

# ویژگی‌های نهفته در فضای پیچش زمانی پویا برای بازشناسی کنش با استفاده از داده‌های سنسور کینکت

نرجس الهدی محمدزاده<sup>۱</sup> و محسن تابع‌جماعت<sup>۲</sup>

## چکیده

هدف از این مقاله، بازشناسی کنش افراد با استفاده از اطلاعات سری‌های زمانی استخراج شده از دنباله‌های اسلکتی به منظور استفاده در سامانه‌های مانیتورینگ فعالیت‌های روزمره‌ی انسانها می‌باشد. به این منظور، هر کنش به صورت یک سری زمانی چند بعدی بیان شده و سپس با استفاده از مفهوم "شبه کرنل مبتنی بر فاصله‌ی پیچش زمانی پویا" به یک فضای برداری نگاشت می‌گردد. در ادامه، به منظور استفاده از نسبت همبستگی-تمایز دنباله‌ها در پروسه‌ی شناسایی، این فضای برداری توسط روش فیشر تنظیم‌شونده به یک فضای تمایزی نگاشت شده و تصمیم‌گیری نهایی در خصوص محتوای حرکت در فضای حاصل انجام می‌پذیرد. بر خلاف سایر روش‌های کرنلی موجود، الگوی همترازی حاصل از پیچش زمانی، موجب می‌شود تا اثر شیفت، و انقباض و انبساط‌های زمانی دنباله‌ها در فضای کرنل به کمترین میزان ممکن کاهش یابد. همچنین، روش ما پیچیدگی‌های محاسباتی و محتوایی موجود در استخراج ویژگی-های استاتیک و دینامیک دنباله‌های حرکتی را حذف نموده و در مقابل، آنها را در قالب الگوی همترازی در فضای برداری کرنل مورد استفاده قرار می‌دهد. نتایج ارزیابی‌ها بر روی سه پایگاه داده‌ی معروف TST، UTKinect و UCFKinect، قابل رقابت بودن عملکرد روش ارائه شده با برترین روشهای موجود در بازشناسی کنشهای انسانی را نشان می‌دهد.

## کلیدواژه‌ها

شناسایی کنش، پیچش زمانی پویا، حقه کرنل، بیان ویژگی نهفته

توان به مانیتورینگ افراد بیمار و سالخورده، نظارت در مکانهای عمومی و سرگرمی اشاره نمود. به طور کلی، از نظر پیچیدگی می‌توان یک عمل را به یکی از سه سطح ژست<sup>۱</sup>، کنش و رفتار<sup>۲</sup> طبقه‌بندی کرد که در این مقاله به سطح میانی یعنی کنش که دنباله‌ای در حد چندین ثانیه از چند ژست است، پرداخته می‌شود. به عنوان مثال به اعمالی از قبیل نشستن، برداشتن شیء، افتادن بر روی زمین کنش اطلاق می‌شود.

در تلاشهای اولیه برای بازشناسی کنش<sup>۳</sup> از ویدئوهای RGB استفاده می‌شد که مزیت سهولت و سرعت در جمع‌آوری داده را

## ۱ مقدمه

بازشناسی کنش انسان یکی از مسائل پژوهشی مهم در بینایی ماشین محسوب می‌شود که دارای کاربردهای بالقوه متعددی در زمینه‌های مختلف از جمله سلامت، امنیت و تعامل انسان با ماشین می‌باشد. از میان کاربردهای متفاوت در این زمینه‌ها می‌

این مقاله در خردادماه ۱۳۹۸ دریافت، در آبان‌ماه ۱۳۹۹ بازنگری و در بهمن‌ماه همان سال پذیرفته شد.

<sup>۱</sup> دانشکده مهندسی برق، دانشگاه صنعتی شریف

رایانامه: [hoda@sharif.edu](mailto:hoda@sharif.edu)

<sup>۲</sup> دانش‌آموخته کارشناسی مهندسی برق، دانشگاه آزاد اسلامی

رایانامه: [m.tabejamaat@sharif.edu](mailto:m.tabejamaat@sharif.edu)

<sup>۱</sup>Gesture  
<sup>۲</sup>Behavior

پروسه‌ی حرکت می‌باشد [۱۶][۱۷][۴۰]. در این میان مسلماً روشی مطلوب است که بتواند همزمان شرایط زیر را برآورده کند: (الف) کمترین وابستگی را به پیش پردازش داده داشته باشد، (ب) پیچیدگی محاسباتی مرحله‌ی استخراج ویژگی آن تا حد امکان کوچک باشد، (ج) تغییرات فضا-زمانی<sup>۱۱</sup> دنباله‌های حرکتی را به صورت همزمان در پروسه‌ی شناسایی لحاظ نماید (د) شیفت زمانی دنباله‌های حرکتی را در نظر بگیرد، و (ه) نسبت همبستگی به تمایز دنباله‌ها را در پروسه‌ی تصمیم‌گیری لحاظ نماید. متأسفانه روش‌های ارائه شده تنها روی یک یا بخشی از شرایط فوق تمرکز نموده و تمامی شرایط را به صورت همزمان برآورده نمی‌نمایند. برای مثال در روش‌های دسته‌ی نخست از نسبت همبستگی به تمایز دنباله‌ها صرف نظر شده که موجب کاهش قدرت تمایزی ویژگی‌ها می‌گردد. در روش‌های دسته‌ی دوم نحوه‌ی شناسایی فریم‌های کلیدی یک دنباله و حساسیت آن به پارامترهای آزاد سیستم، چالشی اساسی را به همراه می‌آورد. مشکل اساسی در روش‌های دسته‌ی سوم، تعریف یک معیار مناسب جهت مقایسه و سنجش زیر منیفلد‌های مختلف می‌باشد. نهایتاً، در روش‌های یادگیری عمیق، زمان بر بودن پروسه‌ی آموزش و نیاز به تعداد قابل توجهی از نمونه‌های آموزشی از چالش‌های اساسی سیستم محسوب می‌گردد. همچنین از آنجا که روش‌های یادگیری ژرف کانونلوشنی تمامی دنباله را به صورت یکجا به ورودی شبکه اعمال می‌نمایند، نمی‌توانند در کاربردهای بی‌درنگ (جاییکه هنوز یک کنش به پایان نرسیده و تمامی فریم‌های آن در دسترس نیست) مورد استفاده واقع شوند.

