعاتی در شکل بدن افراد نهفته است که برای تمایز فعالیت‌ها موثر و کافی است. روش‌های شناسایی فعالیت مبتنی بر pose می‌توانند با استفاده از تخمین pose بعنوان ورودی در مرحله شناسایی فعالیت مورد استفاده قرار بگیرند. این روش دارای این مزیت است که خطاهای ناشی از تخمین نادرست pose تاثیر کمتری روی کیفیت نهایی شناسایی فعالیت دارد [۴].

Poselet

در تشخیص معانی فعالیت، قسمت های بارز بدن در طول فعالیت در نظر گرفته می‌شود که تحت عنوان poselet می‌گویند. معمولا بیشتر از یک قسمت از بدن را شامل می‌شود، مانند نیمی از بدن و دست چپ یا نمای جلویی از پاهای در حال راه رفتن. مزیت اصلی این ویژگی‌ها این است که آنها قسمت‌های مهم بدن که در فعالیت نقش داشته‌اند را ثبت می‌کنند. شکل (۱-۴) مثالی از قسمت-های بارز بدن در راه رفتن را نشان می‌دهد که در آن سطر بالا بخش پاها و سطر پایین بخش تنه و دست‌ها را مشخص می‌کند. هرکدام از قسمت‌ها ممکن است از نظر بصری متفاوت به نظر برسند، اما آنها مفهوم معنایی تقریباً یکسانی دارند.



شکل ۱-۴ مثالی از Poselet مربوط به فعالیت"راه رفتن" [۵].

شی و صحنه

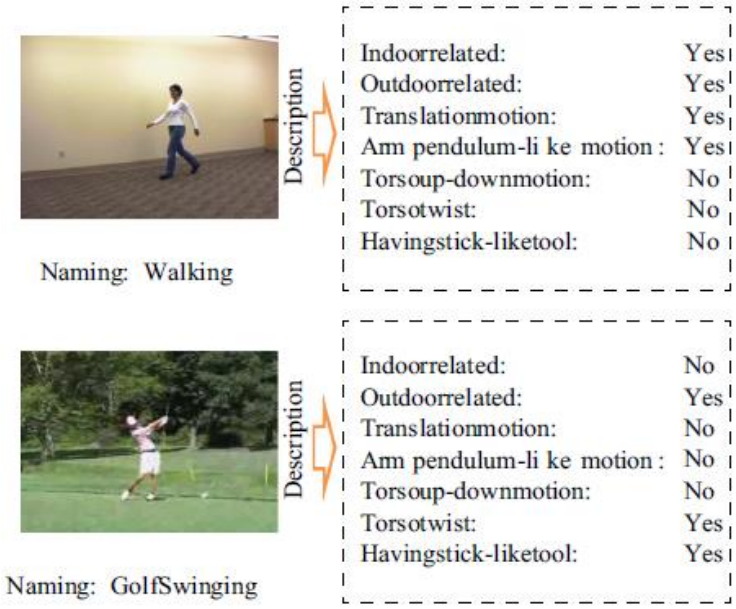
اطلاعات معنایی همچنین می تواند از صحنه‌ای که فعالیت در آن رخ داده است، استخراج شود. برای مثال، اگر زمینه فعالیت بعنوان یک زمین فوتبال شناسایی شود، این احتمال قوی است که فعالیت مربوطه "بازی فوتبال" است. علاوه بر صحنه، اشیای مربوط به تعامل بین انسان و اشیا[1] در شناسایی معنایی فعالیت استفاده می‌شود. با توجه به نوع فعالیت انسان، اشیای مربوط به آن ممکن است وجود داشته باشد. دانستن اشیای مرتبط به شناسایی فعالیت‌های مربوطه کمک می‌کند. بعنوان مثال، یک اسب با انسان احتمالا مربوط به فعالیت "اسب سواری" است، درحالی که یک تلفن با یک فرد می تواند به عمل "تلفن کردن" مرتبط باشد.

اگرچه صحنه می تواند به شناسایی فعالیت کمک کند، اما در مواردی که صحنه شلوغ باشد، ممکن است تاثیر منفی در شناسایی داشته باشد. علاوه براین، ممکن است صحنه شامل فعالیت‌های مختلفی باشد و در نتیجه اطلاعات مفیدی برای ایجاد تمایز بین فعالیت‌ها ارایه نکند.

صفات

صفات یکی دیگر از انواع مهم ویژگی‌های معنایی هستند که می‌توانند بطور مستقیم با خصوصیات بصری فعالیت‌ها مرتبط باشند. صفات، مکان و زمان حرکات فرد را توصیف می‌کنند. برای مثال، حرکت شبیه پاندول بازو و یا الگوی حرکت دوپا، قراردادن یک پا در کنار پای دیگر، مشخصه های بالقوه برای فعالیت "راه رفتن" هستند. صفات دو فعالیت راه رفتن و بازی گلف در شکل (۱-۵) نشان داده شده است.

[1] Human object interaction



شکل ۱-۵ مثالی از صفات فعالیت راه رفتن و گلف بازی[۶].

۱-۶ هدف و ساختار این پایان نامه

در این پایان نامه به منظور ارتقا و بهبود دقت شناسایی فعالیت پیشنهاد می‌کنیم که با ترکیبی از ویژگی‌های معنایی و غیرمعنایی به شناسایی فعالیت بپردازیم.

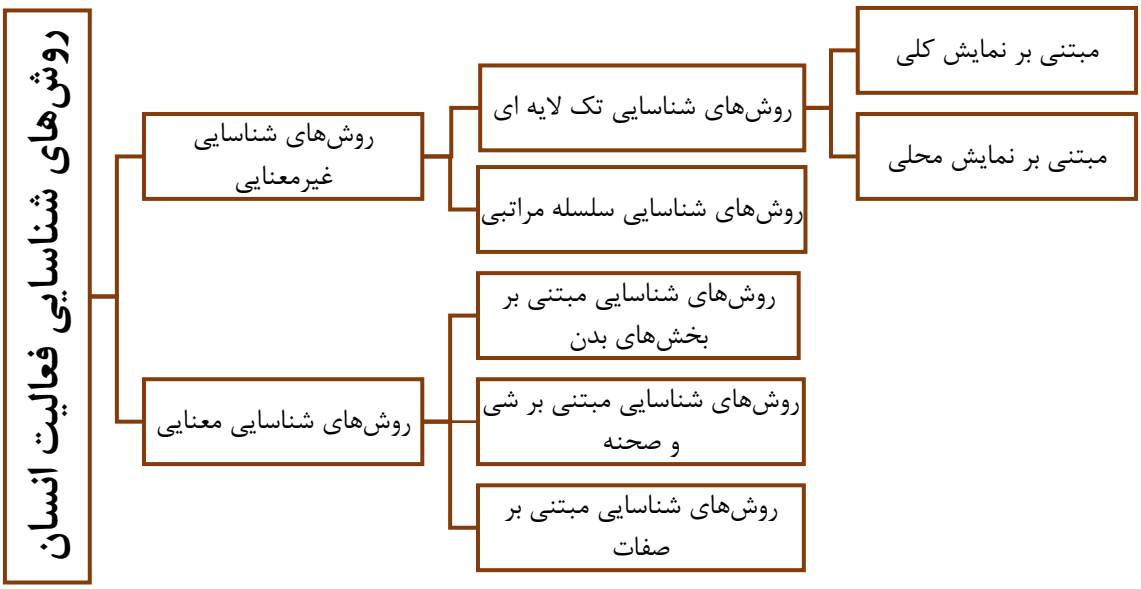
این پایان نامه در ۵ فصل به شکل زیر تدوین شده است: در فصل دوم مروری بر کارهای گذشته در شناسایی فعالیت را بررسی و مقایسه می‌کنیم. سپس در فصل سوم مبانی نظری بکارگرفته شده را شرح می‌دهیم. در فصل چهارم روش پیشنهادی این پایان نامه معرفی خواهد شد. فصل پنجم به نتایج شبیه‌سازی، مقایسه و تحلیل آن‌ها می‌پردازد.

فصل ۲: مروری بر کارهای گذشته

۲-۱ مقدمه

تجزیه و تحلیل حرکت و فعالیت‌ها تاریخچه‌ای طولانی دارد و مورد توجه رشته‌های مختلفی از جمله: روانشناسی، زیست شناسی و علوم رایانه قرار گرفته است [۷].

در سال های اخیر مطالعات زیادی در زمینه شناسایی فعالیت انسان از ویدئو انجام شده است که تمام آنها سعی بر افزایش دقت طبقه‌بندی و حل چالش‌های موجود در این موضوع را دارند. سطوح مختلفی از ویژگی‌ها در روش‌های شناسایی فعالیت استفاده شده است. یکی از راه حل‌ها برای افزایش دقت شناسایی فعالیت‌ها همانطور که در فصل اول گفته شد استفاده از ویژگی‌های معنایی است. ابتدا روش‌های شناسایی غیر معنایی را مرور کرده، سپس با تعریف فضای معنایی و بکارگیری ویژگی‌های معنایی، روش‌های شناسایی معنایی را بررسی و عملکرد روش‌ها را مقایسه می‌کنیم. دسته بندی روش‌های مرور شده در بلوک دیاگرام شکل (۲-۱) نمایش داده شده است.



شکل ۲-۱ طبقه بندی روش‌های شناسایی فعالیت انسانی

۲-۲ روش‌های شناسایی فعالیت غیر معنایی

روش‌های شناسایی فعالیت غیر معنایی را می‌توان به دو دسته تک لایه‌ای و سلسله مراتبی تقسیم کرد، که در ادامه به مرور این روش‌ها می‌پردازیم.

۲-۲-۱ روش‌های شناسایی فعالیت تک لایه‌ای

روش‌های شناسایی فعالیت تک لایه‌ای از داده‌های ویدئویی خام مستقیماً استفاده می‌کنند. در نتیجه بیشتر روش‌های شناسایی تک لایه‌ای با ویدئو یا پایگاه داده‌های ساده مانند KTH سر و کار دارند. روش‌های تک لایه‌ای با شناسایی فعالیت‌های ساده می‌توانند با استفاده از روش‌های سلسله-مراتبی برای شناسایی فعالیت‌های پیچیده‌تر بکار گرفته شوند [۸].

روش‌های شناسایی تک لایه‌ای را می‌توان به دو گروه روش‌های مبتنی بر نمایش کلی و نمایش محلی دسته بندی کرد. روش‌های مرتبط با نمایش‌های کلی در زمینه شناسایی فعالیت عمدتاً بین سال‌های ۱۹۹۷ تا ۲۰۰۷ بسیار مورد مطالعه قرار گرفته‌اند. روش‌های شناسایی مبتنی بر نمایش کلی در چالش‌هایی مانند وابستگی به زاویه دید و انسداد عملکرد خوبی ندارند و همین باعث پیدایش بازنمایی‌های محلی شده است.

۲-۲-۱-۱ روش‌های شناسایی مبتنی بر نمایش کلی

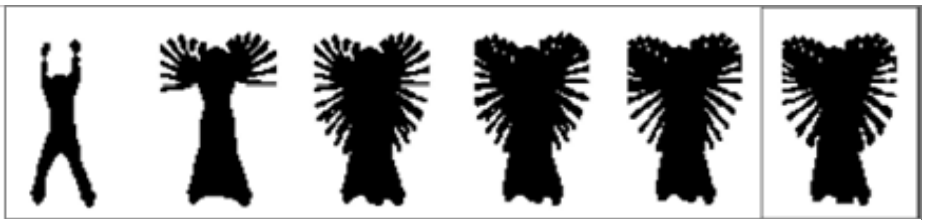
یک ویدئو می‌تواند بعنوان مجموعه‌ای از پیکسل‌ها در فضای سه بعدی XYZ که حجمی در فضا و زمان را تشکیل می‌دهند، نمایش داده شود. این حجم، اطلاعات لازم را برای شناسایی فعالیت‌های رخ‌داده دارا می‌باشد. شناسایی یک فعالیت در ویدئوی ورودی از طریق محاسبه میزان شباهت این حجم یا ویژگی استخراج شده از آن، بین ویدئوی ورودی و مدل فعالیت‌ها صورت می‌پذیرد [۹].

روش‌های شناسایی حجم مکان-زمان از حجم مکان-زمان به عنوان ویژگی یا الگو برای شناسایی استفاده می‌کنند و بوسیله معیار شباهت بین دو حجم ویدئو آموزشی و آزمون کلاس فعالیت

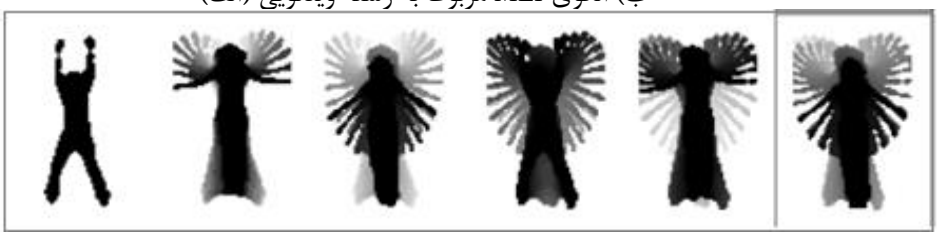
را مشخص می‌کنند. ایرادی که این روش‌های شناسایی فعالیت دارند این است که نسبت به نویز مقاوم نیستند و دارای اطلاعات کمی از پس زمینه هستند [۸].

یکی از کارهای تاثیرگذار در دسته روش‌های شناسایی حجم مکان-زمان توسط Bobick و همکارانش در مرجع [۱۰]، گزارش شده است. در این روش تصویر انرژی حرکت[1](MEI) و تصویر تاریخ حرکت[2](MHI) بعنوان معیار شباهت از هر ویدئو استخراج می‌گردد. در این روش ایده اصلی، کد کردن اطلاعات مربوط به حرکت با استفاده از تعدادی فریم ویدئویی است. فریم‌های بکار رفته تصاویری باینری هستند که حرکت را توصیف می‌کنند. الگوی MHI نشان دهنده چگونگی تغییرات تصویر حرکت است. هر پیکسل در MHI تابعی از تاریخچه اطلاعات زمانی حرکت در آن نقطه است. بطور مثال شدت روشنایی کم، به حرکات اخیر بیشتر مربوط است شکل (۲-۲).

الف) رشته فریم‌های ویدئویی مربوط به فعالیت پریدن



ب) الگوی MEI مربوط به رشته ویدئویی (الف)



ج) الگوی MHI مربوط به رشته ویدئویی (الف)

شکل ۲-۲تصویر تاریخچه حرکت و تصویر انرژی حرکت[۱۰]. الف) رشته فریم‌های ویدئویی مربوط به فعالیت پریدن، ب) تصاویر متناظر با الگوی MEI، ج) تصاویر متناظر با الگوی MHI

[1] Motion Energy Image (MEI)
[2] Motion History Image (MHI)

روش حجم مکان_زمان توسط [۱۱]، توسعه و به عنوان الگوهای تصویر انرژی حرکت معرفی شده است. در این تحقیق ایده اصلی، نمایش فعالیت بوسیله شکل سه بعدی بدست آمده از سایه‌های فعالیت در مکان-زمان است (شکل۲-۳).

(الف) (ب) (ج) (د) (ه)

شکل ۲-۳ الف)، ب)، ج): حجم های مکان-زمان برای سه نوع حرکت [۱۱]. د) و ه): سطوح مکان-زمان مربوط به فعالیت تنیس و پیاده روی [۱۲].

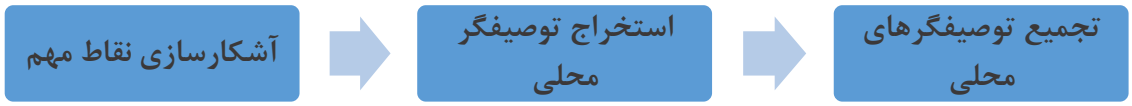
در این روش به منظور طبقه‌بندی، با محاسبه میانگین زمان در هر نقطه، رویه سه بعدی به نقشه دوبعدی تبدیل می‌شود. مطالعه انجام شده در [۱۲]، با در نظر گرفتن شکل و حرکت فرد، فعالیت را مدل‌سازی کردند. با تصویر کردن نقاط مرز بدن فرد در حال انجام دادن فعالیت، تصویر دوبعدی از مرز فعالیت تشکیل می‌شود. به کمک دنباله‌ای از تصاویر مرز ناحیه فعالیت، حجم مکان-زمان[1] STV تولید می‌شود. به کمک برخی توصیفگرهای فعالیت که بر روی STV تعریف شده‌اند، شناسایی فعالیت صورت می‌گیرد (شکل ۲-۳).

۲-۲-۱-۲روش‌های مبتنی بر نمایش محلی

در مرجع [۱۳]، روشی مبتنی بر نمایش محلی برای شناسایی فعالیت پایه‌گذاری شد، این روش از نقاط مهم فضا-زمان[2] (STIPs) بهره برداری می‌کند. روش‌های شناسایی فعالیت مبتنی بر نمایش محلی از روندنمای (۲-۴)، پیروی می‌کنند [۷].

[1] spatiotemporal volume
[2] Spatiotemporal interest point

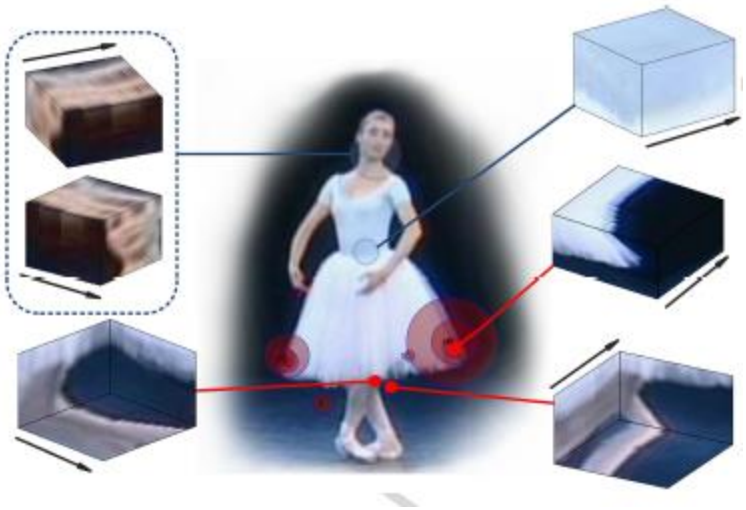
در ادامه ایده‌های اصلی و پیشرفت‌های عمده انجام شده برای هریک از مؤلفه‌های روندنمای (۲-۴) را بصورت جداگانه بررسی می کنیم.



