

# تشخیص فعالیت‌های هنجار و ناهنجار انسان برای نظارت ویدیویی

سروناز صادقی ایوریک<sup>۱</sup> and محمود فضلعلی<sup>۲</sup>

<sup>۱</sup> گروه مهندسی کامپیوتر، واحد بوئین زهرا، دانشگاه آزاد اسلامی، بوئین زهرا، ایران، Sarvenaz. sadeghi@gmail. com

<sup>۲</sup> استادیار دانشگاه شهید بهشتی، تهران، نویسنده مسئول، mmmfazlali@yahoo. com

## چکیده

همزمان با افزایش تعداد استفاده‌کنندگان از دوربین‌های نظارتی در اماکن عمومی این نیاز احساس می‌شود که این دوربین‌ها بطور خودکار به تشخیص فعالیت‌های هنجار و ناهنجار بپردازند تا هم به سرعت تشخیص بالاتری برسند و هم این تشخیص بدور از خستگی و اشتباه انسانی بتواند نیروهای حفاظتی را آگاه نماید. بدین منظور یک توصیفگر یکپارچه معرفی میکنیم که در آن، به اطلاعات پس زمینه و پیش زمینه بطور همزمان تبدیل فوریه اعمال میکنیم تا اطلاعات حرکت را استخراج کرده، با استفاده از فیلتر گابور استخراج ویژگی انجام می‌دهیم، برای کاهش ابعاد از تبدیل فضا و انتخاب ویژگی MRMR استفاده می‌کنیم و در نهایت آموزش با دسته‌بندی کننده ماشین بردار پشتیبان و تشخیص فعالیت را خواهیم داشت. به برتری این روش می‌توان به خاصیت جدا نکردن پس زمینه از پیش زمینه که پاسخ گوی نیاز دآوری در بازی‌های ورزشی نیز هست و استفاده از دسته‌بندی کننده مرحله به مرحله و استفاده از انتخاب ویژگی MRMR برای کاهش ابعاد با توجه به سرعت و دقت اشاره کرد.

## کلمات کلیدی

تشخیص فعالیت، توصیفگر ویدئویی، کاهش ابعاد، استخراج ویژگی.

## ۱ تحقیقات مرتبط

سلسله مراتبی طبقه‌بندی شود. روش‌های تک لایه، بیشتر برای تشخیص عمل مناسب بوده و روش‌های سلسله مراتبی برای تشخیص فعالیت مورد استفاده قرار می‌گیرند[۱]. در سال ۲۰۱۰ نیز فورستی و همکاران[۲] با استفاده از ماشین‌های بردار پشتیبان تک کلاسه مقاله‌ای را تحت عنوان کشف رویداد نظارت محور در جریان‌های تصویری ارائه کردند. در سال ۲۰۱۴ چن و همکاران[۳] یک الگوریتم جدید بر پایه‌ی شبکه برای شناسایی فعالیت انسان در ویدئو با یادگیر SVM ارائه کرد، همین طور در سال ۲۰۱۴ کانگ و همکاران بر پایه‌ی همین دسته بند [۴] کشف رفتار غیر عادی با استفاده عامل‌های ترکیبی در صحنه‌های شلوغ را ارائه دادند. در ۲۰۱۶ [۵] نیز یو و همکاران بر پایه مدل مارکوف، الگوریتم تشخیص فعالیت بی ناظر معرفی کردند. در سال ۲۰۱۵ اوامر و همکاران [۶] با استفاده از مدل گرافیکی و تجزیه ویدئو، مقاله‌ای را تحت عنوان تجزیه فضایی-زمانی ویدئو برای کشف اختلال ارائه کردند. همه‌ی روش‌های ارائه شده در گذشته، روش‌های مبتنی بر حرکت بودند و نیازمند تخمین پس زمینه و آشکارسازی پیش‌زمینه بوده‌اند و یا بر اساس شار نوری ناشی از حرکت انسان عمل می‌کردند. در هر دو حالت، مجموع مراحل پیش‌پردازش، آشکارسازی پیش‌زمینه، و در نهایت انجام الگوریتم دسته‌بندی