از آنجایی که کنش‌های مختلف می‌توانند در تمام و یا تنها در بخش‌هایی از فریم‌های حرکتی با یکدیگر متفاوت باشند، بنابراین توزیع دنباله‌های حرکتی از ساختار غیر خطی تبعیت نموده که عدم توجه به آن کاهش دقت شناسایی سیستم را به همراه خواهد داشت. روش‌های ارائه شده به منظور مقابله با این ساختار عمدتاً به دو دسته‌ی کلی تقسیم بندی می‌شوند: (الف) روش‌های مبتنی بر ویژگی‌های منیفلد و (ب) روش‌های مبتنی بر حقه‌ی کرنل<sup>۱۲</sup>. در روش‌های مبتنی بر منیفلد فرض بر آن است که داده‌ی غیر خطی با ابعاد بزرگ معمولاً روی منیفلد‌های غیر خطی با ابعاد کوچکتر می‌نشینند و بنابراین طبقه‌بندی آن در فضای غیر اقلیدسی بهتر انجام خواهد شد. هرچند چگونگی مقایسه‌ی داده‌های قرار گرفته روی منیفلد، خود چالش بزرگ این روش‌ها محسوب می‌شود. در روش‌های کرنلی، داده‌های غیر خطی با استفاده از حقه‌ی کرنل به فضایی با ابعاد بزرگتر نگاشت می‌شوند تا ساختار غیر خطی داده در این فضا به صورت خطی قابل مجزا شدن گردد. هرچند چگونگی تعیین نوع کرنل و پارامترهای آن از مسائل اساسی و چالش برانگیز این روش‌ها به

به همراه دارد. اما مشکلاتی از قبیل تداخل پس زمینه، زاویه تصویر برداری و تغییر روشنایی، پژوهشگران را به استفاده از اطلاعات سه بعدی صحنه به منظور بازشناسی محتوای کنش ترغیب نمود. مدل‌های اولیه سه بعدی بدن انسان، بر مبنای داده‌های سنسورهای پوشیدنی بود که معایب عدم راحتی، عدم امکان سنجش از راه دور و نیاز به همکاری فرد را به همراه داشت. در سال‌های اخیر ظهور و تجاری سازی سنسورهای تشخیص عمق کینکت<sup>۳</sup> با رفع مشکلات سنسورهای پوشیدنی، زمینه پیشرفت در روش‌های بازشناسی کنش با استفاده از اطلاعات سه بعدی را بوجود آورده است. به دنبال کار مهمی که در [۱] برای استخراج محل مفاصل با استفاده از تصویر عمق صورت پذیرفت، ارائه روش‌هایی برای بازشناسی کنش با استفاده از داده‌های بی‌درنگ<sup>۴</sup> اسکلت بدن انسان مورد توجه فراوان قرار گرفت. در مقایسه با تصاویر عمقی، مدل‌های اسکلتی اطلاعات فشرده و موثرتری را در اختیار کاربر قرار داده و علاوه بر آن معایبی همچون ماسکه شدن اعضای بدن توسط یکدیگر و تاثیر حجم بدن در پروسه‌ی شناسایی را نیز برطرف می‌نمایند. با حل مسئله‌ی دریافت داده، تنها چالش باقیمانده در شناسایی کنش، نحوه‌ی مدل کردن یک دنباله‌ی حرکتی است که در سال‌های اخیر در کانون توجه محققین این حوزه قرار گرفته است. روش‌های ارائه شده در این زمینه عمدتاً به چهار دسته تقسیم بندی می‌گردند: در روش‌های دسته‌ی نخست، حرکت به صورت یک یا مجموعه‌ای از سری‌های زمانی<sup>۵</sup> مدل شده و شناسایی آن با استفاده از تطبیق یک به یک این سری با سری‌های زمانی استخراج شده از نمونه‌های آموزشی انجام می‌پذیرد [۳۲][۳۳][۳۴]. در دسته‌ی دوم، کنش به صورت یک مجموعه از حالت‌های کلیدی<sup>۶</sup> اسکلت بدن (ژست اسکلتی) تعریف شده و به صورت دنباله‌ای از اندیس‌های متناظر با حالات کلیدی رمزنگاری می‌گردد [۲][۱۱][۱۵][۲۱]. اصول مشترک روش‌های دسته‌ی سوم، استفاده از ویژگی‌های منیفلد<sup>۷</sup> در فضای توزیع دنباله‌های حرکتی می‌باشد [۱۳][۲۹]. این روش‌ها سعی می‌نمایند تا با استفاده از ویژگی انحنای فضا غیر اقلیدسی، تصویر روشنتری از میزان مشابهت‌ها در مقایسه‌ی کنش‌های حرکتی را ارائه نمایند. نهایتاً مشخصه‌ی بارز روش‌های دسته‌ی چهارم استفاده از شبکه‌های عصبی کانونلوشنی<sup>۸</sup> و یا بازخوردی<sup>۹</sup> عمدتاً ژرف در مدل سازی

<sup>3</sup>Action Recognition

<sup>4</sup>Kinect Depth Sensor

<sup>5</sup>Real-Time

<sup>6</sup>Time Series

<sup>7</sup>Key-pose

<sup>8</sup>Manifold

<sup>9</sup>Deep Convolutional Neural Networks

<sup>10</sup>Deep Recurrent Neural Networks

<sup>11</sup>Spatio-Temporal

<sup>12</sup>Kernel Trick

چگونگی اجرای آن می‌پردازد. در بخش (۴) نتایج شبیه سازی الگوریتم مزبور و مقایسه‌ی آن با سایر روش های موجود ارائه می‌گردد و نهایتاً بخش (۵) به نتیجه گیری از مقاله اختصاص خواهد داشت.