شکل ۲-۴ روند شناسایی فعالیت مبتنی بر نمایش محلی

- **آشکارسازی نقاط مهم**

برای یافتن نقاط مهم مکان-زمانSTIP، Laptive و همکارانش آشکارساز گوشه هریس[1] [۱۴]، را به آشکارساز هریس سه بعدی ارتقا و گسترش دادند [۱۳]. در آشکارساز گوشه هریس سه بعدی، علاوه بر ساختار فضایی اطلاعات در هر فریم، این اطلاعات در طول زمان نیز در نظر گرفته می‌شوند. ایده اصلی آشکارساز گوشه هریس دو بعدی، پیدا کردن موقعیت‌های مکانی با تغییرات مهم در دو جهت متعامد است. آشکارساز هریس سه بعدی نقاط با تغییرات مکانی زیاد و حرکات غیر یکنواخت را شناسایی می‌کند. به عنوان مثال در شکل (۲-۵) نقاط مهم مکان-زمان مشخص شده است. تغییرات مکانی در طول زمان با فلش نشان داده شده است.



شکل ۲-۵ بخش قرمز رنگ مشخص شده نقاط مهم فضایی زمانی هستند[۱۳].

[1] Harris Corner detector

یکی دیگر از آشکارسازهای نقاط مهم در فضای دوبعدی، آشکارساز هسین[1] است، که این آشکارساز به مدل سه بعدی نیز ارتقا یافته است [۱۵]. در آشکارساز هسین سه بعدی از مشتق دوم روشنایی نقاط برای تشخیص نقاط مهم، استفاده می‌شود.

برخلاف تصاویر، ویدئوهای فعالیت معمولا در محیط های کنترل نشده ضبط شده‌اند در نتیجه احتمال استخراج نقاط مهم در منطقه غیرمرتبط با فعالیت، بالا است. برای مثال، یک دوربین دارای لرزش می تواند نقاط غیر مرتبط را مهم در نظر بگیرد. برای رسیدگی به این موضوع، در مرجع [۱۶]، ویژگی‌های نامرتبط با استفاده از خواص آماری فیلتر می‌شوند. علاوه بر این، ویژگی‌های مکان-زمان به‌دست آمده از پس‌زمینه، بعنوان ویژگی‌های ایستا شناخته شده‌اند [۱۶].

- **توصیفگر محلی**

برای توصیف محلی در یک نقطه مهم، معمولا از پیکسل‌های درون یک مکعب[2] به مرکز نقطه استفاده می‌شود [۱۳،۱۷]. در مراجع [۱۸،۱۹] بصورت جداگانه مطالعاتی درباره انتخاب مکعب‌های دارای شکل ثابت برای شناسایی فعالیت انجام شد و اصطلاح مسیرحرکت[3] معرفی گردید. در ادامه، ابتدا توصیفگرهای محلی گوناگون که بطور گسترده در شناسایی فعالیت استفاده شده‌اند را مورد بحث و بررسی قرار می‌دهیم، سپس مسیرهای متراکم[4] و اهمیت شان را مرور می‌کنیم.

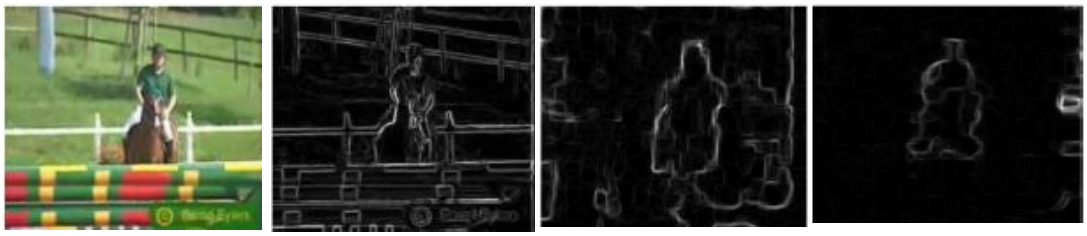
توصیفگرهای حرکت و لبه

Klaser و همکارانش در [۲۰]، استفاده از هیستوگرام گرادیان جهت دار [5] HOG را بعنوان توصیفگر حرکت پیشنهاد دادند.

میدان‌های شار نوری[1] حرکت پیکسل‌ها در ویدئو را کد می‌کنند. در مرجع [۲۱]، هیستوگرام شار نوری[2] HOF در تمامی نواحی محلی بعنوان توصیفگر زمان-مکانی پیشنهاد شده است. توسعه

[1] Hessian
[2] cuboid
[3] trajectory
[4] Dense trajectories
[5] Histogram of oriented gradients

یافته‌تر و مقاوم‌تر ازتوصیفگرHOF در برابر حرکت دوربین، توصیفگر هیستوگرام مرز حرکت[3] MBH است که در [۲۲]، معرفی شده است. هیستوگرام مرز حرکت در سراسر میدان‌های مرز حرکت محاسبه شده است. این توصیفگر از مشتقات مکانی میدان‌های شار نوری به دست می‌آید (شکل ۲-۶).

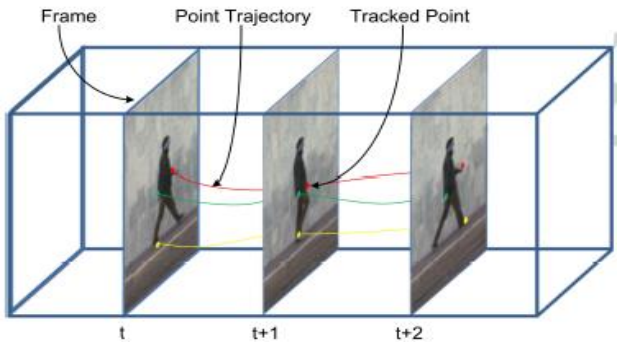


(الف) (ب) (ج) (د)

شکل ۲-۶ الف)فعالیت اسب سواری ب)مشتق مکانی در جهت افقی ج) مشتق مکانی در جهت عمودی د)تصویر حرکت مرز[۷].

از مکعب‌ها تا مسیر های حرکت

بعضی از نقاط مهم مکان_زمانی ممکن است دقیقا درون مکعب گسترش یافته زمانی تعریف‌شده، قرار نداشته باشند. از این رو، ویژگی‌های استخراج شده از مکعب‌ها ممکن است برای توصیف نقاط مهم لازم نباشد. مسیرحرکت، ردیابی صحیح یک ویژگی در طول زمان است. شکل (۲-۷) مفهوم مسیر حرکت را نشان می دهد.



شکل ۲-۷مسیرهای حرکت نقاط ردیابی شده در طول فریم ها [۷].

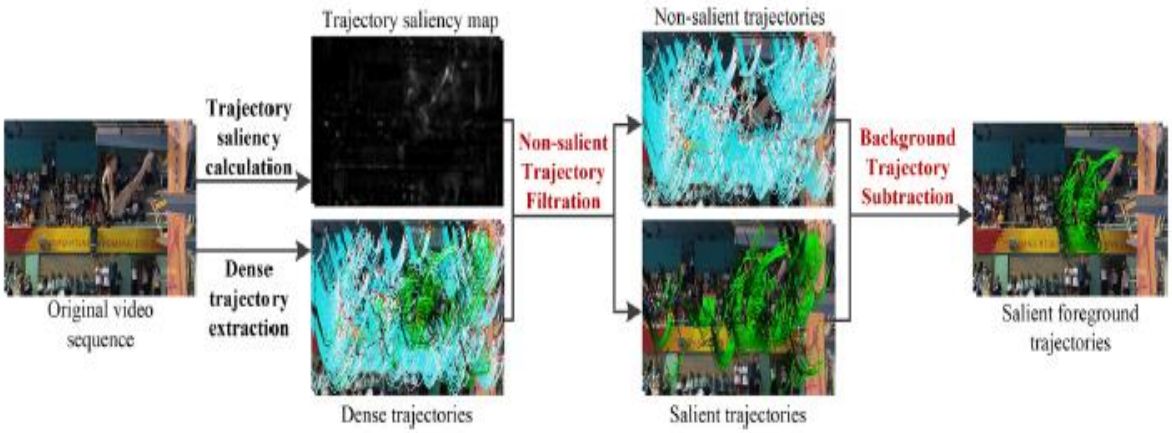
[1] Optical flow fields
[2] Histogram of optical flow
[3] Histogram of motion boundary

در ارزیابی که برای اولین بار در زمینه شناسایی فعالیت توسط wang و همکارانش انجام شد [۲۳] ، مشاهده شد که نمونه برداری متراکم[1] از مکان و زمان نسبت به استخراج نقاط مهم پراکنده[2] عملکرد بهتری دارد. در این روش مسیر حرکت بوسیله شار نوری متراکم استخراج گردید. در این روش همراه نقاطی که بصورت متراکم انتخاب می‌شوند، ویژگی‌های هیستوگرام گرادیان جهت‌دار، هیستوگرام جربان نوری و هیستوگرام مرز حرکت و همچنین مسیر حرکت برای هر ۱۵ فریم متوالی بدست می‌آید. سپس با کنار هم قرار دادن این ویژگی‌ها یک بردار ویژگی به طول ۴۳۶ بدست می‌آید. این روش در مقایسه با روش ردیابی نقاط KLT[3] به دلیل نمونه برداری متراکم، در برابر حرکات نامنظم ناگهانی بهتر عمل می‌کند. شکل (۲-۸) مقایسه این دو روش را نشان می‌دهد. در مرجع [۲۴]، علاوه بر توصیفگرهای همراه مسیر حرکت از توصیفگر مرز حرکت که به اختلاف شار نوری وابسته است، استفاده شده که منجر به بهبود روش قبلی شده است.

شکل ۲-۸مقایسه ردیابی لوکاس-کانادی و مسیرهای حرکت متراکم[۲۳] .

[1] Dense sampling
[2] sparse
[3] Kanade-Lucas-Tomasi(KLT) Tracker

به تازگی در مطالعه انجام شده توسط [۲۵]، روشی برای ارتقا عملکرد شناسایی فعالیت با استفاده از استخراج مسیرهای حرکت شاخص پیش‌زمینه[1] پیشنهاد شده است. این روش بر اساس تشخیص شاخص بودن[2] و بازیابی ماتریس رتبه پایین برای یادگیری ویژگی‌های متمایز کننده از زمینه پیچیده ویدئو ارایه شده است. همچنین روشی برای ترکیب اطلاعات شاخص ظاهر و حرکت به منظور جداسازی مسیر حرکت مهم و غیرمهم پیشنهاد شده‌است که شامل دو مرحله، فیلتر مسیرهای حرکت غیر شاخص و تفریق مسیرحرکت پس زمینه برای استخراج مسیرهای‌حرکت شاخص پیش زمینه است. نشان داده شد که مسیرهای حرکت شاخص، مربوط به پیش‌زمینه و بقیه مربوط به پس‌زمینه هستند. در نهایت ویژگی‌های مسیر حرکت شاخص پیش زمینه بوسیله بردار فیشر کد می شوند. شکل(۲-۹) نحوه استخراج مسیر حرکت شاخص را نشان می‌دهد که در آن خطوط سبز نشان دهنده مسیر حرکت شاخص و خطوط آبی مسیر حرکت غیر شاخص است.



شکل ۲-۹ روش استخراج مسیر حرکت برجسته پیش زمینه[۲۵].

[1] Foreground salient trajectories
[2] saliency

- **تجمیع[1] توصیفگرهای محلی**

پس از استخراج توصیفگرهای محلی از ویدئو می‌بایست در یک فرآیند یادگیری، فعالیت‌ها طبقه‌بندی و مقایسه شوند. از آنجا که تعداد ویژگی‌های محلی هر ویدئو متفاوت از ویدئوی دیگر است و الگوریتم‌های یادگیری مانند ماشین بردار پشتیبان بردارهایی با ابعاد یکسان را بعنوان ورودی می-پذیرند، نیازمند روشی برای تجمیع ویژگی‌های محلی به منظور تبدیل به توصیفگرهای با ابعاد یکسان برای تمام ویدئوها هستیم.

با استفاده از روش کیف کلمات[2] می‌توان فراوانی بردارهای ویژگی محلی مرجع را در هر ویدئو بدست آورد. تئوری کیف‌کلمات در فصل سوم شرح داده می‌شود. مرجع [۱۷]، جزء اولین روش‌هایی است که در آن برای شناسایی فعالیت از کیف کلمات استفاده شده‌است. در فرم اصلی آن، اطلاعات زمانی بوسیله کیف کلمات در نظر گرفته نشده است. برای رفع این نقص، Laptev و همکارانش [۲۱]، شبکه‌های مکان_زمانی را پیشنهاد دادند. ایده اصلی این روش بخش‌بندی ویدئو به چند زیربخش، و کنارهم قرار دادن توصیفگرهای محلی هر زیربخش برای تشکیل "کانال" و مقایسه ویدئوها براساس توصیفگرهای کانال‌های هر ویدئو است.

در مراجع [۲۶،۲۵،۲۳]، برای درکنارهم قرار دادن بردارهای ویژگی از کد کننده بردار فیشر استفاده شده‌است. در روش بردار فیشر کد شده، بردارهای ویژگی استخراج شده از ویدئو در کنارهم قرار می‌گیرند و برای هر مجموعه از بردارهای ویژگی محلی یک بردار فیشر با طول ثابت بدست می-آید. از روش بردار فیشر کد کننده همراه با توصیفگر مسیر حرکت عملکرد خوبی گزارش شده‌است [۲۸،۲۳]. تحلیل و بررسی بردارهای فیشر با جزئیات بیشتر در حوزه شناسایی فعالیت در مرجع [۲۷]، انجام شده‌است.

[1] Aggregation
[2] Bag of words

اخیرا از نمایش تنک[1] در بسیاری از مطالعات حوزه پردازش سیگنال استفاده شده‌است. با بکارگیری این روش می‌توان یک تصویر را به کمک ترکیب خطی تعداد کمی از اتم‌های یک دیکشنری یاد گرفته‌شده تقریب زد. این بازنمایی در حوزه‌های مختلف پردازش تصویر مورد مطالعه قرار گرفته - است. در مرجع [۲۹]، با استفاده از نمایش تنک روشی برای خوشه بندی زیرفضا [۳۰]، بر روی ویژگی مسیر حرکت پیشنهاد داده شده و با استفاده از خوشه‌ها، حرکت‌های محلی شناسایی شده‌است. سپس با بکارگیری حرکات محلی در طبقه بند KGS[2] فعالیت‌ها شناسایی شده‌است.

۲-۲-۲ روش‌های شناسایی فعالیت سلسله مراتبی

همانطور که در مرجع [۱]، شرح داده شده است، روش‌های سلسله مراتبی سعی بر شناسایی فعالیت‌های سطح بالا براساس فعالیت‌های ساده‌تر دارند. به بیان دیگر، یک فعالیت سطح بالا می‌تواند بصورت دنباله‌ای از زیرفعالیت تجزیه شود که به این زیرفعالیت‌ها عمل های بنیادی[3] نیز گویند.

مزیت روش‌های سلسله مراتبی توانایی در مدل کردن ساختارهای پیچیده فعالیت‌های انسان و انعطاف‌پذیری آن‌ها برای فعالیت‌های فردی، تعاملات بین انسان‌ها، اشیا و یا فعالیت‌های گروهی است.

علاوه براین، مدل‌های سلسله مراتبی ارتباط بصری و راحتی را برای بکارگیری دانش قبلی و درک ساختار فعالیت‌ها ارایه می دهند.

در روش سلسله مراتبی آماری از مدل مخفی مارکوف[4] استفاده شده‌است. بدین صورت که سطوح چندتایی حالت های مخفی HMM نمایشی از فعالیت‌های سلسله مراتبی انسان را تشکیل می‌دهد. در[۳۱-۳۳]، از HMM سلسله مراتبی دو لایه برای شناسایی فعالیت استفاده شده است. در مرجع [۳۴]، مدل سطح متوسط برای شناسایی فعالیت توسعه داده شده‌است، که در آن مسیرهای

[1] Sparse representation
[2] Kernel group sparse
[3] Atomic action
[4] Hidden markov model (HMM)

حرکت متراکم برای تشکیل قسمت‌های یک فعالیت خوشه‌بندی شده‌اند و روابط متقابل بوسیله مدل گرافیکی توصیف شده‌اند.

در تحقیق گزارش شده در مرجع [۳۵]، برای شناسایی فعالیت‌های پیچیده یک روش انطباق سلسله مراتبی برای در نظر گرفتن ارتباط مکان-زمان میان نقاط ویژگی، پیشنهاد داده شده‌است.در این روش، انطباق ارتباط مکان-زمان برای اندازه‌گیری شباهت ساختاری میان ویژگی‌های استخراج شده از ویدئوی آزمون و آموزش انجام می‌شود.

در شناسایی فعالیت دو مسئله وابستگی زمانی و وابستگی مکانی نقش کلیدی دارند. اکثر روش‌های اخیر تنها روی یکی از این دو مورد تمرکز کرده‌اند، از این رو توانایی توصیف کارآمد برای شناسایی فعالیت پیچیده را ندارند. Wanru و همکارانش در [۳۶]، برای حل این مشکل یک مدل مکان-زمان سلسله مراتبی HSTM[1] بوسیله مدلسازی همزمان محدودیت‌های مکان-زمان پیشنهاد دادند. HSTM پیشنهاد شده از دو لایه HCRF[2] تشکیل شده که در لایه زیرین ارتباط مکانی در هر فریم و در لایه بالایی، ویژگی‌های سطح بالا را برای دسته بندی ارتباطات زمانی در کل دنباله ویدئویی بکار می‌گیرند. این الگوریتم در شناسایی تعاملات انسان با انسان و انسان با شئ عملکرد خوبی از خود نشان داده است.