همان طور که سیستم‌های نظارتی زمینه‌ی مناسبی برای افزایش تعامل انسان و ماشین محسوب می‌شود، در دهه‌ی اخیر فعالیت‌ها در زمینه‌های تعامل بین انسان و ماشین افزایش چشم‌گیری داشته است. سیستم‌های تشخیص فعالیت ارائه شده چهار مرحله را دنبال می‌کردند در مرحله‌ی پیش‌پردازش، بر روی کاهش محدوده‌ی مطلوب مورد بررسی با استفاده از تفریق پس زمینه و شار نوری اقداماتی انجام شده است. تعداد زیادی از روش‌های مدل‌سازی پس‌زمینه توسط کریستانی و همکارانش در سال ۲۰۱۰ و الهاییان در سال ۲۰۰۸ توسعه داده شده است [۷]. یافتن مقادیر جابجایی بین فریم‌های متوالی از یک توالی تصویر، شار نوری نامیده می‌شوند. اولین روش برای محاسبه‌ی شار نوری متغیر در توالی تصویر توسط هرن و همکاران در سال ۱۹۸۱ معرفی شده است [۸]. حساس‌ترین و مهم‌ترین بخش سیستم، آشکارسازی انسان و تخمین وضعیت بدن می‌باشد. که این بخش خود به دو بخش اصلی تقسیم می‌شوند متداول‌ترین این روش‌ها، روش‌های مبتنی بر اجزا و تجزیه و تحلیل پنجره‌ی شناسایی منفرد است [۹]. در مرحله‌ی آخر هر دونوع روش‌های تشخیص عمل و فعالیت می‌تواند در دو دسته روش‌های تک لایه و روش‌های

مناسب، بسیار زمان‌بر و حافظه‌بر می‌باشد. همین‌طور در برخی از فعالیت‌ها مانند داوری در بازی‌های ورزشی به اطلاعات صحنه و حرکت دست یا پا به صورت هم‌زمان نیاز داریم که با استفاده از روش‌های پیشین مانند حذف پس‌زمینه و استخراج جداگانه وضعیت‌های مختلف بدن به این اطلاعات مهم دست نیافتیم.

## ۲ روش پیشنهادی

روش‌های ارائه شده در گذشته، نیازمند تخمین پس‌زمینه و آشکارسازی پیش‌زمینه بوده‌اند. در این حالت مجموع مراحل پیش‌پردازش، آشکارسازی پیش‌زمینه و در نهایت انجام الگوریتم دسته‌بندی مناسب، بسیار زمان‌بر و حافظه‌بر می‌باشد. لذا ارائه‌ی روشی که نیازمند استخراج جداگانه‌ی وضعیت‌های مختلف بدن در فریم‌های مختلف نباشد و اطلاعات پس‌زمینه و پیش‌زمینه را از هم جدا نکند می‌تواند به بهبود عملکرد سیستم و کاهش زمان مورد نیاز منجر شود. از طرفی جدا نکردن پس‌زمینه از پیش‌زمینه حین استخراج ویژگی در داوری بازی‌های ورزشی نیز مفید عمل می‌کند. این نیاز و برخی شرایط نامطلوب که در بالا به آن اشاره شد ما را بر آن داشت تا یک توصیف‌گر یکپارچه ویدیویی طراحی کنیم که اطلاعات صحنه و حرکت را به صورت هم‌زمان استخراج کرده و سپس با اعمال انتخاب ویژگی و حذف پس‌زمینه را به طرز مطلوبی کاهش دهیم با توجه به اینکه دقت مان نیز کاهش نیابد، که در این مرحله ما با افزایش دقت نیز مواجه شدیم که این نقطه‌ی تمایز روش پیشنهادی را نشان می‌دهد. در این توصیف‌گر، ما مراحل زیر را خواهیم داشت که در شکل ۱؟ چهارچوب روش پیشنهادی نشان داده شده است.