## ۲ مروری بر روش های پیشین

در این بخش مروری کوتاه بر روش های برجسته موجود در زمینه ی شناسایی کنش بر اساس ویژگی های اسکلتی انسانی ارائه می‌گردد. در [۲]، Lv و همکارانش یک سیستم شناسایی کنش مبتنی بر داده های MoCap ارائه نمودند که از فاصله ی میان مفصلی به عنوان ویژگی های هر فریم اسکلت استفاده نموده و سپس ترکیبی از مدل مخفی مارکوف<sup>۱۷</sup> (HMM) و الگوریتم Adaboost چند کلاسه را جهت دسته بندی دنباله ی اسکلتی بکار می‌برد. Guo و همکارانش [۳] یک روش توصیف چرخش و سرعت اشیاء<sup>۱۸</sup> (RRV) را به منظور بیان مسیر حرکتی اعضای مختلف بدن ارائه و از روش پیشین زمانی پویا جهت خنثی نمودن شیفت های زمانی موجود در RRV های مختلف استفاده نمودند. در [۳۲][۳۳][۳۴] از فاصله پیشین زمانی پویا به عنوان یک ویژگی برای طبقه بندی بوسیله نزدیکترین همسایگی استفاده شده است. در [۴]، ماتریس کواریانس دنباله ی زمانی حاصل از تغییرات مختصات مفاصل به عنوان ویژگی های حرکتی استخراج و با استفاده از ماشین بردار پشتیبان با کرنل خطی دسته بندی می‌شد. آنها همچنین از یک روش سلسله مراتبی به منظور استخراج بردار ویژگی های مبتنی بر کواریانس استفاده نمودند تا ترتیب زمانی حرکات را در روند شناسایی سیستم لحاظ نمایند. Lu و همکارانش [۵] از هیستوگرام کلمات رمز استخراج شده از میزان جابجایی فریمها به منظور توصیف یک دنباله ی حرکتی و از الگوریتم دسته بندی بیز به منظور طبقه بندی توصیفگرها استفاده نمودند. در مقاله ی [۶] ابتدا مدل اسکلتی به صفحات  $xy$ ،  $yz$  و  $xz$  افکنش می‌شود و سپس هیستوگرام جابجایی مفاصل به عنوان ویژگی، مورد استفاده قرار می‌گیرد. در [۷]، از دو توصیفگر دنباله ی مختصات های اسکلتی و دنباله ی زوایای اسکلتی که با استفاده از مشتق خط سیر حرکتی اعضای مختلف بدست می‌آیند جهت توصیف کنش افراد استفاده می‌گردد. Wang و همکارانش [۸]، هر کنش را با استفاده از هیستوگرام اندیس های فضا-زمانی بخش های مختلف حرکتی توصیف نموده و از ماشین بردار پشتیبان<sup>۱۹</sup> جهت دسته بندی بردارهای هیستوگرام استفاده می‌نمایند. نویسندگان مقاله ی [۹] بر اهمیت متفاوت هر مفصل در شناسایی کنش های متفاوت تمرکز نموده و نشان دادند که با انتخاب مفاصلی با بیشترین اهمیت، دقت شناسایی روش به میزان قابل توجهی افزایش می‌یابد. انتخاب

شمار می‌رود. علاوه بر آن در روش های کرنلی موجود، ابتدا نیاز است تا سری های زمانی به داده های یک بعدی تبدیل شده و سپس حقه ی کرنل، روی بردار های حاصل اعمال گردد. این روش سبب حساسیت قابل توجه روش های کرنلی نسبت به شیفت زمانی دنباله های حرکتی خواهد شد. از طرف دیگر تبدیل سری های زمانی به بردارهای یک بعدی، عدم قابلیت روشهای کرنلی در شناسایی بی‌درنگ دنباله های حرکتی را به دنبال خواهد داشت، زیرا بردار های آموزشی تک بعدی از کنش های تکمیل شده ی آموزشی، استخراج و بعد مکانی این بردارها در طی فرآیند استخراج حذف شده است. بنابراین مشابهت نمونه ی آزمون با بردارهای استخراج شده، تنها در گرو استخراج بردار ویژگی پس از تکمیل حرکت مذکور خواهد بود.

پیشین زمانی پویا<sup>۱۳</sup> یکی از روشهای متداول در همتراز نمودن<sup>۱۴</sup> سری های زمانی می‌باشد که در کاربرد های شناسایی کنش افراد نیز مورد استفاده قرار گرفته است. با این حال، این روش، ساختار غیر خطی نسبت همبستگی به تمایز دنباله های آموزشی را در نظر نگرفته و بنابراین استفاده از آن به منظور دسته بندی دنباله های حرکتی کارایی مناسبی را به همراه نخواهد داشت.