در مطالعه‌ی انجام شده توسط محمودی[۳۷]، روشی سلسله مراتبی برای شناسایی فعالیت پیشنهاد شده‌است که ترتیب زمانی اطلاعات موجود در ویدئو را در نظر می‌گیرد و می‌تواند ویدئوهای با طول نابرابر را با یکدیگر مقایسه کند. در روش پیشنهادی از مسیر متراکم بهبود یافته بعنوان ویژگی استفاده شده و این ویژگی‌ها بوسیله بردار فیشر ارتقا یافته کد می‌شوند. ساختار ترتیبی ویدئو طی دو مرحله با روش‌های ادغام رتبه و تابع هسته انعطاف پذیر با زمان استخراج می‌شود. در این مطالعه از طبقه بند ماشین بردار پشتیبان برای شناسایی فعالیت‌ها استفاده شده‌است.

[1] Hierarchical Spatio-Temporal Model
[2] Hidden conditional random field

۲-۳ روش‌های شناسایی معنایی فعالیت انسان

در این بخش مرور کلی بر روش‌های مختلف شناسایی فعالیت انسان با استفاده از ویژگی‌های معنایی ارایه می‌کنیم. براساس بهره برداری از معنا، روش‌های شناسایی فعالیت معنایی به ۳ دسته طبقه‌بندی شده‌اند[۴]: روش‌های مبتنی بر توصیف قسمت‌های بدن، روش‌های مبتنی بر توصیف اشیا/صحنه، روش‌های مبتنی بر مشخصه‌های فعالیت.

۲-۳-۱ روش‌های مبتنی بر توصیف قسمت‌های بدن

روش‌های مبتنی بر pose

در بینایی ماشین، تخمین شکل بدن انسان به در کنار هم قرارگرفتن قسمت های مختلف بدن انسان در یک عکس گویند. بسیاری از روش‌های مبتنی بر شکل بدن انسان بوسیله بهم چسباندن مفاصل بدن مدل شده‌اند. موقعیت مکانی مفاصل بدن، شکل بدن انسان را نمایش می‌دهد. Wang و همکارانش[۳۸]، از اتصالات بدن برای ساختار مدل مکانی شکل بدن و همچنین از تغییر شکل زمانی pose برای شناسایی فعالیت‌ها در یک ویدئو استفاده کردند. آنها مکان اتصالات بدن انسان را تخمین زده و بهترین تخمین مفاصل را برای هر فریم بدست آوردند. سپس اتصالات تخمین زده شده را در پنج قسمت از بخش‌های بدن دسته‌بندی کردند و مجموعه‌هایی از دنباله pose متمایز رخ داده از قسمت های بدن در حوزه های مکان و زمان را بدست آوردند. در حالت آزمایش، هیستوگرام‌های مجموعه قسمت‌های تشخیص داده شده بعنوان ورودی طبقه‌بندهای ماشین بردارهای پشتیبان داده می‌شوند.

Chaaraoui و همکارانش [۳۹]، بدن انسان را بوسیله نقاط کانتور نمایش دادند. برای یادگیری pose کلیدی، تمام فریم‌های کلاس فعالیت‌های مشابه را به k خوشه دسته‌بندی کردند که مرکز هر خوشه یک pose کلیدی اولیه را ارایه می‌داد. با تعیین فاصله اقلیدسی میان pose آموزشی و

pose کلیدی اولیه نتایج بدست می‌آیند. منطبق‌ترین pose کلیدی اولیه بعنوان pose کلیدی نهایی قرار می‌گیرد. فرایند یادگیری pose کلیدی برای داده‌های آموزشی هر کلاس فعالیت تکرار می‌شود .

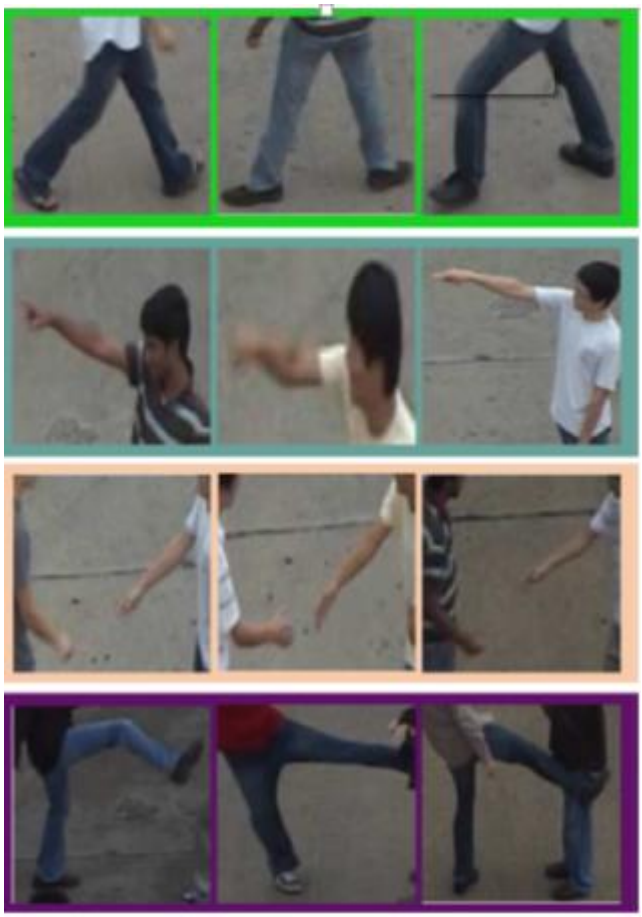
در مرجع [۴۰]، از موقعیت های مکانی اتصالات بدن برای شناسایی فعالیت انسان استفاده شده‌است. در این مقاله تعامل میان بخش‌های یک فرد را تعامل درون فردی و تعامل میان بخش‌های افراد مختلف را بین فردی نامگذاری شده‌است. مکان‌های نسبی اتصال بعنوان ویژگی‌های ارتباط مکانی معنایی برای یادگیری مدل نمایش داده شده‌اند. Vahdatو همکارانش [۴۱]، الگوریتمی برای مدل کردن تعاملات بعنوان دنباله‌هایی از pose های کلیدی پیشنهاد دادند. خوشه بندی میانگین-k برای استخراج poseهای مختلف افراد از ویژگی مسیر حرکت داده شده افراد در ویدئو، استفاده شده‌است. کاندیدهای اولیه pose کلیدی، با معیار نزدیک ترین نمونه‌های آموزشی به مراکز خوشه‌ها استخراج شده‌اند. پارامترهای مدل برای پیداکردن pose کلیدی در یک دنباله آزمون و شناسایی کلاس فعالیت یادگیری شده‌اند.

روش‌های مبتنی بر poselet

مدل‌های رایج تخمین شکل بدن با شناسایی تمام بخش‌های بدن و ساختار ساختمانی شکل بدن انسان سر و کار دارد. انتخاب شکل‌های کلیدی بدن از قوانین خاصی پیروی می‌کند مانند در نظر داشتن نظم زمانی یک فعالیت. بعنوان مثال، در فعالیت "پیاده‌روی" عبور پاها باید میان گام پای چپ و گام پای راست اتفاق بیفتد. از سویی دیگر، روش‌های مبتنی بر poselet قابل اطمینان‌تر هستند چون از زمانی که قسمت‌های مختلف بدن در یک فعالیت قابل مشاهده می‌باشند می‌تواند شناسایی شود.

Paptis و sigal در [۴۲]، مدلی برای شناسایی فعالیت‌های ویدئویی ارائه نمودند. در این روش از فریم‌های کلیدی بعنوان متغیرهای پنهان استفاده شده و بر اساس آن یادگیری در فاز آموزش انجام شده‌است. در این روش اهمیت نظم زمانی فریم‌های کلیدی نیز در نظر گرفته شده‌است.

فریم‌های کلیدی، شامل pose های مهم فرد در فعالیت‌های خاص که در بیشترین حاشیه متمایز کننده یادگیری شده‌اند، هستند. ۲۸ نوع از بخش‌های بارز بدن، برای فعالیت‌های مختلف مثل دست دادن، هل دادن مخاطب، و... معرفی شده‌است [۴۲] (شکل ۲-۱۰).



شکل ۲-۱۰ poselet های فعالیت‌های مختلف[۴۲].

۲-۳-۲ روش‌های مبتنی بر شی و صحنه

با توجه به فعالیت انسان، ممکن است اشیا مرتبط با آن فعالیت وجود داشته باشد. فعالیت-های متفاوت با اشیای متفاوتی ارتباط دارند. دانستن و شناختن اشیا مرتبط به شناسایی فعالیت‌های متناظر با آنها کمک می‌کند. علاوه براین، بعضی از فعالیت‌ها در صحنه‌های خاص انجام می‌شوند، به

طور مثال، شناکردن در آب و رانندگی در جاده. بنابراین استخراج اطلاعات از زمینه فعالیت یا کل صحنه احتمالا برای تحلیل و بررسی و شناسایی فعالیت مفید است.

روش‌های یادگیری با استفاده از توصیفگرهای صحنه برای شناسایی فعالیت استفاده شده-است. Mars zalek و همکارانش [۴۳]، طبقه‌بند مبتنی بر ماشین بردارهای پشتیبان یک فعالیت با صحنه همزمان را بوسیله آموزش چندین توصیفگر صحنه ارتقا دادند. اما، یادگیری براساس توصیفگر نیازمند دانش قبلی راجع به دسته‌بندی‌های صحنه و معمولا به مجموعه داده بستگی دارد. از این رو، این روش نسبت به تغییرات مجموعه داده مقاوم نیست.

بعنوان یک جایگزین برای یادگیری مبتنی بر توصیفگر، zhang و همکارانش[۴۴]، یک روش یادگیری برای شناسایی فعالیت‌ها از صحنه بر طبق توزیع‌های چندجمله‌ایی و دریکله پیشنهاد دادند. آنها در هر فریم نواحی فرد و پس زمینه را جداسازی کردند. سپس، از ناحیه فرد نقاط مهم فضا-زمانی و همینطور رنگ و شکل و ویژگی‌های محلی را از ناحیه پس زمینه شناسایی کردند. این ویژگی‌ها در HOG ، هیستوگرام های رنگ و توصیفگرهای Gist توصیف شده‌اند.

مشکل ویژگی‌های محلی این است که آنها ارتباط زمانی میان ویژگی‌ها را در نظر نمی‌گیرند. liu و همکارانش[۴۵]، با بهره‌مندی از خصوصیت پویایی فعالیت‌ها، روشی مبتنی بر مسیر حرکت با استفاده از پس‌زمینه صحنه پیشنهاد داده‌اند.

با توجه به شناسایی فعالیت مبتنی بر شی، Gupta [۴۶] ، تحقیقی برای شناسایی شی و فعالیت همزمان براساس شکل و حرکت انجام دادند. آنها از ارتباط میان نوع شی، نوع فعالیت و اثر شی برای ارتقا عملکرد شناسایی استفاده کردند.

۲-۳-۳ روش‌های مبتنی بر صفات[1]

صفات یکی از ویژگی‌های فضای معنایی هستند. گاهی اوقات، ممکن است یک صفت به بیشتر از یک فعالیت تخصیص یابد. بعنوان مثال، "سواری‌کردن" صفتی است که می‌تواند هم به "دوچرخه‌سواری" و هم "اسب‌سواری" مربوط باشد.

از این رو، این موضوع برای انتخاب صفات متمایز کننده مهم است که نتیجه دقیق‌تری در شناسایی فعالیت به دنبال دارد[۴].

Lin و همکارانش [۶]، صفات فعالیت را برای شناسایی فعالیت انسان از ویدئو معرفی‌کردند. آنها صفات را بعنوان متغیرهای پنهانی مدل‌کردند و مسئله طبقه‌بندی را با استفاده از ماشین بردارهای پشتیبان خطی پنهان انجام دادند. در این روش متمایزکننده‌ترین صفات برای هر فعالیت انتخاب شده‌است. وجود و یا عدم وجود هر صفت در یک فعالیت با یک بردار باینری بیان شده که کلاس فعالیت را مشخص می‌کند.

۲-۴ جمع‌بندی

روش‌های معنایی و غیرمعنایی هرکدام مزایا و معایبی دارند. بطور خاص، روش مبتنی بر بخش‌های بدن بسیار به تفاوت شکل بدن در فعالیت‌ها وابسته است. اگر شکل بدن در فعالیت به میزان قابل توجه متفاوت از سایر فعالیت‌ها نباشد، این روش‌ها با شکست روبرو می‌شود. بعنوان مثال، دویدن در مقابل راه‌رفتن. روش‌های شناسایی مبتنی بر شی و صحنه معمولاً به مجموعه داده بستگی دارد. از این رو، این روش‌ها نسبت به تغییرات مجموعه داده‌های یک کلاس فعالیت مقاوم نیستند.

از سویی دیگر، روش‌های غیرمعنایی از ویژگی‌های سطح پایین و یا متوسط استفاده می‌کنند و از این رو به شکل بدن و صفات مشترک در فعالیت‌های متفاوت کمتر حساس هستند. این روش‌ها برای فعالیت‌های ساده، مناسب و ایده‌آل هستند، و در وضعیت های پیچیده شکست می‌خورند.

[1] attributes

در جدول۲-۱، مقایسه ای بین دقت شناسایی روش‌های معنایی و روش‌های غیرمعنایی بر روی پایگاه داده‌های گوناگون انجام شده است. عملکرد روش‌ها بر اساس متوسط دقت AP[1] قضاوت شده است.

جدول ۲-۱عملکرد روش‌های مختلف بر روی پایگاه داده‌های گوناگون.

روش‌های شناسایی	پایگاه داده	معنایی	غیرمعنایی	متوسط دقت
Ryoo [35]	UT-Interaction		✓	۷۰/۸
Xu [36]	UT-Interaction		✓	۹۴/۱۷
Vahdat [41]	UT-Interaction	✓		۹۳/۳
Meng [40]	UT-Interaction	✓		۸۷/۷
Raptis [42]	UT-Interaction	✓		۹۳/۳
Liu[6]	Olampic sports	✓		۷۴/۳۸
[37]محمودی	Olampic sports		✓	۸۵/۰۷
Wang[24]	Olampic sports		✓	۷۷/۲
Yi, Y [25]	Olampic sports		✓	۸۵/۴۶
Cho[29]	Olampic sports		✓	۸۱/۳
Wang[38]	UCF Sport	✓		۹۰
Rapits[34]	UCF Sport		✓	۷۹/۴
Cho [29]	UCF Sport		✓	۸۹/۷

[1] Average precision

فصل ۳: مباحث نظری

۳-۱ مقدمه

در این فصل ابتدا مباحث نظری ویژگی‌های استفاده شده در این پایان نامه به منظور شناسایی فعالیت را معرفی کرده و سپس الگوریتم خوشه بندی بکار گرفته شده و طبقه بند مورد استفاده را شرح می دهیم.

۳-۲ استخراج ویژگی

۳-۲-۱ هیستوگرام گرادیان جهت‌دار

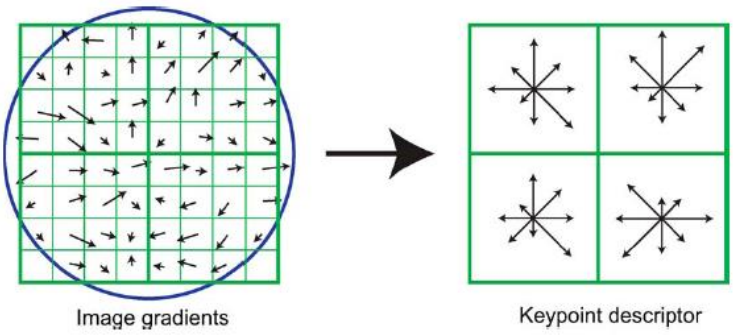
یکی از پرکاربردترین توصیفگرهای محلی در پردازش تصویر، هیستوگرام گرادیان جهت‌دار است. این توصیفگر پنجره‌ای از تصویر را بوسیله توزیع گرادیان‌های شدت روشنایی یا جهت لبه‌ها توصیف می‌کند. تصویر ورودی به ناحیه‌های کوچک بهم چسبیده که سلول نامیده می‌شود، تقسیم شده، برای پیکسل‌های درون هر سلول ، اندازه و زاویه گرادیان محاسبه می‌شود. زاویه ۰تا ۳۶۰ درجه را معمولا به ۸ انبارک تقسیم نموده، تعداد اندازه گرادیان برای بردارهایی که در هر انبارک قرار می-گیرند با هم جمع می‌شوند. در نتیجه هیستوگرامی به تعداد انبارک‌های گرادیان برای هر سلول ساخته شده و در نهایت توصیفگر کل تصویر از کنار هم قرار گرفتن هیستوگرام سلول‌ها تشکیل می‌شود. گرادیان تصویر در یک جهت، بیانگر تغییرات شدت روشنایی در آن جهت است. اندازه و جهت گرادیان از رابطه‌های زیر بدست می آیند[۴۷].

$$\boldsymbol{M(x,y) = \sqrt{[\,I(x+1,y) - I(x-1,y)]^2 + [\,I(x,y+1) - I(x,y-1)]^2}} \qquad (۱-۳)$$

$$\theta(x,y) = tan^{-1}\frac{I(x,y+1)-I(x,y-1)}{I(x+1,y)-I(x-1,y)} \qquad (۲-۳)$$

که در آن I(x,y) شدت روشنایی پیکسلی در مختصات (x,y) است.

تعداد سلول ها نقش مهمی در نتایج آزمایشات دارند. در شکل (۳-۱)، تصویر را به ۴زیرناحیه تقسیم کردند اما در تحقیق [۴۷]، نشان دادند که با تقسیم کردن تصویر به آرایه های ۴*۴ به نتیجه بهتری در توصیف ناحیه می‌توان رسید. توصیفگر نهایی تصویر حول هر نقطه مهم، برداری به ابعاد ۴*۴*۸ =۱۲۸ است.