در ابتدا مجموعه فریم‌های ویدیویی سه بعدی را برش می‌دهیم و از سه برش در آزمایشات استفاده می‌کنیم. برای محاسبات کمتر و اینکه انتخاب تعداد بیشتر از سه برش در بهبود بیشتر کارایی تأثیری ندارند. برش فیلم‌های ویدیویی را به حوزه فرکانس می‌بریم زیرا تشخیص فعالیت در حوزه‌ی فرکانس راحت‌تر است. در این جا فرض بر آن است که دوربین در یک مکان ثابت قرار گرفته است. بنابراین اطلاعات پس‌زمینه در تمام فریم‌های ویدئو یکسان است. حال چون کانولوشن در حوزه اولیه<sup>۱</sup> معادل ضرب در حوزه‌ی فرکانس است و چون اطلاعات پس‌زمینه را داریم به راحتی می‌توانیم حرکت را در حوزه‌ی فرکانس از اطلاعات پس‌زمینه جدا کنیم. برای بردن مجموعه فریم‌های ویدئو به حوزه‌ی فرکانس از تبدیل فوریه‌ی گسسته استفاده کردیم. تبدیل فوریه‌ی سه بعدی  $f(x, y, t)$  بر روی فضا و زمان به صورت زیر محاسبه می‌شود:

$$F(f_x, f_y, f_t) = \frac{1}{MNT} \sum_{x=0}^M \sum_{y=0}^N \sum_{t=0}^T f(x, y, t) e^{-2\pi i \left( \frac{x f_x}{M} + \frac{y f_y}{N} + \frac{t f_t}{T} \right)} \quad (۱)$$

$M$ ،  $N$  و  $T$  به ترتیب عرض، ارتفاع و مدت زمان برش ویدئو می‌باشد و  $x$  و  $y$  و  $t$  موقعیت‌های فضایی و زمانی هر نقطه در حجم ایجاد شده می‌باشد. برای استخراج ویژگی از بانک فیلتر گابور سه بعدی استفاده کردیم. فیلتر گابور از ساختار چشم انسان الگوبرداری شده است و یک توصیف‌گر مناسب جهت حرکت می‌باشد. این فیلتر لبه‌ها را در جهات مختلف یافته و یک شی و یا حرکات مربوط به آن را به خوبی توصیف می‌کند که نمایش فیلترها را در شکل (۲) مشاهده می‌کنیم.

تابع انتقال هر فیلتر سه بعدی، مطابق با یک فرکانس فضایی  $f_{r0}$  در امتداد مشخص شده با زوایای قطبی  $\theta_0$  و سمتی  $\phi_0$  و در یک سیستم مختصات کروی

می‌تواند به صورت زیر بیان شود:

$$G(f_r, \theta, \phi) = \exp\left\{-\frac{(f_r - f_{r0})^2}{2\sigma_r^2} - \frac{(\theta - \theta_0)^2}{2\sigma_\theta^2} - \frac{(\phi - \phi_0)^2}{2\sigma_\phi^2}\right\} \quad (۲)$$

که

$$\phi = \arccos\left(\frac{f_z}{\sqrt{f_x^2 + f_y^2 + f_z^2}}\right), \theta = \arctan\left(\frac{f_y}{f_x}\right), f_r = \sqrt{f_x^2 + f_y^2 + f_z^2}$$

است. پارامترهای  $\sigma_r$ ،  $\sigma_\theta$  و  $\sigma_\phi$  به ترتیب پهنای باند شعاعی و زاویه‌ای هستند که کشیدگی فیلتر را در دامنه‌ی فرکانسی فضایی-زمانی تعیین می‌کنند. تأثیر فیلترسازی طیف فرکانسی را در شکل (۳) مشاهده می‌کنید:

در مرحله‌ی چهارم اعمال تبدیل فوریه‌ی گسسته سه بعدی معکوس، ویژگی‌های استخراج شده در حوزه‌ی فرکانس هستند و باید به حوزه‌ی اولیه بازگردند. (۳)

$$\sum_{u=0}^{M-1} \sum_{V=0}^{N-1} \sum_{t=0}^{T-1} F(f_x, f_y, f_t) e^{-2\pi i \left( \frac{x f_x}{M} + \frac{y f_y}{N} + \frac{t f_t}{T} \right)} F(x, y, t) = \frac{1}{MNT} \sum_{x=0}^M \sum_{y=0}^N \sum_{t=0}^T f(x, y, t) \quad (۳)$$

در مرحله‌ی پنجم از زیرحجم‌های تولید شده در هر برش میانگین گرفته و آنها را به مرحله‌ی بعد (الحاق) می‌فرستیم. زیرا به اطلاعات زیرحجم‌ها در هر برش نیاز داریم تا با کمک آنها به سیستم مان آموزش دهیم.