با در نظر داشتن معایب فوق، مقاله ی حاضر یک روش شناسایی کنش با استفاده از ویژگی های مبتنی بر فاصله ی پیشین زمانی پویا را ارائه می‌نماید. به این منظور، ابتدا دنباله های حرکتی توسط یک کرنل مبتنی بر فاصله ی پیشین زمانی به یک فضای برداری بهینه انتقال می‌یابند. این عمل ابعاد سری های زمانی را به طور موثری کاهش داده و همزمان اثر شیفت و انقباض و انبساط های زمانی در دنباله های مشابه را نیز حذف می‌نماید. به علاوه ویژگیهای استخراج شده به گونه ای است که می‌توان از آنها در آموزش طبقه بند و یادگیری ویژگیهای تمایزی استفاده کرد. از آنجا که روش ارائه شده تنها یکی از دنباله های موجود در هر کلاس را به عنوان نمونه ی مرجع کلاس در محاسبه ی بردار کرنل مورد استفاده قرار می‌دهد، پیچیدگی محاسباتی آن به شدت کاهش یافته و در شرایطی می‌تواند در کاربردهایی درنگ نیز به کار گرفته شود. در ادامه، از روش تحلیل افتراقی فیشر تنظیم شونده<sup>۱۵</sup> برای افکنش مجدد فضای برداری به یک فضای تمایزی استفاده می‌شود. این عمل سبب می‌شود تا نسبت همبستگی به تمایز دنباله ها در پروسه ی تصمیم گیری لحاظ شده و علاوه بر آن تصمیم گیری در ابعادی به مراتب کوچکتر از ابعاد فضای برداری انجام می‌شود که خود موجبات تقلیل اثر نفرین ابعاد<sup>۱۶</sup> موجود در فضای برداری را فراهم می‌آورد.

در ادامه بخش های ارائه شده ی مقاله به ترتیب زیر خواهند بود: در بخش (۲) مروری بر کارهای انجام شده در زمینه شناسایی کنش افراد ارائه می‌گردد. بخش (۳) به معرفی روش پیشنهادی و

<sup>17</sup>Hidden Markov Model

<sup>18</sup>Rotation and Relative Velocity

<sup>19</sup>Support Vector Machine

<sup>13</sup>Dynamic Time Warping

<sup>14</sup>Alignment

<sup>15</sup>Regularized Fisher Discriminant Analysis

<sup>16</sup>Curse of Dimensionality

همچون نویز و انسداد از دست رفته‌اند، استفاده کرده‌اند. این قیدها یکی ثابت بودن طول استخوانها و دیگری محدودیت جابجایی هر استخوان حول مفصلش است. در [۴۴]، Papadopoulis و همکاران، از روش کیف لغات برای تشخیص کنش استفاده کرده‌اند که در آن لغت‌ها از خوشه‌بندی مسیره‌های حرکت مفاصل بدست آمده‌اند. موفقیت استفاده از شبکه‌های عصبی کانولوشنی و بازخوردی عمدتاً ژرف در شناسایی محتوای ویدئو سبب شد تا این شبکه‌ها به طور گسترده در شناسایی کنش افراد از طریق ویژگی‌های اسکلتی نیز بکار گرفته شوند. در این زمینه Du و همکارانش [۱۶] از یک روش سلسله مراتبی مبتنی بر شبکه‌های عصبی بازخوردی (RNN) به منظور یادگیری حالت‌های اسکلتی بدن استفاده می‌نمایند. در این روش اسکلت بدن به ۵ عضو تقسیم بندی شده و مجموعه‌ای از شبکه‌های RNN موجود در لایه‌های مختلف یک ابر شبکه، ویژگی‌های محلی، متقابل و جامع این اعضا را به صورت همزمان آنالیز می‌نمایند. در [۱۷]، Li و همکاران، از یک شبکه‌ی عصبی کانولوشنی به منظور استخراج الگوهای هم‌رخدادی در سیگنالهای مختصات مفاصل استفاده می‌شود. همانند مقاله‌ی [۱۶]، نویسندگان این مقاله نیز سعی نموده‌اند تا از یک روش سلسله مراتبی در یادگیری ویژگی‌های مفاصل استفاده نمایند. بر این اساس ویژگی‌های تغییرات هر مفصل در لایه‌ی اول و ویژگی‌های برهم نهی مفاصل در لایه‌های بعدی شبکه، رمز نگاری شده و مورد استفاده قرار می‌گیرد. در [۴۱]، Naveenkumar و همکاران از سه زیر شبکه ژرف برای استخراج ویژگی‌های دینامیکی زمانی و مکانی که مربوط به موقعیت‌های اسکلت در هر فریم و تغییرات آن در طول دنباله است، استفاده کرده‌اند. در آخر این سه شبکه برای تشخیص کنش در قالب یادگیری تجمعی به-کار رفته‌اند. در [۴۵]، Cho و همکاران، از شبکه‌های خودتوجه<sup>۲۵</sup> برای استخراج همبستگی‌های بلند مدت در مسیر موقعیت مفاصل استفاده کرده‌اند. نویسندگان ادعا می‌کنند که این همبستگی‌ها با استفاده از شبکه‌های بازگشتی معمولی قابل استخراج نیستند. در [۴۶]، Liu و همکاران، از شبکه‌های کانولوشنی برای استخراج ویژگی‌های مکانی در هر فریم و سپس در سطح بالاتر از روش فوریه‌ی همی برای استخراج ویژگی‌های زمانی دنباله‌های اسکلت بهره می‌برند.