شکل ۳-۱دسته بندی جهت های گرادیان به ۸ قسمت[۴۷].

۳-۲-۲ هیستوگرام جریان نوری

جریان نوری یکی از مفاهیم پایه در پردازش ویدئو است. این ویژگی بصورت برداری در مختصات دو بعدی که نشان دهنده جهت حرکت هر پیکسل در فریم‌های متوالی ویدئو است، تعریف می‌شود. اگر اندازه بردار جریان نوری در قسمتی از تصویر از یک آستانه‌ای بیشتر بود آن بخش دارای حرکت در نظر گرفته شده، در غیر اینصورت ساکن در نظر گرفته می‌شود. جریان نوری بین دو فریم متوالی محاسبه می‌شود و فرض اصلی در این الگوریتم ثابت بودن شدت روشنایی پیکسل‌های متناظر بین دو فریم است. فرض کنیم پیکسلی به مختصات (x,y,t) به اندازه‌ی (dx,dy,dt) حرکت کرده‌است. در این صورت اگر I شدت روشنایی این پیکسل باشد داریم:

$$\boldsymbol{I(x,y,t) = I(x+dx, y+dy, t+dt)} \qquad \textbf{(۳-۳)}$$

از طرفی با استفاده از بسط تیلور می‌توان عبارت سمت راست رابطه‌ی (۳-۳) را به صورت رابطه‌ی (۳-۴) نوشت.

$$\boldsymbol{I(x+dx, y+dy, t+dt) \cong I(x,y,t) + \frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt} \qquad \textbf{(۴-۳)}$$

از مقایسه‌ی روابط (۳-۳) و (۳-۴) داریم:

$$\boldsymbol{\frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt = 0} \qquad \textbf{(۵-۳)}$$

با تقسیم طرفین رابطه‌ی (۳-۵) به dt داریم:

$$\boldsymbol{\frac{\partial I}{\partial x}v_x + \frac{\partial I}{\partial y}v_y + \frac{\partial I}{\partial t} = 0} \qquad \textbf{(۶-۳)}$$

که V_x و V_y مؤلفه‌های سرعت یا همان جریان نوری در جهت‌های x وy هستند. در معادله‌ی (۳-۶) دو مجهول وجود دارد و به تنهایی قابل حل نیست. برای حل آن می‌توان از روش لوکاس-کانادی[1] [۴۸]، استفاده نمود. در این روش فرض می‌شود در یک تکه‌ی ۳*۳ اطراف هر نقطه از تصویر، جریان نوری با آن نقطه مشابه است. بنابراین در پنجره ۳*۳، ۹ نقطه داریم که همگی V_x و V_y یکسانی دارند. لازم است دستگاه معادلات به فرم (۳-۷) را حل نمود.

$$\boldsymbol{S\begin{bmatrix} v_x \\ v_t \end{bmatrix} + B = 0} \qquad \textbf{(۷-۳)}$$

که در آن:

$$\boldsymbol{S = \begin{bmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \end{bmatrix}, B = \begin{bmatrix} \frac{\partial I}{\partial t} \end{bmatrix}} \qquad \textbf{(۸-۳)}$$

پاسخ معادله‌ی (۳-۷) به صورت (۳-۹) خواهد بود.

[1] Lucas-Kanade

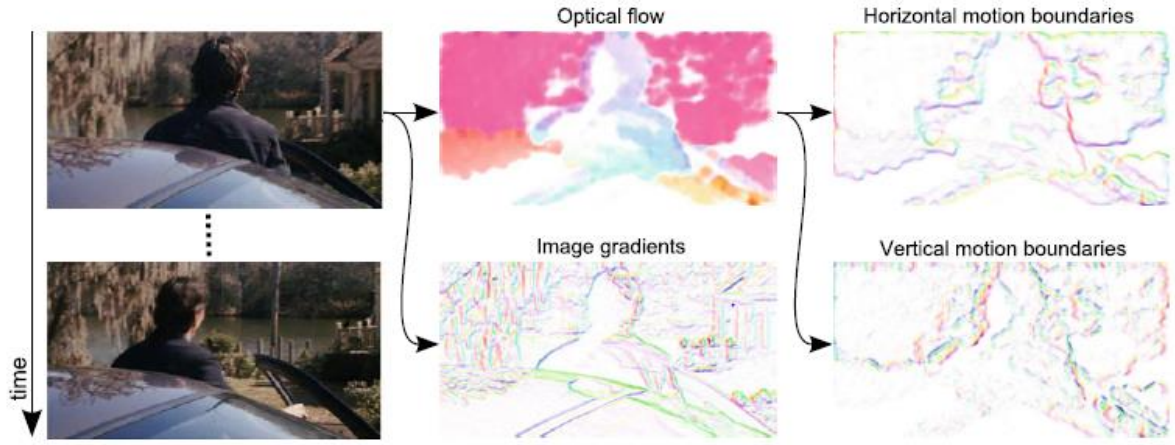
$$\begin{bmatrix} v_x \\ v_t \end{bmatrix} = (S^T S)^{-1} S^T B \qquad (۹-۳)$$

پس از بدست آوردن بردار جریان نوری هر نقطه از فریم‌های ویدئو، جهت این بردارها به ۸ جهت کوانتیزه می‌شود و هیستوگرام جهت‌ها برای هر تکه از تصویر محاسبه می‌شود. انبارک نهم که نشان دهنده نقاط با اندازه جریان نوری تقریبا صفر هستند نیز به هیستوگرام اضافه می‌شود [۴۹،۳۷].

۳-۲-۳ هیستوگرام مرز حرکت

جریان نوری اندازه حرکت میان دو فریم را نشان می دهد، که این حرکت می‌تواند ناشی از حرکت شی پیش‌زمینه و یا حرکت ناشی از دوربین و حتی تلفیق این دو حرکت باشد. جریان نوری نسبت به حرکت دوربین مقاوم نیست و اگر حرکت دوربین بعنوان حرکت فعالیت در نظر گرفته شود، ممکن است یک فعالیت اشتباه طبقه‌بندی شود. Dalal در[۲۳]، توصیفگر هیستوگرام مرز حرکت را بوسیله محاسبه مشتق‌های مؤلفه‌های عمودی و افقی جریان نوری بصورت جداگانه معرفی کرد. این توصیفگر می تواند حرکات ناشی از دوربین را حذف کند. در این روش دو مؤلفه جریان نوری V_x و V_y بعنوان تصاویر مستقل در نظر گرفته شده و گرادیان‌های محلی آن‌ها بصورت جداگانه محاسبه و مطابق روش مطرح شده در قسمت هیستوگرام گرادیان‌های جهت دار، برای هریک از تصاویر محاسبه می‌شود. هیستوگرام مرز حرکت در راستای سطرها و ستون‌ها به ترتیب MBH_x و MBH_y اشاره می‌شوند که توصیفگر مرز حرکت می‌باشند.

در شکل (۳-۲) دوربین از راست به چپ در حرکت است و فرد در حال دور شدن از دوربین است. شار نوری (بالا-وسط) نشان دهنده حرکات ثابت پس‌زمینه ناشی از حرکت دوربین است. مرز حرکت (راست)، حرکت مرتبط میان فرد و پس زمینه را نشان می‌دهد.

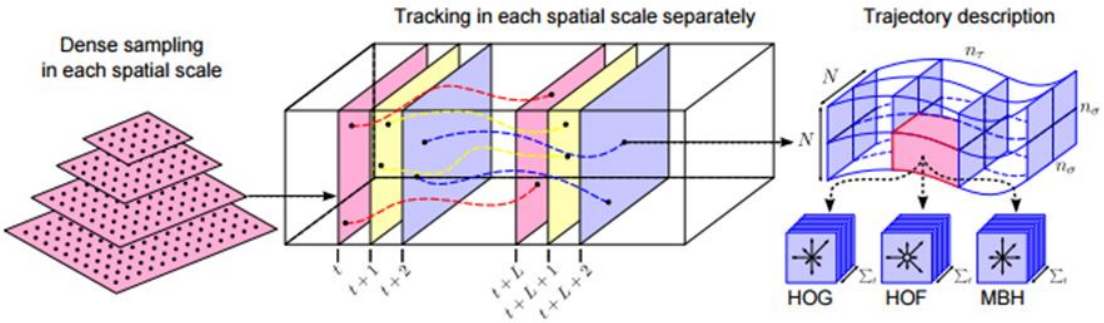


شکل ۳-۲ نمایش اطلاعات ذخیره شده بوسیله توصیفگرهای HOG,HOF,MBHبرای رشته ویدئویی نمونه[۲۴].

۳-۲-۴ مسیر حرکت متراکم

ویژگی مسیر حرکت متراکم برای اولین بار در مرجع [۲۳]، معرفی شد. در این تحقیق مسیر حرکت متراکم همراه با ویژگی‌های HOG,HOF,MBH برای چند مقیاس مکانی استخراج شد. همانطور که در شکل (۳-۳) نشان داده شده، ابتدا نقاط ویژگی روی یک فضای شبکه‌ای با فاصله w پیکسل نمونه برداری شده اند و در هر مقیاس جداگانه ردیابی می‌شوند. بصورت تجربی مقدار w را در این تحقیق ۵ پیکسل در نظر گرفته‌اند و ۸ مقیاس فضایی-مکانی برای استخراج مسیر حرکت متراکم بکار رفته است.

به ازای هر کدام از مقیاس‌های مکانی، هر نقطه $\boldsymbol{P_t}$ به مختصات (x,y) در فریم t بوسیله فیلتر میانه M و جریان شار نوری محاسبه می‌شود و هر پیکسل مطابق رابطه (۳-۱۰) تا فریم بعدی ردیابی می‌شود تا مختصات جدید نقطه در فریم بعدی بوسیله رابطه (۳-۱۱) بدست آید.



شکل ۳-۳ مراحل استخراج ویژگی مسیر متراکم [۲۳].

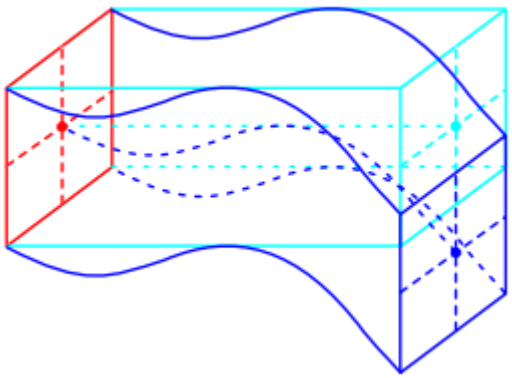
$$\boldsymbol{\omega(x_t, y_t) = (u_t, v_t)} \qquad \textbf{(۱۰-۳)}$$

$$\boldsymbol{P_{t+1} = (x_{t+1}, y_{t+1}) = (x_t, y_t) + (M * w_t)|_{(x_t, y_t)}} \qquad \textbf{(۱۱-۳)}$$

در این مرحله نقطه نمونه برداری شده در طول L فریم ردیابی شده و با کنار هم قرار دادن نقاط ردیابی شده مسیرحرکت آن نقطه بصورت رابطه (۳-۱۲) تشکیل می‌شود. که در تحقیق[۲۳]، بصورت تجربی، مقدار **L=15** بکار گرفته شده است.

$$\boldsymbol{T = (P_t, P_{t+1}, P_{t+2}, \dots)} \qquad \textbf{(۱۲-۳)}$$

در مرحله بعدی به منظور استفاده از اطلاعات حرکت و ترکیب آن‌ها با توصیفگر مسیر متراکم، در اطراف هر یک از نقاط ردیابی شده مسیر حرکت که شامل ۱۵ نقطه است، حجم زمانی-فضایی به ابعاد N*N و L فریم مانند شکل (۳-۴) در نظر گرفته شده است. در این مقاله N=32 در نظر گرفته شده است.



شکل ۳-۴حجم خمیده اطراف مسیر حرکت[۲۴].

در نهایت برای محاسبه ویژگی‌های محلیHOG وHOF و MBH، حجم بدست آمده در اطراف مسیر حرکت را به پنجره های کوچکتری با ابعاد $n_\tau * n_\sigma * n_\sigma$، همانطور که در تصویر سمت راست شکل (۳-۳) نشان داده شده است، تقسیم کرده و ویژگی‌های محلی ذکر شده را در هر یک از این پنجره‌ها محاسبه می‌کنند. در این مقاله n_σ =۲ و n_τ =۳ پیشنهاد شده است.

در نهایت در هریک از این پنجره‌های کوچک یک بردار ۴۲۶بعدی استخراج می‌شود که ۳۰ مؤلفه‌ی آن مربوط به مسیر حرکت، ۹۶ مؤلفه مربوط به HOG، ۱۰۸ مؤلفه مربوط به HOF، ۹۶ مؤلفه مربوط به MBH_x و ۹۶ مؤلفه‌ی دیگر مربوط به MBH_y است.

۳-۳ کد کردن ویژگی‌ها

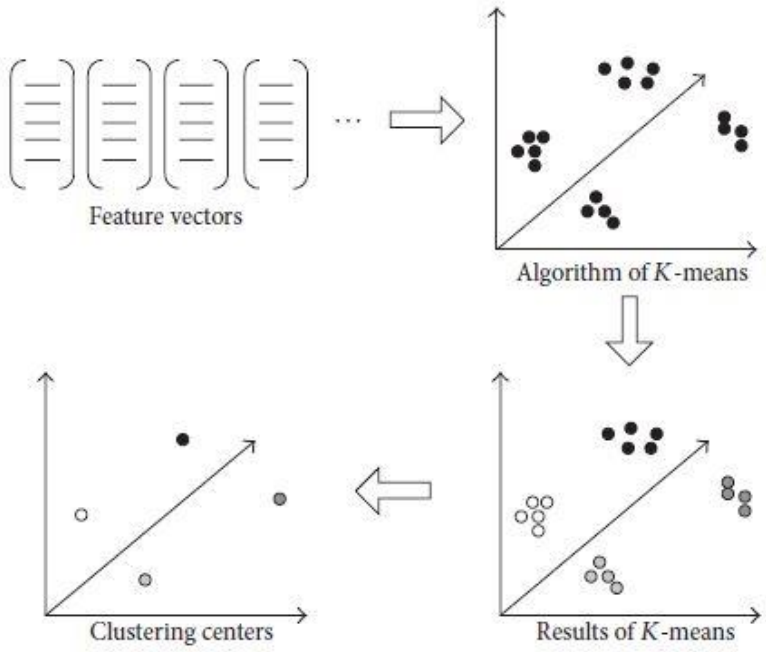
به منظور شناسایی فعالیت ما نیازمند روشی برای یادگیری ویژگی‌های استخراج شده هستیم، یکی از روش‌های پرکاربرد برای کد کردن ویژگی‌ها که در این پایان نامه بکار گرفته شده، استفاده از روش کیف کلمات است. برای این کار ابتدا یک لغت نامه تشکیل داده، سپس هر یک از ویژگی‌های استخراج شده به یکی از کلمات کد نسبت داده می‌شود.

۳-۳-۱ کیف کلمات

در این روش بردارهای ویژگی محلی هر بخش از مدل مکان-زمان ویدئو استخراج شده و با تشکیل کیف کلمات، هیستوگرام وقوع کلمات بعنوان نماینده هر ویدئو بدست می‌آید. تشکیل کیف کلمات شامل مراحل زیر است:

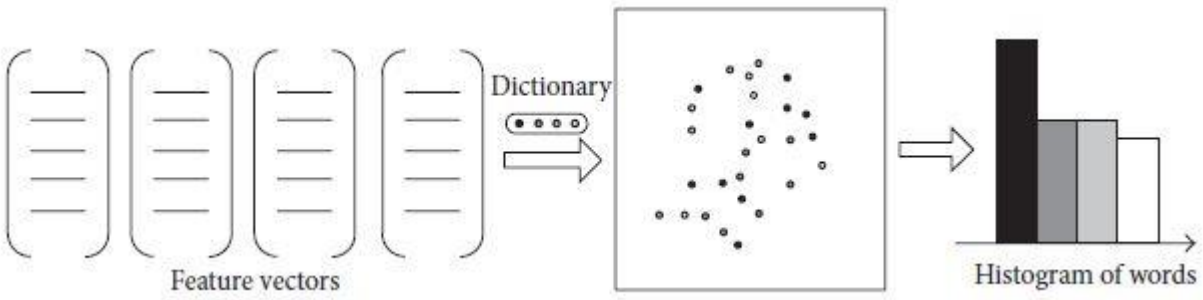
- استخراج بردارهای ویژگی تمام ویدئوهای آموزشی
- تشکیل کتاب کد ویژگی‌های استخراج شده

در این گام بردارهای ویژگی استخراج شده را با یک الگوریتم خوشه‌بندی مانند k- میانگین خوشه-بندی کرده و مراکز خوشه‌ها بعنوان کلمات بصری[1] در نظر گرفته می‌شود. هیستوگرام کلمات، کتاب کد را تشکیل می‌دهند. روند تشکیل کتاب کد در شکل (۵-۳) نشان داده شده‌است.



شکل ۳-۵ مراحل تشکیل کتاب کد

- انتساب ویژگی‌های محلی هر ویدئو به نزدیک‌ترین کلمات بصری کتاب کد (با استفاده از فاصله اقلیدسی)
- تشکیل هیستوگرام وقوع کلمات کد بعنوان توصیفگر برای هر ویدئو



شکل ۳-۶ روند تشکیل کیف کلمات برای هر ویدیو

[1] Visual word

۳-۴ خوشه بندی

۳-۴-۱ خوشه‌بندی انتشار وابستگی

خوشه‌بندی انتشار وابستگی (AP[1] Clustering) یکی از الگوریتم‌های خوشه‌بندی بر اساس مفهوم انتقال پیام بین داده‌ها می‌باشد. بر خلاف بسیاری از الگوریتم‌های خوشه‌بندی از جمله K-means، این الگوریتم نیازی به مشخص‌شدن تعداد خوشه‌ها قبل از اجرای الگوریتم ندارد. همچنین در K-means بدلیل اینکه مقدار اولیه مراکز خوشه‌ها را بصورت تصادفی تعیین می‌کنند برای دستیابی به نتیجه بهتر در خوشه‌بندی، الگوریتم K-means را با مقادیر تصادفی اولیه متفاوتی اجرا می‌کنند که در خوشه‌بندی AP به چنین مکانیزمی نیاز نیست.