در مرحله‌ی ششم میانگین زیرحجم‌های بدست آمده از مرحله قبل به هم الحاق می‌شوند. در مرحله‌ی بعد نیاز است طول بردارهای ویژگی تولید شده کاهش یابند. در این مرحله چون طول بردار ویژگی تولید شده به ۱۰۴،۴۴۸ می‌رسد بنابراین نیاز به کاهش بعد خواهیم داشت. برای کاهش ابعاد بردارهای ویژگی بدست آمده، ابتدا PCA<sup>۲</sup> اعمال کردیم ولی کاهش ابعاد با این روش ما را اقناع نکرد بنابراین از انتخاب ویژگی MRMR استفاده کردیم سپس به MRMR اعمال شده PCA اعمال کردیم که ببینیم آیا به ابعاد کمتر با دقت بالاتر می‌رسیم؟ پاسخ پرسش‌های مطرح شده را قسمت نتایج آزمایشات بررسی خواهیم کرد.

در روش MRMR انتخاب ویژگی‌ها بر اساس ارتباط میان ویژگی و مفهوم (یا بر چسب)  $c$  تعریف می‌گردد. در «بیشترین ارتباط»<sup>۳</sup> سعی بر تقریب بیشه‌ای از ارتباط میان ویژگی‌های انتخاب شده و بر چسب است که این دقیقاً همان چیزی است که «بیشترین وابستگی»<sup>۴</sup> هم به دنبال آن بود. تفاوت این دو در نحوه‌ی محاسبه‌ی  $D$  است، به طوریکه در این روش  $D$  به صورت میانگین همه‌ی اطلاعات مشترک میان ویژگی‌های  $X_i$  و بر چسب کلاس  $C$  تعریف می‌گردد. این روابط به فرم ریاضی ذیل تعریف شده است:

$$\max D(S, c), \quad D = \frac{1}{|S|} \sum_{X_i \in S} I(X_i; c) \quad (۴)$$

حال اگرچه مرتبط ترین ویژگی‌ها بدین طریق بدست می‌آیند اما ممکن است دو ویژگی وجود داشته باشند که وابستگی شدیدی به هم داشته باشند و قدرت دسته‌بندی در صورت حذف یکی از آنها هیچ تغییری نکند و یا اینکه تغییری بسیار جزئی و قابل چشم‌پوشی داشته باشد. در اینجا فاز دوم یعنی کمینه‌سازی افزودگی<sup>۵</sup> به فاز قبلی اضافه می‌گردد تا بتوان ویژگی‌هایی را یافت که علاوه بر ارتباط نزدیک آنها با مفهوم مورد نظر ( $c$ )، در مقابل یکدیگر دارای انحصار متقابل باشند، یعنی

## پانویس ها

<sup>1</sup> Spatial

<sup>2</sup> Principal Component Analysis

<sup>3</sup> Max-Relevance

<sup>4</sup> Max-Dependency

<sup>5</sup> Mutual Information-based Feature Selection

<sup>6</sup> Feature Selection

<sup>7</sup> Joint Mutual Information

<sup>8</sup> Mutual Information-based Feature Selection

ارتباط وابستگی شدیدی با یکدیگر نداشته باشند به عبارت ریاضی می توان گفت:

$$\min R(S), \quad R = \frac{1}{|S|^2} \sum_{X_i, X_j \in S} I(x_i; x_j) \quad (5)$$

دو محدودیت بالا در جوار هم معیاری به نام حداکثر ارتباط و حداقل افزونگی را تعریف می کند که به اختصار MRMR نامیده می شود عملگر  $\varphi(D, R)$  جهت ترکیب دو فاز مطرح شده به کار می رود که هدف آن بهینه سازی  $R, D$  به صورت همزمان است و به فرم ذیل تعریف می گردد:

$$\max \Phi = (D, R), \Phi = D - R \quad (6)$$

یکی از محاسن بارز روش MRMR نسبت به «بیشترین وابستگی» آنست که در MRMR از تخمین چگالی احتمال چند گانه  $p(x_1, x_2, \dots, x_m)$  و همچنین  $p(x_1, x_2, \dots, x_m, c)$  اجتناب می گردد و در عوض چگالی احتمال توام  $p(x_i, x_j)$  و همچنین  $p(x_i, c)$  محاسبه می گردد که کاری بس ساده تر و ممکن تر است. در مرحله ی آخر با SVM سلسله مراتبی دسته بندی کرده و تشخیص رفتار خواهیم داشت. به هر داده در مرحله ی اول اعمال SVM حداقل یک برچسب زدیم سپس در مرحله ی دوم از احتمالاتی که در مرحله ی اول به داده های برچسب خورده نسبت داده بودیم استفاده کردیم به عنوان ورودی های جدید به SVM مرحله دوم که در این مرحله به داده های برچسب نخورده مرحله اول برچسب بزنیم.

## ۳ نتیجه گیری و کارهای آینده

همواره طراحی سیستمی که بدور خستگی، اشتباه، کوتاهی و با قدرت پردازش مطلوب همزمان باشد و بتواند فعالیت های انسان ها را در محیط های حساس تحت نظارت کاملاً هوشمند قرار دهد و فعالیت های غیرعادی را از عادی تشخیص دهد و نیروهای حفاظتی را آگاه نماید می تواند امنیت بهتری را برای انسان ها به همراه بیاورد از این رو ما در این مقاله یک توصیف گر یکپارچه برای ویدئو به منظور استفاده در تشخیص فعالیت غیرعادی انسان معرفی کردیم که در آن پس از برش فیلم ها به آنها تبدیل فوریه اعمال کردیم تا اطلاعات حرکت را به حوزه ی فرکانس برده و در آنجا به استخراج اطلاعات حرکت پردازیم به طیف فرکانسی تولید شده فیلتر گاپور اعمال کرده و استخراج ویژگی انجام دادیم. سپس کاهش بعد با انتخاب ویژگی MRMR و تبدیل فضا و آموزش با SVM و تشخیص رفتار را داشتیم. با توجه به نتایج بدست آمده MRMR مقام نخست را از نظر میزان دقت در مقابل کاهش بعد داشته است. در میان پیش پردازش در داده ها، می توان به انتخاب زیر مجموعه ویژگی و انتخاب ویژگی،<sup>۶</sup> اشاره کرد. که در میان روش های انتخاب ویژگی، MRMR که شالوده ی این مقاله را تشکیل می داد، میتوان به سایر روش های انتخاب ویژگی مانند JMI<sup>۷</sup> و MIFS<sup>۸</sup> و ترکیب MIFS & MRMR و اعمال دوباره ی PCA به هر کدام از آنها اشاره کرد. که اعمال هر کدام از آنها می تواند منجر به افزایش دقت و سرعت تشخیص و کاهش بعد بیشتر بشود که برای کارهای آینده توصیه می شود. در مرحله ی یادگیری ما از SVM استفاده کردیم ولی اگر یک یادگیری بتواند با دقت بهتری داده ها را دسته بندی کند نیز می تواند مطلوب باشد. سایر الگوریتم های یادگیر که می توان برای افزایش دقت دسته بندی پیشنهاد کرد: روش های مبتنی بر درخت تصمیم، روش های مبتنی بر قانون، استدلال مبتنی بر حافظه، شبکه های عصبی، روش های مبتنی بر نظریه ییز، ماشین های بردار پشتیبان. از طرفی برای دسته بندی بهتر که هیچ داده ای بی برچسب نماند ما از ترکیب دسته بندی کننده ها استفاده کردیم که یکی از اهداف این مقاله را تشکیل می داد برای این منظور از SVM سلسله مراتبی در تمام مراحل استفاده کردیم. در اینجا می توان پس از اعمال مرحله ی اول SVM در مراحل بعدی از سایر الگوریتم های دسته بندی مانند درخت تصمیم یا شبکه های عصبی یا روش های مبتنی بر قانون و... نیز استفاده کرد. همچنین از جنبه های نوآوری سیستم پیشنهادی می توان به کاهش بعد با انتخاب ویژگی MRMR و SVM سلسله مراتبی و جدا نکردن پس زمینه از پیش زمینه حین استخراج ویژگی اشاره کرد.