### ۳ فاصله پیچش زمانی پویا به عنوان تابع کرنل و چالش بار محاسباتی

#### ۳-۱- پیچش زمانی پویا

روش پیچش زمانی پویا [۳۱] روشی برای همتراز کردن دو سیگنال زمانی است. هدف این همترازی آن است که دو سیگنال

مفاصل بر اساس واریانس زوایای بین مفصلی و یا ماکزیمم سرعت زاویه‌ای آنها انجام می‌پذیرد. Weng و همکارانش [۱۰]، ابتدا هر دنباله‌ی حرکتی را به  $n$  مرحله تقسیم بندی می‌نمایند. سپس فاصله‌ی میان توصیفگر هر مفصل در آن مرحله تا هر یک از کلاسها محاسبه می‌گردد. نهایتاً ماتریس فضا-زمانی این فواصل برای هر مفصل در طول بازه‌ی زمانی هر مرحله محاسبه و به عنوان توصیفگر سری زمانی کنش مربوطه به کار می‌رود. در [۱۱]، هر کنش به عنوان دنباله‌ای از فریم‌های کلیدی که توسط آموزش یک ماشین بردار پشتیبان نهفته بدست می‌آیند بیان می‌گردد. در این روش موقعیت فریم‌های کلیدی به عنوان متغیر نهفته در ماشین بردار پشتیبان محسوب می‌گردد. Evangelidis و همکارانش در [۱۲] موقعیت‌های نسبی دسته‌های چهارتایی از مفاصل با نام Quad را با استفاده از بیان سلسله مراتبی بردارهای فیشر<sup>۲۰</sup> (FV) این دسته‌ها کد نموده و از ماشین بردار پشتیبان با کرنل خطی به منظور دسته بندی بردارهای فیشر استخراج شده استفاده می‌نمایند. Vemulapalli و همکارانش [۱۳]، چرخشها و انتقال‌های بخشهای مختلف اسکلت در فضای اقلیدسی سه بعدی را به عنوان یک نقطه در گروه Lie معرفی نمودند. از آنجا که گروه Lie یک منیفلد خمیده را تشکیل می‌دهد، مقایسه‌ی دنباله‌های قرار گرفته روی این منیفلد<sup>۲۱</sup> توسط انطباق ویژگیها در فضای تانژانت منیفلد صورت می‌پذیرد. در [۱۴]، Yang و همکارانش ترکیبی از ویژگی‌های حالات ایستا، حرکات جزئی و حرکات کلی اسکلت بدن افراد را به عنوان مشخصه‌های یک دنباله حرکتی استخراج می‌نمایند. سپس با اعمال روش تحلیل مولفه‌های اساسی (PCA) روی بردارهای ترکیبی حاصل، مجموعه‌ای از ویژگی‌های تمایزی با نام EigenJoints را استخراج و از الگوریتم نزدیکترین همسایه بی‌زی غیر پارامتری<sup>۲۲</sup> (NBNN) به منظور دسته بندی ویژگیها استفاده می‌نمایند. نویسندگان مقاله‌ی [۱۵]، از هیستوگرام توزیع کروی مفاصل به عنوان ویژگی فریم‌های حرکت استفاده نموده و پس از اعمال روش تحلیل افتراقی خطی<sup>۲۳</sup> (LDA) از خوشه بندی ویژگیها به منظور رمزنگاری ویژگیهای حاصل بهره می‌گیرند. در نهایت مدل HMM جهت مدل سازی پروسه‌ی حرکت استفاده می‌گردد. در [۴۲]، تابع جماعت و همکاران، برای حل مشکل تشخیص کنش در حالت برخط (در حین انجام کنش، وقتی هنوز اجرای کنش به پایان نرسیده) از روش تشریح مساعی<sup>۲۴</sup> برای بیان هر فریم از روی فریم‌های نمونه‌های آموزشی استفاده کرده‌اند. این روش همچنین در مواقعی که بعضی از فریم‌ها از دست رفته‌اند، کاربرد دارد. در [۴۳]، Nie و همکاران، از دو قید برای بازیابی موقعیت مفاصل هنگامی که به دلایلی

<sup>20</sup>Fisher Vectors

<sup>21</sup>Manifold

<sup>22</sup>Non-parametric Bayesian Nearest Neighbor

<sup>23</sup>Linear Discriminant Analysis

<sup>24</sup>Collaborative Representation

<sup>25</sup>Self-attention

را حداقل نماید که در آن  $d$  یک تابع فاصله برای بردارهای  $\vec{x}_i$  و  $\vec{y}_k$  می‌باشد.

در تلاشهایی که در گذشته برای استفاده از پیچش زمانی پویا برای بازشناسی کنش صورت گرفته [۳۲][۳۳][۳۴]، عمدتاً از طبقه‌بند نزدیکترین همسایه استفاده شده است. به بیانی دیگر کارهای گذشته از فاصله پیچش زمانی مستقیماً برای بازشناسی کنش استفاده کرده‌اند و قادر نبوده‌اند ویژگیهای نهفته در فضای پیچش زمانی را برای آموزش طبقه‌بند های پیشرفته استخراج کنند. استخراج یا بکارگیری این ویژگیها برای آموزش یک طبقه‌بند، این مزیت مهم را به همراه دارد که طبقه‌بند می‌تواند تفاوت‌های بین کلاسی را از شباهت‌های میان کلاسی کنش‌ها متمایز نماید و در نتیجه بازشناسی را با دقت بالاتری انجام دهد. از طرف دیگر طبقه‌بند نزدیکترین همسایه برای اینکه بهترین عملکرد خود را داشته باشد نیاز دارد که فاصله نمونه تست را با تمام نمونه‌های آموزشی محاسبه نماید که با توجه به زمانبر بودن الگوریتم پیچش زمانی پویا، بار محاسباتی بسیار سنگینی را ایجاد می‌نماید. در ادامه ما به مرور روشهای کرنلی کاهش بعد می-پردازیم تا راه حلی برای بکارگیری ویژگیهایی که در فضای پیچش زمانی پویا قرار دارند ارائه دهیم.