فرض کنیم $x_1, \dots, x_n$ نقطه داده‌های ورودی باشند. S تابع شباهت بین دو نقطه داده را می‌سنجد بطوریکه داشته باشیم $S(x_i, x_j) > S(x_i, x_k)$ اگر x_i به x_j شباهت بیشتری نسبت به x_k داشته باشد. تابع شباهت S را بصورت زیر انتخاب می‌کنیم که شرط بالا رعایت شود.

$$S(i,k) = -\|x_i - x_k\|^2 \qquad (۱۳-۳)$$

مفهوم انتقال پیام با دو ماتریس وظیفه[2] R و ماتریس دسترسی[3] A بشکل زیر انجام می‌شود.

- ماتریس وظیفه R: مقدار R(i,k) نشان دهنده این است که x_k به چه میزان مرکز مناسبی برای x_i می‌باشد. مقدار بزرگتر نشان دهنده بهتر بودن است.
- ماتریس دسترسی A: مقدار A(i,k) نشان دهنده این است که برای x_i چقدر مناسب است که x_k به عنوان مرکز انتخاب شود. مقدار بزرگتر نشان دهنده بهتر بودن است.

الگوریتم خوشه‌بندی AP بصورت زیر است:

دو ماتریس A و R مقدار دهی اولیه صفر می‌شوند.

۱. ابتدا ماتریس وظیفه بروزرسانی می‌شود.

[1] Affinity propagation
[2] Responsibility
[3] Availability

$$R(i,k) \leftarrow S(i,k) - \max_{k \neq k'}\{A(i,k') + S(i,k')\}$$

۲. سپس ماتریس دسترسی بروزرسانی می‌شود.

$$A(i,k) \leftarrow \min\left(0, R(k,k) + \sum_{i' \notin \{i,k\}} \max\left(0, R(i',k)\right)\right) \quad for\ i \neq k$$

$$A(k,k) \leftarrow \sum_{i' \neq k} \max(0, R(i',k))$$

مراحل ۱ و ۲ تا زمانی اجرا می‌شود که یکی از دو شرط زیر یا هردو شرط اتفاق بیفتد.

- تعداد تکرار از حد مشخصی بیشتر شود.
- خوشه‌های پیدا شده در چند تکرار متوالی تغییر نکنند.

مراکز خوشه‌ها بصورت زیر از دو ماتریس A و R انتخاب می‌شود.

$$i\ is\ center\ if\ \left(R(i,i) + A(i,i)\right) > 0$$

۳-۵ طبقه‌بندی

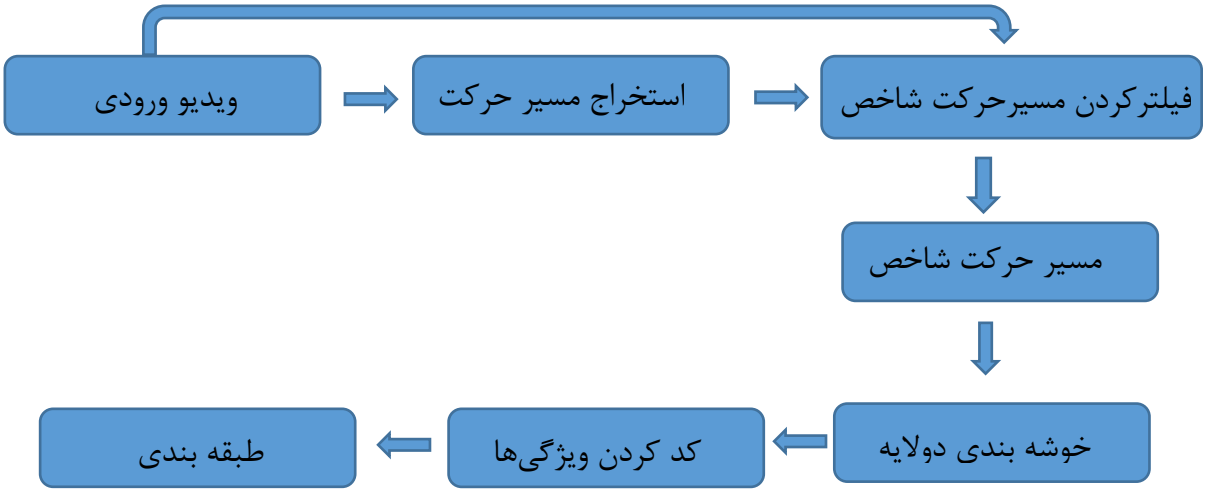
بعد از استخراج ویژگی‌های موردنظر و تشکیل کیف کلمات، هر ویدئو آموزشی را با یک بردار ویژگی d بعدی نشان می‌دهیم. که d تعداد مراکز خوشه بندی استفاده شده در الگوریتم کیف کلمات است. برای طبقه‌بندی ویدئوهای آزمون از طبقه بند K نزدیک‌ترین همسایه ها استفاده می‌کنیم. در این روش با داشتن بردارهای ویژگی مجموعه آموزش و نمونه آزمون، K نزدیک‌ترین داده‌ها را در بین مجموعه آموزش پیدا می‌کند.

فرض کنیم $D_n = \{x_1, x_2, \dots, x_n\}$، n نمونه در مجموعه آموزش باشند که کلاس آن‌ها مشخص است. هر نمونه با یک بردار ویژگی در فضای d بعدی مشخص می‌شود. اگر نمونه آزمون x باشد که این نمونه هم در فضای d بعدی است برای عمل طبقه‌بندی کافی است K نزدیک‌ترین همسایه ها را از نمونه‌های آموزش مشخص کنیم و کلاس آن را به نمونه آزمون با تعداد بیشتر نسبت دهیم.

فصل ۴: روش پیشنهادی

۴-۱ مقدمه

در این فصل ابتدا روش پیشنهادی را شرح داده سپس پایگاه داده‌های استفاده شده را معرفی می‌کنیم. شکل (۴-۱) روند مراحل روش پیشنهادی در شناسایی ویدیو را نشان می‌دهد. در ادامه هر یک از مراحل را شرح می دهیم.



شکل ۴-۱ روندنمای روش پیشنهادی

۴-۲ شرح مراحل روش پیشنهادی

۴-۲-۱ استخراج مسیر حرکت

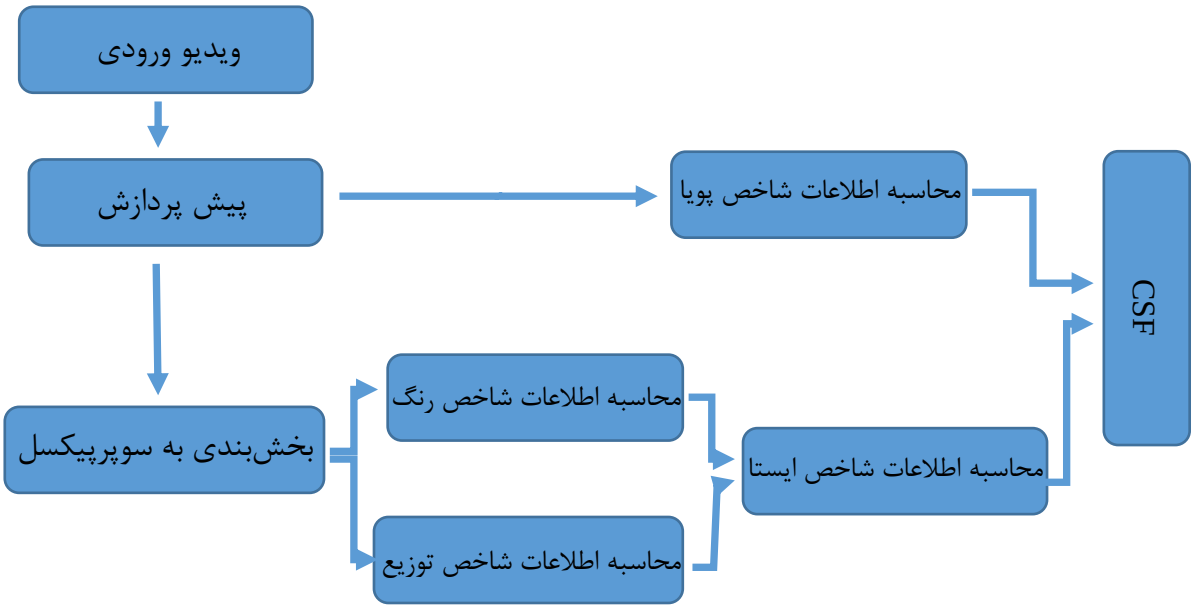
در ابتدا ویژگی مسیر حرکت متراکم از تمام ویدئوهای آموزشی و آزمون استخراج می‌شود. همانطور که در بخش ۳-۲-۴ توضیح داده شد، با در نظر گرفتن حجم خمیده‌ای اطراف مسیر حرکت ویژگی-های HOG وHOF و MBH استخراج شده و بردار ویژگی بدست می‌آید. هر بردار ویژگی ۴۲۶ مؤلفه دارد که ۳۰ مؤلفه‌ی آن مربوط به مسیر حرکت، ۹۶ مؤلفه مربوط به HOG، ۱۰۸مؤلفه مربوط به HOF، ۹۶ مؤلفه مربوط به MBH_x و ۹۶ مؤلفه مربوط به MBH_y است. معمولاً ویژگی مسیر حرکت بوسیله روش نمونه‌برداری متراکم بدست می‌آید. این روش نمونه برداری در صحنه‌های پیچیده، مسیرهای حرکت زائد و نامرتبط را استخراج می‌کند. برای رفع این مشکل مسیرهای حرکت زائد را بوسیله فیلتر حذف می‌کنیم. در بخش بعدی نحوه محاسبه این فیلتر شرح داده می‌شود.

در این پایان نامه گام نمونه برداری نقاط ویژگی را w=۱۰ پیکسل در نظر گرفته‌ایم.

۴-۲-۲فیلترکردن مسیرحرکت شاخص

در حالت کلی، تمام ویژگی‌ها برای طبقه‌بندی متمایزکننده نیستند. ویژگی‌های پس‌زمینه و نواحی ثابت پیش‌زمینه نقش زیادی در موفقیت شناسایی فعالیت ندارند، در نتیجه شناسایی نواحی دارای حرکت و مدل کردن اطلاعات مکانی آنها اهمیت زیادی دارد. به همین منظور برای هرس کردن[۱] ویژگی‌های زائد از فیلتر شاخص سنجی مبتنی بر کنتراست[۲] (CSF) استفاده می‌کنیم.

برای محاسبه CSF از اطلاعات شاخص ایستا[۳] مانند رنگ و توزیع[۴] و اطلاعات شاخص پویا[۵] استفاده شده‌است. نحوه محاسبه CSF در روندنمای (۴-۲) نشان داده شده و در ادامه هر یک از مراحل روندنما را شرح می‌دهیم.



شکل ۴-۲ نحوه محاسبه CSF

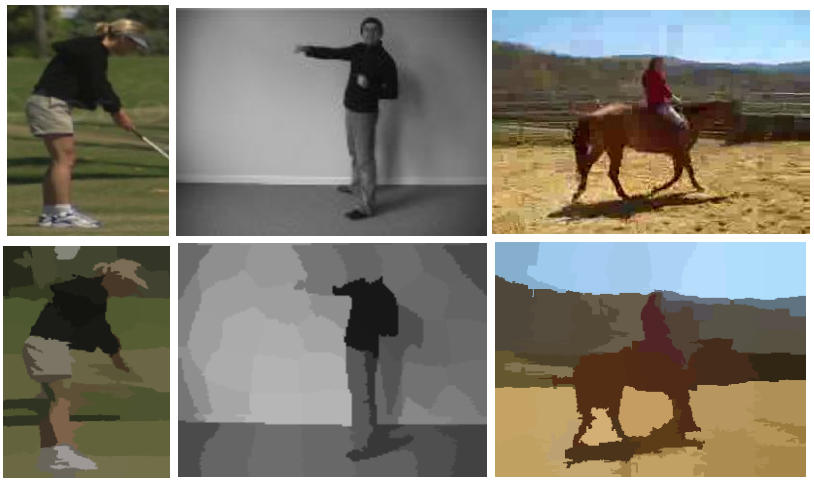
[1] prune
[2] Contrast based saliency filter
[3] Static saliency
[4] Distribution
[5] Dynamic saliency

- **پیش پردازش**

فریم ها در رزولوشن پایین، محتوای کلی تصویر و در رزولوشن بالا جزئیات محلی را توصیف می‌کنند، در نتیجه ابتدا ۸ لایه هرم با ضریب $\frac{۱}{\sqrt{۲}}$ برای هر فریم تشکیل می‌دهیم. ویژگی های استخراج شده در رزولوشن‌های متفاوت به دلیل ترکیب اطلاعات محلی و سراسری، متمایز کننده‌تر هستند.

- **بخش‌بندی فریم‌ها به سوپرپیکسل**

به منظور کاهش پیچیدگی محاسباتی و کاهش زمان در محاسبه نگاشت شاخص ایستا[۱] از خوشه‌بندی SLIC[۲] پیکسل‌ها برای تولید سوپرپیکسل‌ها استفاده شده‌است [۵۰]. ایده اصلی این الگوریتم خوشه-بندی پیکسل‌ها براساس شباهت رنگ و نزدیکی فاصله مکانی است. ورودی این الگوریتم تعداد سوپرپیکسل‌ها است. تعداد سوپرپیکسل‌ها وابسته به ابعاد فریم‌ها در نظر گرفته شده‌است. برای فریم-های مقیاس ۷ و ۸ تعداد سوپرپیکسل‌ها را ۱۰۰، مقیاس ۶ و ۵ به ترتیب ۱۵۰ و ۳۰۰ در نظر می-گیریم. برای سایر مقیاس‌ها تعداد سوپرپیکسل‌های مقیاس iام از فرمول $۲۵۰*۱.۲۵^{(۸-i)}$ بدست می‌آید. نمونه‌ای از نتایج پیاده سازی الگوریتم SLIC در شکل (۴-۳) نشان داده شده‌است.



(الف) (ب) (ج)

شکل ۴-۳ نمایش سوپرپیکسل‌های فعالیت‌های الف)بازی‌گلف، ب)بوکس و ج)اسب‌سواری

[1] static saliency map
[۲] Simple Linear Iterative Clustering

برای محاسبه میزان اطلاعات شاخص هر پیکسل ابتدا اطلاعات شاخص ایستا و پویای هر فریم را بدست‌آورده سپس با ترکیب آنها اطلاعات شاخص فریم‌ها را بدست می‌آوریم.

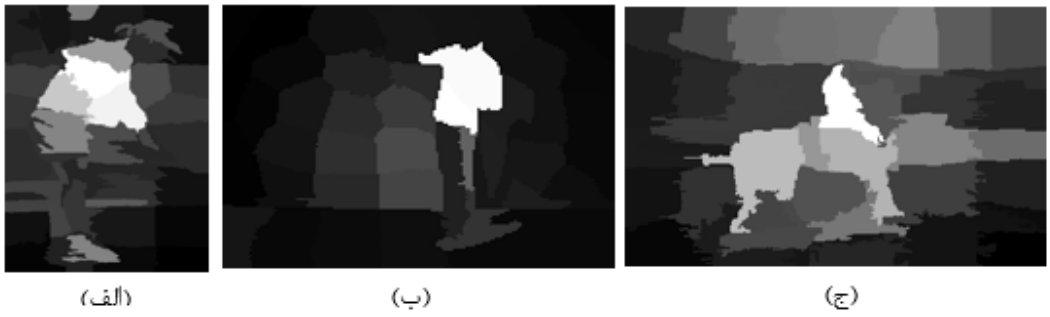
- **محاسبه اطلاعات شاخص مبتنی بر کنتراست رنگ[1]**

$$U_i = \sum_{j=1}^{K} \|c_i - c_j\|^2 w_{ij}{}^{(p)} \qquad (۱-۴)$$

سوپرپیکسل i با مکان P_i و مقدار رنگ c_i را در نظر بگیرید، اطلاعات شاخص مبتنی بر کنتراست رنگ سوپرپیکسل i بصورت رابطه (۴-۱) تعریف می‌شود.

$$w_{ij}{}^{(p)} = \frac{1}{Z_i} exp\left(-\frac{1}{2\sigma_p^2}\|P_i - P_j\|^2\right) \qquad (۲-۴)$$

که K تعداد سوپرپیکسل ها و $w_{ij}{}^{(p)}$ پارامتر تنظیم کنتراست رنگ است که بصورت رابطه (۴-۲) تعریف می‌شود. σ_p محدوده شاخص رنگ را کنترل می‌کند، که در این تحقیق $\sigma_p =$ 0.25نظرگرفته شده‌است. Z_i ضریب نرمال ساز است طوری که مجموع $w_{ij}{}^{(p)}$ برای تمام سوپرپیکسل‌ها ۱ شود. مثال‌هایی از اطلاعات شاخص مبتنی بر کنتراست رنگ در شکل (۴-۴) نشان داده شده‌است.



شکل ۴-۴ نمایش اطلاعات شاخص مبتنی بر کنتراست رنگ بر روی فعالیت‌های الف)گلف، ب) بوکس و ج)اسب سواری

[1] Color contrast saliency

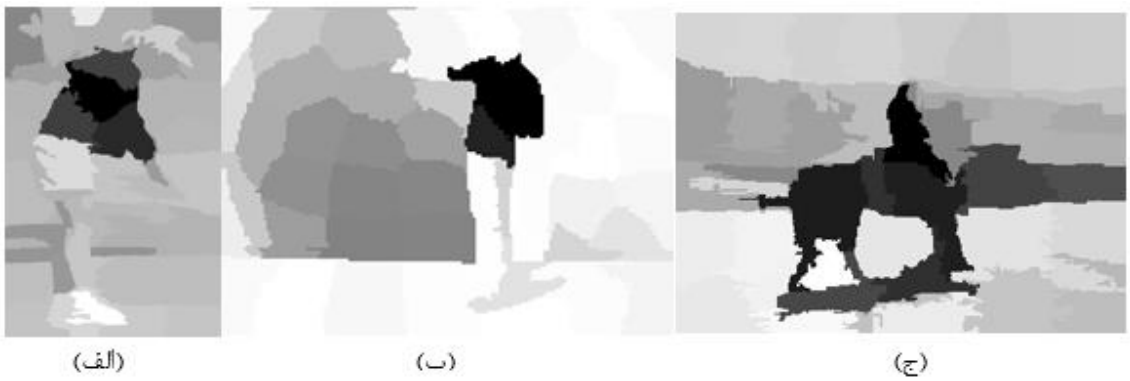
- **محاسبه اطلاعات شاخص مبتنی بر کنتراست توزیع**[1]

اطلاعات شاخص کنتراست توزیع یک سوپرپیکسل با استفاده از رابطه (۴-۳) بدست می‌آید:

$$D_i = \sum_{j=1}^{K} \left\| P_j - \varphi_i \right\|^2 w_{ij}{}^{(c)} \qquad (۳-۴)$$

$$w_{ij}{}^{(c)} = \frac{1}{Z_i} exp\left(-\frac{1}{2\sigma_c^2} \left\| c_i - c_j \right\|^2\right) \qquad (۴-۴)$$

$$\varphi_i = \sum_{j=1}^{K} w_{ij}^{(c)} P_j \qquad (۵-۴)$$



شکل ۴-۵ نمایش اطلاعات شاخص مبتنی بر کنتراست توزیع فعالیت‌های الف) بازی‌گلف، ب) بوکس و ج) اسب‌سواری

$w_{ij}{}^{(c)}$ تعریف شده در رابطه (۴-۴) پارامتر تنظیم کنتراست است. برای بکارگیری اطلاعات رنگ برای تنظیم کنتراست توزیع، $\sigma_c = ۲۰$ تنظیم می شود و φ_i متوسط وزن دار مکان c_i است. نمونه ای از پیاده‌سازی اطلاعات شاخص مبتنی بر کنتراست توزیع در شکل (۴-۵) نشان داده شده‌است.

سپس با ترکیب اطلاعات شاخص مبتنی بر کنتراست رنگ و اطلاعات شاخص مبتنی بر کنتراست توزیع، اطلاعات شاخص ایستا هر سوپرپیکسل از رابطه (۴-۶) بدست می‌آید:

$$S_i = \overline{U}_i . exp(-k . \overline{D}_i) \qquad (۶-۴)$$

[1] Distribution_contrast saliency

که $\overline{U_i}$ و $\overline{D_i}$ به ترتیب نرمالیزه شدهU_i و D_i هستند. در شرایط واقعی تاثیرD_i بیشتر از U_i است. از این رو، فرم نمایی برای افزایش تاثیر D بکار گرفته شده‌است. k برای کنترل افزایش تاثیر استفاده شده-است، در این تحقیق مطابق مرجع [۵۰]، k=۶ تنظیم شده است.

در نهایت اطلاعات شاخص ایستا هر پیکسل بوسیله درون یابی خطی به دست می‌آید، بعنوان مثال اطلاعات شاخص ایستا پیکسل x_{fi} در فریم f بصورت رابطه (۴-۷) تعریف شده است.

$$C_s(X_{fi}) = \sum_{j=1}^{K} w_{ij} S_j \qquad (۴-۷)$$

$$w_{ij} = \frac{1}{Z_i} exp\left(-\frac{1}{2}\left(\alpha\|c_i - c_j\|^2 + \beta\|P_i - P_j\|^2\right)\right) \qquad (۴-۸)$$

که c_i و P_i به ترتیب مقدار رنگ و مکان پیکسل هستند. $\alpha = \beta = \frac{۱}{۳۰}$ مشابه مرجع [۵۰]، در نظر گرفته شده‌است. نتایج پیاده سازی اطلاعات شاخص ایستا در شکل (۴-۶) نمایش داده شده-است.

شکل ۴-۶ نمایش اطلاعات شاخص ایستا فعالیت‌های الف)گلف، ب)بوکس و ج)اسب‌سواری

- **محاسبه اطلاعات شاخص پویا**

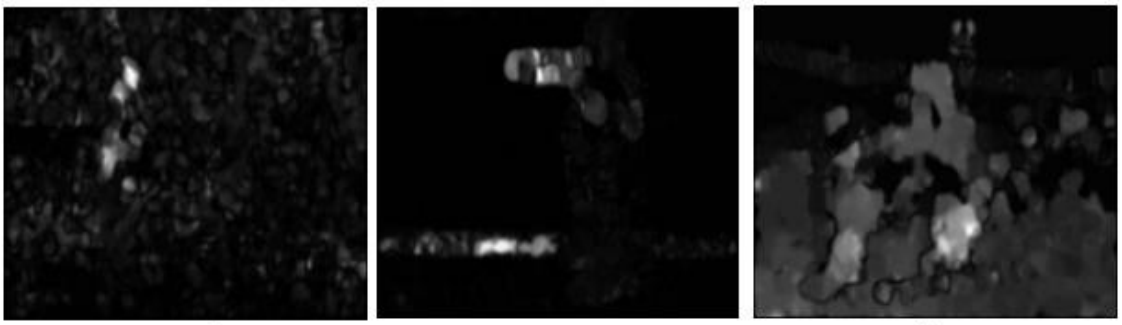
به کمک اطلاعات شاخص ایستا مبتنی بر کنتراست[1]، نواحی شاخص در یک فریم تشخیص داده می-شود. ولی شناسایی فعالیت انسان در ویدیوها مبتنی بر فریم‌های متوالی است. به منظور استفاده از

[1] Contrast based static saliency

اطلاعات پویا، اطلاعات شاخص پویا مبتنی بر کنتراست شار نوری[1] را روی مقیاس‌های هشت‌گانه هر فریم محاسبه می‌کنیم. شار نوری، حرکت هدف را میان فریم‌های متوالی مدل می‌کند. بوسیله محاسبه نقشه اطلاعات شاخص شار نوری[2]، اکثر نواحی دارای حرکت شاخص، در نقشه شار نوری تشخیص داده می‌شوند. با استفاده از رابطه (۴-۹)، اطلاعات شاخص پویا پیکسل X_{fi} در فریم f بوسیله محاسبه هیستوگرام‌های شارنوری دو فریم متوالی و سپس مقایسه بوسیله فاصله chi_square بدست می‌آید.

$$C_d(X_{fi}) = \chi^2\left(h(X_{fi}), h\left(A(X_{fi})\right)\right) \qquad (۴-۹)$$

که $h(X_{fi})$ بردار هیستوگرام شار نوری پیکسل X_{fi} و $h\left(A(X_{fi})\right)$ بردار شارنوری متوسط پیکسل‌های فریم f است. مثال‌هایی از اطلاعات شاخص پویا در شکل (۴-۷) نشان داده شده‌است.



شکل ۴-۷ نمایش اطلاعات شاخص پویا بر روی فعالیت‌های الف)گلف، ب)بوکس و ج) اسب سواری

برای ویدئوهای با پس زمینه ساده، اطلاعات شاخص ایستا از اطلاعات شاخص پویا کارآمدتر است، اما برای ویدیوهای واقعی که دارای پس زمینه شلوغ هستند، اطلاعات شاخص ایستا نواحی دارای حرکت انسان را از نواحی ثابت تشخیص نمی‌دهد. از این رو، ترکیب اطلاعات ایستا و پویا در شناسایی فعالیت

[1] optical_flow contrast based dynamic saliency
[2] saliency optical flow map

ویدئوهای واقعی کارآمد است. به همین منظور اطلاعات شاخص ایستا و پویا بصورت رابطه (۴-۱۰) ترکیب می‌شوند.

$$C_c(X_{fi}) = \overline{C_s}(X_{fi}).e^{-\alpha\overline{C_d}(X_{fi})} + \overline{C_d}(X_{fi}).e^{-\beta\overline{C_s}(X_{fi})} \qquad (۴-۱۰)$$

که در آن، $\overline{C_s}$ و $\overline{C_d}$ اطلاعات شاخص ایستا و پویا بعد از نرمال سازی هستند. α و β پارامترهای کنترل کننده تاثیر اطلاعات شاخص ایستا و پویا هستند که در این تحقیق ۱ در نظر گرفته شده‌است. نتایج پیاده سازی ترکیب اطلاعات شاخص بر روی چند فعالیت در شکل (۴-۸) نشان داده شده است.

شکل ۴-۸ نمایش اطلاعات شاخص ترکیبی فعالیت‌های الف) بازی گلف، ب) بوکس و ج) اسب سواری

۴-۲-۳ استخراج مسیر حرکت شاخص

پس از استخراج مسیرهای متراکم، اطلاعات شاخص ترکیبی را برای حذف مسیرهای متراکم غیرمرتبط و زائد بکار می‌گیریم. برای این کار ابتدا باید مقدار اطلاعات شاخص ترکیبی پیکسل‌های تشکیل دهنده یک مسیر حرکت را بدست آورد. از آنجاییکه مؤلفه‌های بردار ویژگی مسیر حرکت، بیانگر جابجایی[1] پیکسل‌های مسیرحرکت است، برای محاسبه اطلاعات شاخص مسیر حرکت از مختصات میانگین (m_x, m_y)، پیکسل‌های مسیرحرکت استفاده می‌کنیم. به این صورت که پنجره‌ای به ابعاد d*d حول میانگین مختصات نقاط مسیر حرکت در فریم‌های متوالی در نظر گرفته و میزان

[1] displacement

اطلاعات شاخص ترکیبی هر یک از نقاط پنجره را بوسیله رابطه (۴-۱۰) محاسبه می‌کنیم. مقدار بیشینه مقادیر درون پنجره را بعنوان میزان اطلاعات شاخص ترکیبی نقطه مسیر حرکت در آن فریم در نظر می‌گیریم. این فرایند را برای تمام مقیاس‌ها انجام داده و محاسبات در مقیاسی که مسیر حرکت استخراج شده، انجام می‌شود. مقدار پارامتر d را در این تحقیق، ۳۱ در نظر گرفته‌ایم.

مؤلفه‌های دوم و سوم بردار اطلاعات مکانی مسیر حرکت که بیانگر میانگین مختصات نقاط مسیر حرکت هستند، در فرم بدون مقیاس دهی ثبت شده است، از این رو مختصات جدید میانگین نقاط مسیر حرکت در مقیاس S را از رابطه (۴-۱۱) و ابعاد جدید پنجره در مقیاس S را از رابطه (۴-۱۲) بدست می‌آوریم. سپس با استفاده از رابطه (۴-۱۳) میزان شاخص بودن مسیر حرکت بدست می-آید.

$$\left(m_{xs}, m_{ys}\right) = \left(\frac{m_x}{(\sqrt{2})^{s-1}}, \frac{m_y}{(\sqrt{2})^{s-1}}\right) \qquad (۴-۱۱)$$

$$d_s = \frac{d}{(\sqrt{2})^{s-1}} \qquad (۴-۱۲)$$

$$S_c(t_i) = \frac{1}{L}\sum_{f=1}^{L} \overline{C_c}\left(X_{fi}\right) \qquad (۴-۱۳)$$

$$t_i = [X_{1i} \dots X_{Li}] \qquad (۴-۱۴)$$

که در آن، t_i مسیر حرکت در پیکسل i تعریف می‌شود، L طول مسیر حرکت که در این تحقیق ۱۵در نظر گرفته شده‌است و $\overline{C_c}$ اطلاعات شاخص ترکیبی بعد از نرمال سازی است. در نهایت مسیرهای حرکتی که میزان شاخص بودن آنها از یک آستانه‌ای بیشتر باشد بعنوان مسیر حرکت شاخص در نظر گرفته می‌شود. مسیر حرکت شاخص بوسیله رابطه (۴-۱۵) تعیین می‌شود.

$$T = \left\{ t_i \middle| t_i \in T_1, S_c(t_i) > \mu \frac{1}{L}\sum_{f=1}^{L} E\left(\bar{C}_c\left(X_{fi}\right)\right) \right\} \qquad (۴-۱۵)$$

$$E\left(\bar{C}_c(X_{fi})\right) = \frac{1}{W \times h}\sum_{u_{f=1}}^{w}\sum_{v_f=1}^{h} C_c(X_f) \quad (۱۶-۴)$$

$$X_{fi} = [u_{fi}, v_{fi}]^T \quad (۱۷-۴)$$

T_1 مجموعه مسیرهای حرکت متراکم استخراج شده است، μ میزان فیلترینگ شدت روشنایی را کنترل می‌کند و $\mu = ۱.۴$ در نظر گرفته شده‌است. $E\left(\bar{C}_c(X_{fi})\right)$ متوسط اطلاعات شاخص ترکیبی فریمی که پیکسل i در آن فریم وجود دارد، تعریف می‌شود. W و h، عرض و ارتفاع هر فریم است.

۴-۲-۴ خوشه‌بندی دولایه

در این بخش به منظورتولید ویژگی معنایی از یک خوشه‌بندی دو لایه استفاده می‌کنیم. ویژگی‌های بخش‌های متحرک[1] و اشیای مرتبط با فعالیت را بعنوان ویژگی‌های معنایی انتخاب می‌کنیم. تعریف و استخراج ویژگی‌های معنایی امری نسبتاً سخت و با نظارت است چراکه بخش‌های متحرک درطول زمان و مکان نامشخص هستند. همانطور که می‌دانیم مسیرهای حرکت بخش‌های متحرک متناظر، اطلاعات و الگوهای مکانی مشابه‌ی مانند متوسط مختصات نقاط تشکیل دهنده مسیرحرکت و شکل مسیرحرکت دارند. با گروه‌بندی مسیرحرکت‌هایی که دارای اطلاعات مکانی مشابه هستند، می‌توان ویژگی سطح پایین مسیرحرکت را به ویژگی معنایی سطح متوسط تبدیل کرد. برای این کار الگوریتم خوشه بندی دو لایه‌ای را برای به دست آوردن بخش‌های دارای حرکت یک فعالیت پیشنهاد می‌دهیم.

روش خوشه بندی دو لایه پیشنهادی شامل مراحل زیر است:

۱. نمونه برداری تصادفی از مسیرهای حرکت شاخص یک ویدئو
۲. خوشه بندی AP بر اساس اطلاعات مکانی در لایه اول
۳. تعیین مراکز خوشه بندی بوسیله k-means در لایه دوم
۴. گروه بندی تمام مسیرهای حرکت شاخص

[1] Moving parts

تعداد مسیرهای حرکت استخراج شده از ویدئوها معمولا بسیار زیاد است و این باعث افزایش زمان فرآیند خوشه‌بندی می‌شود. به همین منظور n مسیرحرکت شاخص هر ویدیو را بصورت تصادفی انتخاب می‌کنیم. در این مطالعه n=۵۰۰۰ در نظر گرفته شده‌است. اگر تعداد مسیرهای حرکت کمتر از n باشد، تمام مسیرهای حرکت ویدیو انتخاب می‌شوند. در فرآیند الگوریتم خوشه‌بندی ۲ لایه پیشنهادی، از اطلاعات مکانی و شکل مسیرحرکت استفاده می‌کنیم. بر خلاف مرجع [۲۳] که مسیرحرکت را ۳۰ بعدی در نظر گرفته است، مسیرحرکت را با برداری ۴۰ بعدی نشان می‌دهیم. ۱۰ مولفه اول این بردار شامل شماره فریم انتهایی مسیرحرکت (۱ مولفه)، مختصات نقطه شروع مسیرحرکت (۳ مولفه)، متوسط مختصات نقاط تشکیل دهنده مسیرحرکت (۲ مولفه)، واریانس مختصات (۲ مولفه)، مجموع جابجایی نقاط مسیرحرکت (۱ مولفه) و مقیاس فریمی که مسیرحرکت در آن مقیاس استخراج شده است(۱ مولفه). ۳۰ مولفه بعدی شکل توصیفگر مسیرحرکت را نشان می‌دهند. هریک از مؤلفه‌های بردار ویژگی مسیر حرکت را بین $[۰,۱]$ نرمال می‌کنیم.

در لایه اول الگوریتم پیشنهادی، از روش خوشه‌بندی انتشار وابستگی برای خوشه‌بندی مسیرهای حرکتی که بصورت تصادفی انتخاب شده، استفاده می‌کنیم. تنها ورودی الگوریتم AP ماتریس شباهت داده های ورودی است، در نتیجه انتخاب معیار شباهت مناسب برای خوشه‌بندی مهم است. فرض کنید n مسیرحرکت شاخص را با X_1 ... X_n نمایش دهیم. معیار شباهت کسینوسی را برای اندازه‌گیری فاصله میان دو بردارهای توصیفگر مسیرحرکت بکار گرفته‌ایم. اگر $i \neq j$ باشد، شباهت میان مسیرحرکت i و j از رابطه (۴-۱۸) بدست می‌آید.

$$d_{ij} = 1 - \cos\theta = 1 - \frac{X_i.X_j}{\|X_i\|.\|X_j\|} = 1 - \frac{\sum_{k=1}^{40} x_{ik} \times x_{jk}}{\sqrt{\sum_{k=1}^{40} (x_{ik})^2} \times \sqrt{\sum_{k=1}^{40} (x_{jk})^2}} \qquad (۴-۱۸)$$

بعد از محاسبه معیار شباهت کسینوسی میان مسیرهای حرکت شاخص، و d_{ii} که صفر در نظر گرفته‌ایم، ماتریسی $n \times n$ از فواصل بین مسیرهای حرکت بدست می‌آید که بصورت رابطه (۴-۱۹) تعریف می‌شود.

$$D = \{(i, j, d_{ij}) | i = 1, \dots, n ; j = 1, \dots, n\} \quad (۴-۱۹)$$

از آنجاییکه تعداد خوشه ها در خوشه‌بندی AP مشخص نیست، از الگوریتم k-means برای تعیین مراکز خوشه‌ها در لایه دوم استفاده می‌شود. تعداد مراکز خوشه الگوریتم k-means، C خوشه تعیین می‌شود. در نهایت بوسیله محاسبه معیار شباهت کسینوسی با مراکز خوشه، مسیرهای حرکت شاخص ویدیو در خوشه‌های متناظر خود دسته‌بندی شده و C گروه را تشکیل می‌دهند. در این تحقیق C= ۳ در نظر می‌گیریم.