### ۳-۲- استفاده از ماتریس کرنل برای کاهش بعد

در روش معروف تحلیل مولفه‌های اساسی [۳۵] از ماتریس کواریانس داده‌ها، برای یافتن راستاهای بهینه برای کاهش ابعاد استفاده می‌شود. اگر ماتریس داده‌های آموزشی را به صورت

$$\mathbf{X} = [\vec{x}^{(1)} \dots \vec{x}^{(N)}]^T \quad (4)$$

در نظر بگیریم که در آن نمونه‌ها در سطرهای ماتریس قرار دارند ( $N$  تعداد نمونه‌های آموزشی)، ماتریس کواریانس را می‌توان به صورت

$$\mathbf{C} = (\mathbf{X} - \bar{\mathbf{X}})^T (\mathbf{X} - \bar{\mathbf{X}}) \quad (5)$$

نوشت، که در آن  $\bar{\mathbf{X}}$  ماتریسی است که میانگین نمونه‌ها در تمام سطرهای آن تکرار شده است. با حل یک مساله بهینه سازی ساده نشان داده می‌شود که بردارهای ویژه ماتریس کواریانس، راستاهای متعامدی هستند که بیشینه واریانس داده‌ها را در بردارند. در نتیجه با تصویر کردن داده‌ها در این راستاها کاهش بعد بهینه از نظر معیار کمینه مجموع مربعات خطای بازسازی حاصل می‌شود. به عبارتی دیگر، اگر ستونهای ماتریس

$$\mathbf{U} = [\vec{u}_1 \dots \vec{u}_k] \quad (6)$$

در بردارنده  $k$  بردار ویژه اول (از نظر بزرگی مقادیر ویژه متناظر) ماتریس کواریانس باشند، ویژگیهای کاهش بعد یافته برای نمونه  $\vec{x}_i$  به صورت

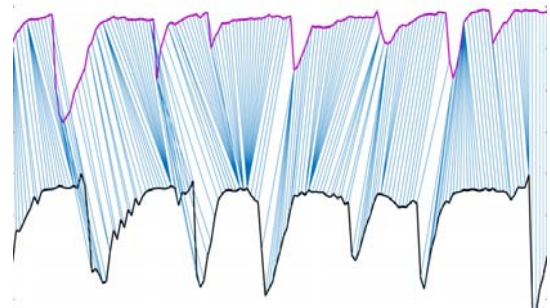
$$\vec{z} = \mathbf{U}^T \vec{x} \quad (7)$$

بدست آورده می‌شود.

که دارای الگوی های مشابهی از تغییرات در طول زمان هستند ولی این تغییرات با سرعت‌های متفاوت و یا تاخیرهای متفاوتی انجام شده‌اند، تطابق نسبتاً کاملی با یکدیگر پیدا کنند. برای بدست آمدن این تطابق هر کدام از سیگنالها ممکن است دچار انبساط‌ها یا انقباض‌های محلی متفاوتی شوند. این روش با استفاده از روش برنامه ریزی پویا، یک تناظر غیر خطی بین اندیس‌های زمانی دو سیگنال فراهم می‌آورد، به طوری که مجموع فواصل بین مقادیر دو سیگنال در اندیس‌های متناظر حداقل شود. شکل ۱ مثالی از همترازی دو سیگنال را نشان می‌دهد.

روش پیچش زمانی پویا را می‌توان به راحتی برای سریهای زمانی چند بعدی<sup>۲۶</sup> تعمیم داد. در این حالت هزینه ای که باید حداقل شود مجموع فواصل بین مقادیر چند بعدی در اندیس‌های متناظر خواهد بود. اگر  $\mathbf{X}$  و  $\mathbf{Y}$  دو سری زمانی  $p$  بعدی به طول به ترتیب  $M$  و  $N$  به صورت

$$\begin{aligned} \mathbf{X}_{p \times M} &= [\vec{x}_1, \dots, \vec{x}_M] \\ \mathbf{Y}_{p \times N} &= [\vec{y}_1, \dots, \vec{y}_N] \end{aligned} \quad (1)$$



شکل ۱: فرآیند همترازی دو سیگنال تقریباً مشابه توسط روش پیچش زمانی پویا

باشند، پیچش زمانی پویا به دنبال پیدا کردن مسیر پیچش به صورت جفت اندیس‌های  $\{(i_1, j_1), \dots, (i_k, j_k), \dots, (i_K, j_K)\}$  است به طوری که شرایط زیر برای این مسیر برقرار باشد:

$$\begin{aligned} i_1 &= j_1 = 1 \\ i_k &= M \\ j_k &= N \\ i_{k-1} &\leq i_k, \quad 1 < k \leq M \\ j_{k-1} &\leq j_k, \quad 1 < k \leq N \end{aligned} \quad (2)$$

و تابع هزینه

$$cost = \sum_{k=1}^K d(\vec{x}_{i_k}, \vec{y}_{j_k}) \quad (3)$$

<sup>26</sup>Multi-Variate

تعریف می‌شود که در آن  $n_k$  تعداد نمونه‌های کلاس  $k$  است و  $K$  تعداد کلاس‌ها می‌باشد. ماتریس کواریانس داخل کلاسی نیز به صورت

$$W = \frac{1}{n} \sum_{k=1}^K \sum_{y \in \Omega_k} (x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T \quad (11)$$