۴-۲-۵ کد کردن ویژگی ها

بعد از گروه بندی مسیرهای حرکت، تمام مسیرهای حرکت شاخص یک ویدیو به C گروه تقسیم شده است، بعلاوه یک گروه که از تمام مسیرهای حرکت ویدیو تشکیل می‌شود، در مجموع C+1 گروه وجود دارد. در بعضی از سناریوهای پیچیده، قسمت‌های متحرک فعالیت در یک نقطه متمرکز شده‌اند که باعث گروه بندی نامطلوب مسیرهای حرکت می‌شود. در نتیجه، نمایش کل مسیرهای حرکت به عنوان یک گروه برای حل این مشکل در نظر گرفته می‌شود.

در این بخش بردارهای ویژگی گروه‌های ساخته شده را بوسیله الگوریتم کیف کلمات کد گذاری می‌کنیم. در نتیجه برای هر ویدیو ۱۲=(C+1)*۳ هیستوگرام وقوع کلمات به دست می‌آید.

برای تولید کتاب کد و کاهش زمان ۱۰۰.۰۰۰ ویژگی شاخص بصورت تصادفی از ویدیوهای آموزشی انتخاب می‌کنیم. سپس بوسیله خوشه بندی k-means با تعداد مراکز خوشه ۴۰۰۰ و ۱۰۰۰ مراکز خوشه را پیدا کرده و کتاب کد را تشکیل می‌دهیم. برای افزایش دقت، ۵ بار خوشه بندی را انجام داده و نتیجه دارای کمترین خطا را نگه می‌داریم. در نهایت هر ویدیو بوسیله ۱۲ بردار ۴۰۰۰ مولفه‌ای نمایش داده می‌شود.

۴-۲-۶طبقه بندی

بعد از کد کردن ویژگی‌ها، با استفاده از طبقه‌بند k نزدیک‌ترین همسایه KNN[1] به شناسایی فعالیت‌های ویدئویی می‌پردازیم. در این روش با داشتن بردارهای ویژگی مجموعه آموزش و نمونه آزمون نزدیک‌ترین داده را در بین مجموعه آموزش پیدا می‌کند. برای عمل کلاسه‌بندی کافی است k نزدیک‌ترین همسایه را از نمونه‌های آموزش مشخص کنیم و کلاس آن را به نمونه آزمون نسبت دهیم. برای پیدا کردن نزدیک‌ترین همسایه از معیار فاصله اقلیدسی استفاده کرده‌ایم.

۴-۳ معرفی پایگاه داده

این تحقیق بر روی پایگاه داده UCF sports [۵۱]، اجرا شده و نتایج آن با مطالعات پیشین مقایسه می‌شود.

۴-۳-۱ پایگاه داده UCF sports

این پایگاه داده در سال ۲۰۰۸ توسط دانشگاه فلوریدای مرکزی[2] (UCF) معرفی شده و شامل ۳۰۰ ویدیوی ورزشی مختلف است که از شبکه های تلویزیونی جمع آوری شده و به ۱۰ فعالیت تقسیم شده است. از جمله این فعالیت‌ها می‌توان شیرجه زدن، وزنه برداری، اسب سواری و ... را نام برد. چالش‌هایی مانند زاویه دید مختلف، حرکت دوربین و تنوع صحنه در این پایگاه داده وجود دارد. فعالیت‌های این پایگاه داده در شکل (۴-۹) نشان داده شده‌است.

[1] K-neaest neighbour
[2] University of Central Florida

شکل ۴-۹ نمایش فعالیت‌های پایگاه داده UCF sports [۵۱].

فصل ۵: نتایج تجربی و مقایسه

۵-۱ مقدمه

در این فصل روش پیشنهادی را بر روی پایگاه داده UCF sports که در بخش ۴-۳ معرفی کرده‌ایم، اعمال کرده و نتایج شبیه‌سازی را گزارش می‌کنیم و در پایان روش‌پیشنهادی را با سایر روش‌ها مقایسه می‌کنیم.

۵-۲ معیار ارزیابی

برای ارزیابی نتایج این تحقیق، از معیار متوسط صحت[1] استفاده می‌شود تا امکان مقایسه این نتایج با تحقیقات گذشته فراهم گردد.

$$\text{صحت} = \frac{\text{تعداد داده‌هایی که دسته‌ی آنها به درستی شناسایی شد}}{\text{تعداد کل داده‌ها}}$$

۵-۳ ارزیابی نتایج

در این پایان‌نامه روشی برای شناسایی فعالیت‌های انسانی با استفاده از ترکیب ویژگی‌های معنایی و غیرمعنایی ارائه شد. در این روش ابتدا توصیفگرهای مسیرحرکت، HOG، HOF و MBH از ویدیوها استخراج شد. سپس با استفاده از اطلاعات شاخص نقاط تشکیل دهنده مسیرحرکت، مسیرهای حرکت شاخص تعیین شدند. با بکارگیری الگوریتم خوشه‌بندی دولایه، مسیرهای حرکت شاخص را گروه‌بندی کرده، که هر گروه بخش متحرکی از فعالیت را نشان می‌دهد. مسیرهای حرکت معنایی هر گروه را با استفاده از الگوریتم کیف کلمات کد کرده و در نهایت با استفاده از طبقه‌بند K نزدیک‌ترین همسایگی فعالیت‌ها شناسایی شدند.

ارزیابی روش پیشنهادی شامل سه بخش: ارزیابی با استفاده از تمام مسیرهای حرکت (Dense trajectories)، ارزیابی با استفاده از مسیرهای حرکت شاخص (salient trajectories) و

[1] Average accuracy

ارزیابی با استفاده از مسیرهای حرکت معنایی (Semantic trajectories) است. در این ارزیابی‌ها از ۴ توصیفگر مسیرحرکت، **HOG**، **HOF**، **MBH** استفاده کرده‌ایم.

از آنجاییکه اطلاعات مسیر حرکت در خوشه بندی دو لایه، برای تشکیل مسیرهای حرکت معنایی، بکار گرفته شده است، همچنین به دلیل حجم محاسباتی زیاد، در ارزیابی با استفاده از مسیرهای حرکت معنایی، توصیفگر مسیرحرکت بکار گرفته نشده است.

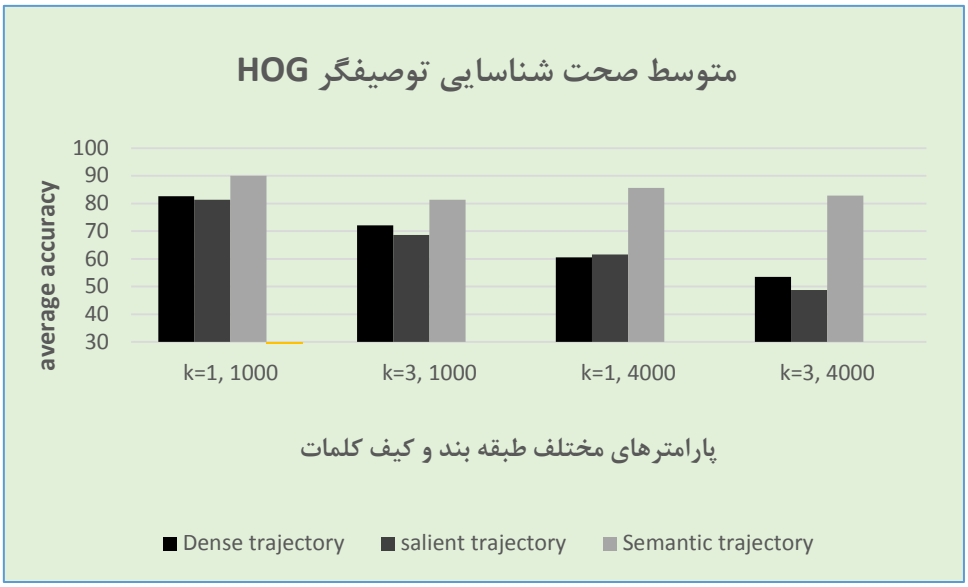
در جدول ۵-۱ بهترین عملکرد ارائه شده با استفاده از روش پیشنهادی و پارامترهای مختلف طبقه بند و کیف کلمات روی پایگاه داده UCF sports مشاهده می‌شود. k پارامتر طبقه بند KNN است و تعداد مراکز خوشه الگوریتم کیف کلمات را ۱۰۰۰ و ۴۰۰۰ در نظر گرفته‌ایم.

همانطور که مشاهده می‌شود استفاده از ویژگی معنایی صحت را نسبت به روش شناسایی با استفاده از مسیرهای حرکت شاخص و تمام مسیرهای حرکت افزایش داده است.

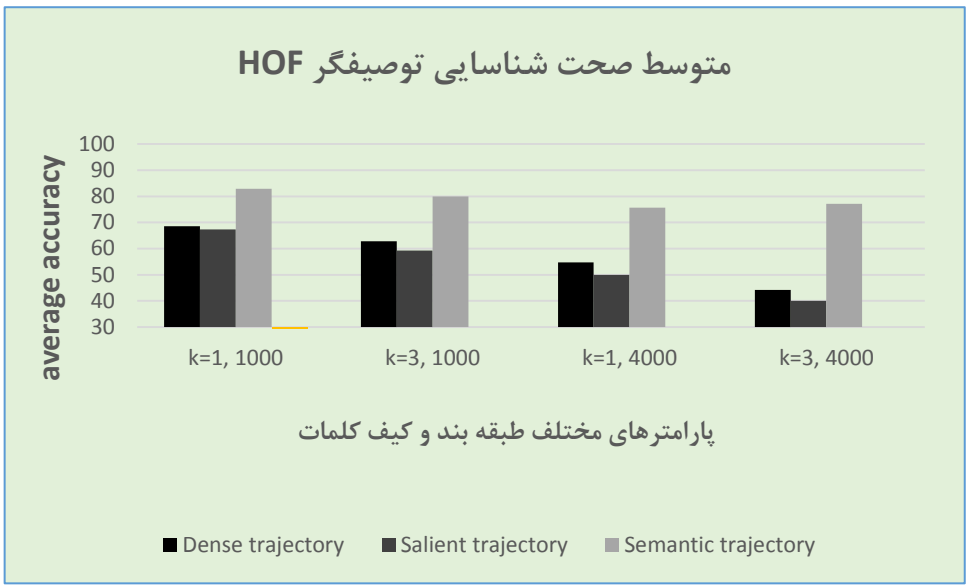
جدول ۵-۱ نتایج ارزیابی برحسب درصد بر روی پایگاه داده UCF sports

توصیفگر	Dense trajectories				Salient trajectories				Semantic trajectories			
	1000 k=1	1000 k=3	4000 k=1	4000 k=3	1000 k=1	1000 k=3	4000 k=1	4000 k=3	1000 k=1	1000 k=3	4000 k=1	4000 k=3
Trajectory	۶۷/۵	۶۵/۱	۵۵/۸	۵۱/۲	۶۷/۴	۶۴	۵۲/۳	۴۳	–	–	–	–
HOG	۸۲/۶	۷۲/۱	۶۰/۵	۵۳/۵	۸۱/۴	۶۸/۶	۶۱/۶	۴۸/۸	۹۰	۸۱/۴	۸۵/۷	۸۲/۹
HOF	۶۸/۶	۶۲/۸	۵۴/۷	۴۴/۲	۶۷/۴	۵۹/۳	۵۰	۳۸/۸	۸۲/۹	۸۰	۷۵/۷	۷۷/۱
MBH	۷۴/۴	۷۰/۹	۵۷	۵۱/۲	۶۹/۸	۶۴	۴۷/۷	۴۴/۲	۹۲/۹	۹۲/۹	۷۷/۱	۷۵/۷
HOG+HOF +MBH	۷۵/۶	۶۰/۵	۵۸/۱	۴۷/۷	۶۵/۱	۶۰/۵	۵۴/۷	۴۴/۲	۸۸/۶	۹۰	۸۵/۷	۸۴/۳

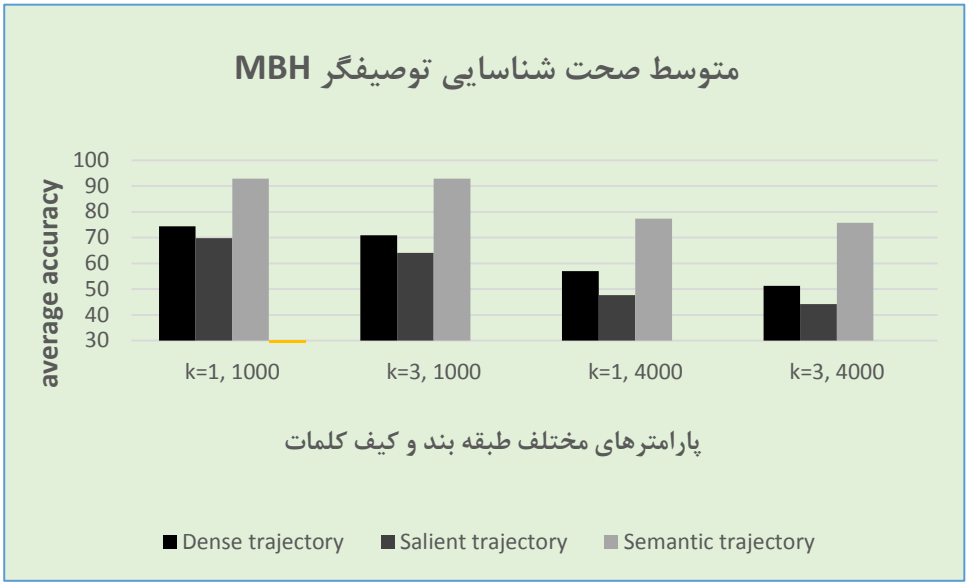
همانطور که در بخش ۳-۳ گفته شد، یکی از چالش‌های پایگاه داده UCF sports، حرکت دوربین است و می‌دانیم که توصیفگر MBH نسبت به حرکت دوربین مقاوم است. همان طور که انتظار می-رفت، توصیفگر MBH بیشترین کارآیی را در شناسایی فعالیت‌های پایگاه داده UCF sports داشته-است. عملکرد توصیفگرهای مختلف در شکل‌های (۵-۱) تا (۵-۴) نشان داده شده‌است



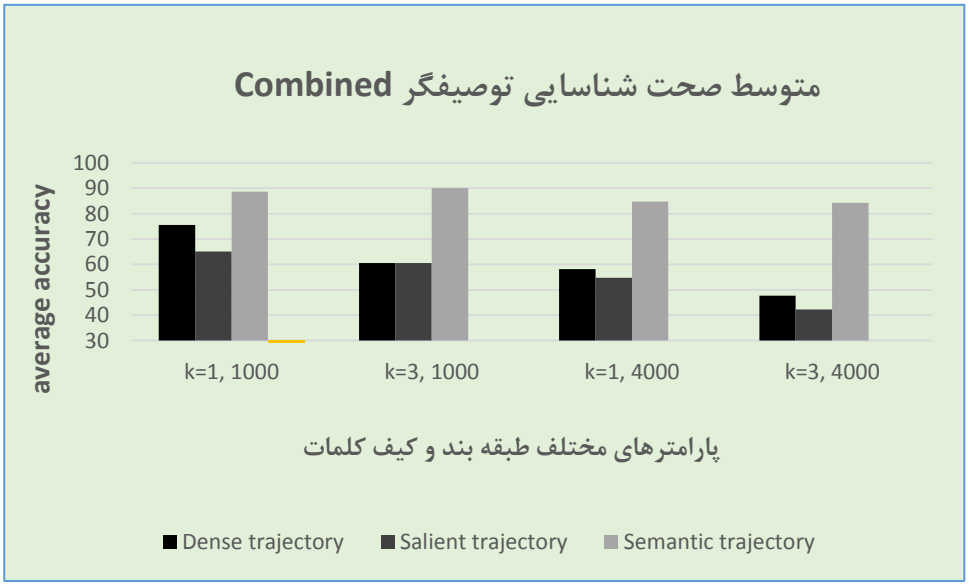
شکل ۵-۱عملکرد توصیفگر HOG در ارزیابی روش پیشنهادی



شکل ۵-۲ عملکرد توصیفگر HOF در ارزیابی روش پیشنهادی



شکل ۵-۳ عملکرد توصیفگرMBH در ارزیابی روش پیشنهادی



شکل ۵-۴ عملکرد توصیفگر Combined در ارزیابی روش پیشنهادی

جدول (۵-۲) صحت بدست آمده برای هر یک از فعالیت‌های پایگاه داده UCF sports را نشان می-دهد. در ارزیابی مسیر حرکت معنایی اکثر فعالیت‌ها بهتر شناسایی می‌شوند. برخی از فعالیت‌ها در ارزیابی با استفاده از مسیر حرکت شاخص بهتر از روش شناسایی با استفاده از تمام مسیرهای حرکت، تشخیص داده می‌شوند.