قابل تعریف است که در آن  $y^{(i)}$  به برچسب نمونه  $x^{(i)}$  اشاره دارد. راستاهای مورد نظر در روش فیشر، راستاهای متعامدی هستند که نسبت زیر را که به نسبت فیشر معروف است، بیشینه می‌کنند:

$$r = \frac{w^T B w}{w^T W w} \quad (12)$$

اگر ماتریس  $W$  غیر تکین باشد، بردارهای ویژه متناظر با بزرگترین مقادیر ویژه ماتریس  $W^{-1}B$  راستاهای فیشر هستند. زمانی که تعداد ویژگی‌ها از تعداد نمونه‌ها بیشتر باشد، با مشکل تکین بودن یا تقریباً تکین بودن ماتریس کواریانس مواجه خواهیم شد. در این صورت می‌توان با جایگزین کردن ماتریس  $W$  با ماتریس  $W + \sigma^2 I_D$  که در آن  $I_D$  ماتریس همانی با سایز  $p$  و  $\sigma^2$  پارامتر تنظیم می‌باشد، راستاهای فیشر را بدست آورد.

## ۴ استخراج ویژگی‌های نهفته در فضای پیچش زمانی

در روش ارائه شده هر کنش به صورت مجموعه‌ای از سری‌های زمانی یا به عبارتی یک سری زمانی چند بعدی بیان می‌شود. این سری زمانی چند بعدی مشخص کننده تغییرات موقعیت سه بعدی مفاصل بدن در طول زمان است. هر بعد این سری زمانی در واقع موقعیت یکی از سه مختصات  $x$ ،  $y$  و  $z$  مربوط به هر کدام از مفاصل اسکلت را در طول زمان بیان می‌کند. به عنوان مثال برای یک اسکلت با تعداد ۲۰ مفصل و طول اکشن برابر ۱۰۰ فریم، سری زمانی مورد نظر دارای ۶۰ بعد یا زیر سری هر کدام به طول ۱۰۰، خواهد بود.

نکته‌ای که حائز اهمیت و در واقع یک از چالش‌های مهم در بازشناسی کنش است این است که نه تنها کنش‌های متفاوت، دارای طول (مدت انجام) متفاوتی از یکدیگر هستند، بلکه کنش‌های یکسان (مثلاً نشستن) که توسط افراد متفاوت (بلکه حتی یک فرد در دفعات متفاوت) انجام می‌شود نیز دارای طول متفاوت هستند. علاوه بر این، افراد متفاوت قسمت‌های مختلف کنش‌های یکسان را ممکن است با سرعت‌های متفاوتی انجام دهند. به طور مثال، در کنش پرتاب یک جسم با دست، فردی ممکن است مرحله بالا آوردن دست را آهسته تر از فرد دیگر، ولی رها کردن جسم را با سرعت بیشتری از آن فرد انجام دهد. در نتیجه حتی نمونه برداری از کنشها برای یکسان سازی طول آنها نیز نمی‌تواند تناظر خوبی بین فریمها بوجود آورد. لذا استخراج ویژگی‌های متداولی مانند ویژگی‌های موجکها و ویژگی‌های فرکانسی

در روش کرنلی کاهش بعد [۳۶]، بجای ماتریس کواریانس، از ماتریس کرنل نرمالیزه استفاده می‌شود. ماتریس کرنل (غیر نرمالیزه) به صورت

$$K_{ij} = K(x^{(i)}, x^{(j)}) \quad (8)$$

تعریف می‌شود که در آن  $K$  تابع کرنل مورد نظر و  $x^{(i)}$  نمونه‌ی  $i$ ام در مجموعه‌ی آموزشی است. می‌توان نشان داد که ماتریس کرنل نرمالیزه به صورت

$$K' = K - \frac{1}{N} K - \frac{1}{N} K^T + \frac{1}{N} K \frac{1}{N} K^T \quad (9)$$

قابل بیان است که در آن  $1_N$  ماتریس  $N \times N$  با تمام درایه‌ها برابر یک است.

حال اگر نمونه‌ها به یک فضای با بعد بالای دلخواه نگاشت شوند ولی تنها ضرب داخلی بین نمونه‌ها در این فضا به صورت یک تابع کرنل مشخص شده باشد، می‌توانیم با استفاده از روش کرنلی تحلیل مولفه‌های اساسی، در فضای جدید کاهش بعد را انجام دهیم. در واقع می‌توانیم کاهش بعد در فضای جدید را بدون نیاز به دانستن بردار ویژگی نمونه‌ها در فضای جدید انجام دهیم. به بیانی دیگر، حقه کرنل ما را قادر می‌سازد که با ویژگی‌هایی که در فضاهای غیر اقلیدسی نهفته است کار کنیم و تنها کافیت ضرب داخلی در این فضاها را بدانیم.

در این مقاله تلاش می‌کنیم تا برای طبقه‌بندی کنشها از ویژگی‌هایی که در فضای پیچش زمانی پویا نهفته است استفاده کنیم زیرا در این فضا کنش‌های مربوط به کلاسهای یکسان فارغ از سرعت‌ها و تاخیرهای انجام متفاوت، نزدیک به هم قرار دارند و در نتیجه طبقه‌بندی کنش‌ها بهتر صورت می‌گیرد. در نتیجه با توجه به حقه کرنل، استفاده از یک ضرب داخلی که تابعی که فاصله پیچش زمانی پویا باشد راه حل بسیار خوبی به نظر میرسد زیرا این ضرب داخلی ویژگی‌های نهفته در فضای پیچش زمانی پویا را در اختیار ما قرار می‌دهد. اما چالش اصلی در استفاده کردن از فاصله پیچش زمانی به عنوان تابع کرنل، محاسبات سنگین آن است، چراکه لازم است فاصله پیچش زمانی هر نمونه با تمام نمونه‌های دیگر بدست آید.