جدول ۵-۲ نتایج ارزیابی روش پیشنهادی بر روی فعالیت‌های پایگاه داده UCF sports

Semantic trajectory				Salient trajectory				Dense trajectory				توصیفگر / نام فعالیت	
HOG	HOF	MBH	COM	HOG	HOF	MBH	COM	HOG	HOF	MBH	COM		
۱۰۰	۱۰۰	۱۰۰	۱۰۰	۶۲	۷۵	۸۷	۶۲	۵۰	۶۲	۳۸	۵۰	1000 K=1	Diving
۸۸/۹	۶۶/۷	۶۶/۷	۶۶/۷	۲۵	۳۸	۲۵	۳۸	۳۸	۳۸	۳۸	۳۸	4000 K=1	
۱۰۰	۸۰	۸۰	۸۰	۱۰۰	۱۰۰	۸۰	۹۰	۹۰	۹۰	۸۰	۹۰	1000 K=1	Golf-
۱۰۰	۸۰	۸۰	۱۰۰	۱۰۰	۸۰	۷۰	۹۰	۱۰۰	۷۰	۹۰	۹۰	4000 K=1	
۱۰۰	۶۶/۷	۱۰۰	۸۸/۹	۶۶	۳۴	۵۰	۳۴	۷۵	۲۵	۳۴	۳۴	1000 K=1	kicking
۷۷/۸	۶۶/۷	۷۷/۸	۸۸/۶	۳۳	۱۷	۲۵	۱۷	۳۴	۳۴	۵۰	۴۲	4000 K=1	
۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	1000 K=1	lifting
۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	4000 K=1	
۷۱/۴	۴۲/۹	۸۵/۷	۷۱/۴	۵۷	۲۹	۴۲	۲۹	۴۳	۲۹	۸۵	۲۹	1000 K=1	Ridingh
۷۱	۲۹	۷۱/۴	۵۷/۱	۵۷	۲۹	۴۳	۴۳	۴۲	۲۹	۵۷	۴۳	4000 K=1	
۸۳/۳	۶۶/۷	۸۳/۳	۶۶/۷	۸۵	۴۲	۴۲	۲۹	۵۷	۲۹	۴۳	۲۹	1000 K=1	Run-
۸۳/۳	۸۳/۳	۸۳/۳	۸۳/۳	۴۳	۲۹	۲۹	۲۹	۲۹	۲۹	۲۹	۲۹	4000 K=1	
۸۰	۸۰	۶۰	۱۰۰	۵۷	۵۷	۴۲	۵۷	۱۶	۲۹	۴۳	۲۹	1000 K=1	Skatebo
۱۰۰	۶۰	۲۰	۱۰۰	۲۹	۱۴	۲۸	۲۹	۱۴	۲۹	۱۴	۴۳	4000 K=1	
۱۰۰	۹۰.۹	۱۰۰	۹۰/۹	۱۰۰	۸۳	۹۱	۱۰۰	۱۰۰	۸۳	۹۱	۸۳	1000 K=1	Swing
۱۰۰	۸۱/۸	۷۲/۷	۹۰/۹	۱۰۰	۵۹	۶۶	۷۵	۱۰۰	۶۷	۷۵	۷۵	4000 K=1	
۸۳/۳	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۷۱	۸۵	۸۵	۸۵	۷۱	۷۱	۸۵	1000 K=1	Swing
۸۳/۳	۱۰۰	۱۰۰	۱۰۰	۵۷	۷۱	۷۱	۷۱	۸۵	۷۱	۷۱	۸۵	4000 K=1	
۷۷/۸	۱۰۰	۱۰۰	۸۸/۹	۸۴	۸۴	۷۶	۶۹	۹۲	۷۶	۷۶	۷۶	1000 K=1	Walk
۶۶/۷	۱۰۰	۱۰۰	۸۸/۹	۶۹	۷۶	۶۱	۶۱	۶۱	۸۴	۵۴	۵۴	4000 K=1	

۵-۴ مقایسه با سایر روش‌ها

در جدول ۵-۳ مقایسه‌ای بین الگوریتم پیشنهادی با سایر الگوریتم‌هایی که از پایگاه داده UCF sport استفاده می‌کنند، انجام شده است. همانطور که مشاهده می‌شود، روش پیشنهادی نسبت به روش‌های دیگر فعالیت‌های بیشتری را به درستی شناسایی کرده است.

جدول ۵-۳ مقایسه الگوریتم پیشنهادی با سایر روش‌ها

مرجع	متوسط صحت
[52]Abdul azim	٪۸۹.۹۷
[29]Cho	٪۹۰.۳
[25]Yang Yi	٪۹۱.۳۷
[24]Wang	٪۸۹.۱
روش پیشنهادی	**٪۹۲.۹**

۵-۵ پیشنهادها و کارهای آینده

- استفاده از ویژگی‌های معنایی دیگر مانند زمینه[1] محل وقوع فعالیت و یا شی مرتبط با آن
- ارتقا الگوریتم کیف کلمات با استفاده از اطلاعات شاخص مسیر حرکت بعنوان ضریب در معیار فاصله خوشه‌بندی k-میانگین

[1] context

واژگان و اصطلاحات

Single-Layered	تک لایه‌ای
Hierarchical	سلسله مراتبی
Space-Time	مکان-زمان
Similarity Measure	معیار شباهت
Motion Energy Image (MEI)	تصویر انرژی حرکت
Motion History Image (MHI)	تصویر تاریخچه حرکت
Spatiotemporal Volume (STV)	حجم مکان-زمان
Spatiotemporal Interest points (STIPs)	نقاط مهم فضا-زمان
Harris Corner Detector	آشکارساز گوشه هریس
Hessian Detector	آشکارساز هسین
Occlusion	انسداد
Cubiod	مکعب
Trajectory	مسیر حرکت
Dense Trajectory	مسیر حرکت متراکم
Histogram of oriented gradients (HOG)	هیستوگرام گرادیان جهت دار
Histogram of optical flow (HOF)	هیستوگرام شار نوری
Histogram of motion boundary	هیستوگرام مرز حرکت
Optical flow field	میدان شار نوری
Dense sampling	نمونه برداری متراکم
Sparse	پراکنده
Kanade-Lucas-Tomasi (KLT) Tracker	ردیاب لوکاس-کانادی
Irregular Abrupt Motion	حرکت نامنظم ناگهانی
Salient trajectory	مسیر حرکت شاخص
Saliency	شاخص
Aggregation	تجمیع
Bag ofwords	کیف کلمات
Fisher vector	بردار فیشر
Sparse representation	نمایش تنک
Atomic action	عمل بنیادی

Hidden markov model (HMM)	مدل مخفی مارکوف
Hierarchical Spatio-Temporal Model (HSTM)	مدل مکان-زمانی سلسله مراتبی
Hidden Conditional Random Field(HCRF)	میدان تصادفی شرطی مخفی
Semanticfeature	ویژگی معنایی
attribute	مشخصه
Semantic space	فضای معنایی
Support Vector Machine	ماشین بردار پشتیبان
Average Precision (AP)	متوسط دقت
Code Book	کتاب کد
Visual words	کلمات بصری
Affinity propagation Clustering	خوشه‌بندی انتشار وابستگی
Availability matrix	ماتریس دسترسی
Responsibility matrix	ماتریس وظیفه
Contrast based saliency filter (CSF)	فیلتر شاخص سنجی مبتنی بر کنتراست
Static saliency	اطلاعات شاخص ایستا
Dynamic saliency	اطلاعات شاخص پویا
Distribution	توزیع
Simple Linear Iterative Clustering (SLIC)	خوشه‌بندی متناوب خطی ساده
Color contrast saliency	اطلاعات شاخص مبتنی بر کنتراست رنگ
Distribution contrast saliency	اطلاعات شاخص مبتنی بر کنتراست توزیع
optical flow contrast based dynamic saliency	اطلاعات شاخص پویا مبتنی برکنتراست شارنوری
Moving parts	بخش‌های متحرک
K-nearest neighbour classifier	طبقه‌بند k نزدیک‌ترین همسایه
average accuracy	متوسط صحت
Context	زمینه

مراجع

[1]. Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. *ACM Computing Surveys (CSUR), 43*(3), 16.

[2]. Vishwakarma, S., & Agrawal, A. (2013). A survey on activity recognition and behavior understanding in video surveillance. *The Visual Computer*, *29*(10), 983-1009.

[3]. Fu, Y. (Ed.). (2015). *Human Activity Recognition and Prediction*. Springer.

[4]. Ziaeefard, M., & Bergevin, R. (2015). Semantic human activity recognition: a literature review. *Pattern Recognition*, *48*(8), 2329-2345.

[5]. Maji, S., Bourdev, L., & Malik, J. (2011, June). Action recognition from a distributed representation of pose and appearance. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (pp. 3177-3184).

[6]. Liu, J., Kuipers, B., & Savarese, S. (2011, June). Recognizing human actions by attributes. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (pp. 3337-3344).

[7]. Herath, S., Harandi, M., & Porikli, F. (2017). Going deeper into action recognition: A survey. *Image and Vision Computing*, *60*, 4-21.

[8]. Cheng, G., Wan, Y., Saudagar, A. N., Namuduri, K., & Buckles, B. P. (2015). Advances in human action recognition: A survey. *arXiv preprint arXiv:1501.05964*.

[9]. Pusiol, G. T. (2012). *Discovery of human activities in video*(Doctoral dissertation, Université Nice Sophia Antipolis).

[10]. Bobick, A. F., & Davis, J. W. (2001). The recognition of human movement using temporal templates. *IEEE Transactions on pattern analysis and machine intelligence*, *23*(3), 257-267.

[11]. Blank, M., Gorelick, L., Shechtman, E., Irani, M., & Basri, R. (2005, October). Actions as space-time shapes. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on* (Vol. 2, pp. 1395-1402). IEEE.

[12]. Yilmaz, A., & Shah, M. (2005, June). Actions sketch: A novel action representation. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 1, pp. 984-989). IEEE.

[13]. Laptev, I. (2005). On space-time interest points. *International journal of computer vision*, *64*(2-3), 107-123.

[14]. Harris, C., & Stephens, M. (1988, August). A combined corner and edge detector. In *Alvey vision conference* (Vol. 15, No. 50, pp. 10-5244).

[15]. Willems, G., Tuytelaars, T., & Van Gool, L. (2008). An efficient dense and scale-invariant spatio-temporal interest point detector. *Computer Vision–ECCV 2008*, 650-663.

[16]. Liu, J., Luo, J., & Shah, M. (2009, June). Recognizing realistic actions from videos "in the wild". In *Computer vision and pattern recognition, 2009. CVPR 2009. IEEE conference on*(pp. 1996-2003). IEEE.

[17]. Dollár, P., Rabaud, V., Cottrell, G., & Belongie, S. (2005, October). Behavior recognition via sparse spatio-temporal features. In *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on* (pp. 65-72). IEEE

[18]. Messing, R., Pal, C., & Kautz, H. (2009, September). Activity recognition using the velocity histories of tracked keypoints. In *Computer Vision, 2009 IEEE 12th International Conference on*(pp. 104-111). IEEE.

[19]. Matikainen, P., Hebert, M., & Sukthankar, R. (2009, September). Trajectons: Action recognition through the motion analysis of tracked features. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on* (pp. 514-521). IEEE.

[20]. Klaser, A., Marszałek, M., & Schmid, C. (2008, September). A spatio-temporal descriptor based on 3d-gradients. In *BMVC 2008-19th British Machine Vision Conference* (pp. 275-1). British Machine Vision Association.

[21]. Laptev, I., Marszalek, M., Schmid, C., & Rozenfeld, B. (2008, June). Learning realistic human actions from movies. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8). IEEE.

[22]. Dalal, N., Triggs, B., & Schmid, C. (2006, May). Human detection using oriented histograms of flow and appearance. In *European conference on computer vision* (pp. 428-441). Springer, Berlin, Heidelberg.

[23]. Wang, H., Kläser, A., Schmid, C., & Liu, C. L. (2011, June). Action recognition by dense trajectories. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*(pp. 3169-3176). IEEE.

[24]. Wang, H., Kläser, A., Schmid, C., & Liu, C. L. (2013). Dense trajectories and motion boundary descriptors for action recognition. *International journal of computer vision*, *103*(1), 60-79

[25]. Yi, Y., Zheng, Z., & Lin, M. (2017). Realistic action recognition with salient foreground trajectories. *Expert Systems with Applications*, *75*, 44-55.

[26]. Peng, X., Zou, C., Qiao, Y., & Peng, Q. (2014, September). Action recognition with stacked fisher vectors. In *European Conference on Computer Vision* (pp. 581-595). Springer, Cham.

[27]. Oneata, D., Verbeek, J., & Schmid, C. (2013). Action and event recognition with fisher vectors on a compact feature set. In *Proceedings of the IEEE international conference on computer vision* (pp. 1817-1824).

[28]. Wang, H., & Schmid, C. (2013). Action recognition with improved trajectories. In *Proceedings of the IEEE international conference on computer vision* (pp. 3551-3558).

[29]. Cho, J., Lee, M., Chang, H. J., & Oh, S. (2014). Robust action recognition using local motion and group sparsity. *Pattern Recognition*, *47*(5), 1813-1825.

[30]. Zhu, Y., Zhao, X., Fu, Y., & Liu, Y. (2010, November). Sparse coding on local spatial-temporal volumes for human action recognition. In *Asian Conference on Computer Vision* (pp. 660-671). Springer, Berlin, Heidelberg.

[31]. Oliver, N., Horvitz, E., & Garg, A. (2002, October). Layered representations for human activity recognition. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces* (p. 3). IEEE Computer Society.

[32]. Yu, E., & Aggarwal, J. K. (2006, August). Detection of fence climbing from monocular video. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* (Vol. 1, pp. 375-378). IEEE.

[33]. Zhang, D., Gatica-Perez, D., Bengio, S., McCowan, I., & Lathoud, G. (2004, December). Modeling individual and group actions in meetings: a two-layer hmm framework. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on* (pp. 117-117). IEEE.

[34]. Raptis, M., Kokkinos, I., & Soatto, S. (2012, June). Discovering discriminative action parts from mid-level video representations. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 1242-1249). IEEE.

[35]. Ryoo, M. S., & Aggarwal, J. K. (2009, September). Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In *Computer vision, 2009 ieee 12th international conference on* (pp. 1593-1600). IEEE.

[36]. Xu, W., Miao, Z., Zhang, X. P., & Tian, Y. (2017). A Hierarchical Spatio-Temporal Model for Human Activity Recognition. *IEEE Transactions on Multimedia*.

[37]. محمودی, ا. (۱۳۹۶). "تشخیص فعالیت های انسان مبتنی بر پردازش رشته های ویدیویی." پایان نامه

[38]. Wang, C., Wang, Y., & Yuille, A. L. (2013). An approach to pose-based action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 915-922).

[39]. Chaaraoui, A. A., Climent-Pérez, P., & Flórez-Revuelta, F. (2013). Silhouette-based human action recognition using sequences of key poses. *Pattern Recognition Letters*, *34*(15), 1799-1807.

[40]. Meng, L., Qing, L., Yang, P., Miao, J., Chen, X., & Metaxas, D. N. (2012, November). Activity recognition based on semantic spatial relation. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 609-612). IEEE.

[41]. Vahdat, A., Gao, B., Ranjbar, M., & Mori, G. (2011, November). A discriminative key pose sequence model for recognizing human interactions. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* (pp. 1729-1736).

[42]. Raptis, M., & Sigal, L. (2013). Poselet key-framing: A model for human activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2650-2657).

[43]. Marszalek, M., Laptev, I., & Schmid, C. (2009, June). Actions in context. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 2929-2936).

[44]. Zhang, Y., Qu, W., & Wang, D. (2014). Action-scene model for human action recognition from videos. *AASRI Procedia*, *6*, 111-117.

[45]. Liu, J., Xiang, H., Shi, Y., & Yu, D. (2012, November). Action recognition with trajectory and scene. In *Digital Home (ICDH), 2012 Fourth International Conference on* (pp. 63-68).

[46]. Gupta, A., & Davis, L. S. (2007, June). Objects in action: An approach for combining action understanding and object perception. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on* (pp. 1-8).

[47]. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91-110.

[48]. Bouguet, J. Y. (2001). Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. *Intel Corporation*, *5*(1-10), 4.

[49]. Beauchemin, S. S., & Barron, J. L. (1995). The computation of optical flow. *ACM computing surveys (CSUR)*, *27*(3), 433-466.

[50]. Perazzi, F., Krähenbühl, P., Pritch, Y., & Hornung, A. (2012, June). Saliency filters: Contrast based filtering for salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 733-740).

[51]. Rodriguez, M. D., Ahmed, J., & Shah, M. (2008, June). Action mach a spatio-temporal maximum average correlation height filter for action recognition. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8).

[52]. Abdul-Azim, H. A., & Hemayed, E. E. (2015). Human action recognition using trajectory-based representation. *Egyptian Informatics Journal*, *16*(2), 187-198.

Abstract

Nowadays with the increasing concern on security problems and necessity for studying human activitiy, requests for an intelligent system which automatically recognizes human activities increases.

In manual inspection in which a person watches given video for recognizing human activities is inefficient mainly because of tiredness. Hence the need for an automatic system for activity recognition is demanded more than before.

Activity recognition is a challenging task due to camera motion, clutter in video background and inter personal differences. Another challenge in human activity recognition is performing an activity in different forms. Thus, we need utilize semantic information about the activity that occurred and address these challenges. Unlike low level features, semantic features describe inherent characteristics of activities. Therefore semantics make the recognition task more reliable especially when the same activities look visually different.

In this thesis, we propose an approach for extract semantic features and utilize them aligned by non-semantic features for human activity recognition. In the proposed method dense trajectories, HOG, HOF and MBH are used as non-semantic features. Then salient trajectories are determined. Salient trajectories are clustered using two layers clustering method then the obtained clusters construct semantic features. These features are encoded by bag of word algorithm. Finally by using the KNN classifier activities are recognized. The implementation of the proposed method result in 92.9% average accuracy on UCF sports dataset.

Keywords: **Human activity recognition, semantic recognition, semantic features, Dense Trajectory, Bag of Words, UCF sports dataset**

Shahrood University of Technology
Faculty of Electrical Engineering and Robotic
M.Sc. Thesis in Telecommunication Systems Engineering

Human Activity recognition using semantic and non-semantic approaches

By: Zahra Shavakandy

Supervisor:
Dr Alireza Ahmadifard

January 2018