## ۳-۳- استخراج ویژگی‌های تمایزی با استفاده از روش تحلیل افتراقی فیشر تنظیم شونده

در روش تحلیل افتراقی فیشر، راستاهایی که در آنها میانگین کلاسها از یکدیگر فاصله زیادی دارند و در عین حال نمونه‌های هر کلاس از یکدیگر فاصله کمی دارند، به منظور افکنش نمونه‌ها بر روی این راستاها بدست می‌آیند. فرض کنید  $\mu_k$  میانگین نمونه‌های کلاس  $k$  و  $\mu$  میانگین تمام نمونه‌ها باشد. در این صورت ماتریس کواریانس بین کلاسی به صورت

$$B = \frac{1}{n} \sum_{k=1}^K n_k (\mu_k - \mu)(\mu_k - \mu)^T \quad (10)$$

دسته دوم ویژگیهای شبه کرنلی بین کنشی در نظر می‌گیریم. برای این کار، از همان همترازی استفاده شده در مرحله قبل یعنی همترازی  $p$  بعدی ( $p$  سه برابر تعداد مفاصل است) استفاده می‌کنیم و فاصله بین دو سری زمانی سه بعدی برای هر کدام از مفاصل را به عنوان دسته دوم ویژگیهای شبه کرنلی بین کنشی به صورت زیر در نظر می‌گیریم:

$$\Phi_2(X) = \left[ \sum_{k=1}^K a(\vec{x}_{1,t_k}, \vec{\theta}_{1,t_k}^1), \dots, \sum_{k=1}^K a(\vec{x}_{1,t_k}, \vec{\theta}_{1,t_k}^2), \dots, \sum_{k=1}^K a(\vec{x}_{1,t_k}, \vec{\theta}_{1,t_k}^g), \dots, \sum_{k=1}^K a(\vec{x}_{1,t_k}, \vec{\theta}_{1,t_k}^p) \right]^T \quad (14)$$

که در آن  $\vec{x}_{1,t_k}$  و  $\vec{\theta}_{1,t_k}^g$  بردارهای سه بعدی مشخص کننده موقعیت مفصل  $t$ -ام به ترتیب در فریم  $t_k$  مربوط به نمونه کنش  $X$  و فریم  $t_k$  مربوط به نمونه مرجع کلاس  $g$ ، تعداد مفاصل و  $K$  تعداد جفت اندیس‌های همترازی است. علت عدم استفاده از نماد  $(\cdot)$  DTW در معادله بالا آن است که سری زمانی های هر مفصل به طور جداگانه بین دو نمونه کنش همتراز نمی‌شوند، بلکه از همان جفت اندیس‌های متناظر بدست آمده در همترازی کلی (چند بعدی) دو نمونه کنش استفاده می‌شود.

#### ۴-۲- ویژگیهای شبه کرنلی میان اسکلتی

دسته سوم ویژگیهایی که معرفی می‌کنیم، فاصله بین دودودی سری های زمانی مفاصل در یک نمونه کنش است. در اینجا دیگر نیازی به انجام پیچش زمانی پویا نیست و صرفاً فواصل بین مفاصل هر اسکلت در طول زمان با هم جمع می‌شوند:

$$\Phi_3(X) = \left[ \sum_{i=1}^M a(\vec{x}_{1,i}, \vec{x}_{2,i}), \dots, \sum_{i=1}^M a(\vec{x}_{1,i}, \vec{x}_{2,i}), \dots, \sum_{i=1}^M a(\vec{x}_{1,i}, \vec{x}_{2,i}) \right]^T \quad (15)$$

از آنجا که فاصله بعضی مفاصل از هم، مانند مفصل آرنج و شانه، در طول زمان تغییر نمی‌کند، در اینجا فقط جفت مفاصلی که فاصله متغیر از هم دارند را در نظر می‌گیریم مانند مفصل مچ دست راست با مچ پای راست و غیره.

#### ۴-۳- طبقه‌بندی کنش‌ها

سه دسته ویژگی معرفی شده در بالا در کنار هم قرار گرفته و بردار ویژگی نهایی را به صورت زیر تشکیل می‌دهند:

$$\Phi(X) = [\alpha_1 \Phi_1(X)^T, \alpha_2 \Phi_2(X)^T, \alpha_3 \Phi_3(X)^T]^T \quad (16)$$

که مبنای آن بر تناظر زمانی نمونه‌ها استوار است، نمی‌تواند برای آموزش طبقه‌بندها موثر واقع شود، هرچند طبقه‌بند بسیار قوی باشد.

هدف از این مقاله ارائه روشی است که بتواند ویژگیهایی را برای آموزش طبقه‌بند استفاده کند که در فضای پیچش زمانی قرار دارند و در نتیجه نسبت به سرعت‌ها و تاخیرهای متفاوت کنش‌ها مقاوم باشند.

#### ۴-۱- ویژگیهای شبه کرنلی بین کنشی

در روش ارائه شده از فاصله پیچش زمانی بین دو سری زمانی به عنوان تابع کرنل برای کاهش بعد استفاده می‌شود. در نتیجه ویژگیهای کاهش بعد یافته ای که برای هر نمونه بدست می‌آیند، ویژگیهای نهفته در فضای پیچش زمانی خواهند بود. نکته ای که لازم است اشاره گردد این است که در ادبیات روشهای کرنلی، مرسوم است که ابتدا تابعی از جنس فاصله به تابعی از جنس ضرب داخلی تبدیل شود تا بتوان از آن به عنوان کرنل استفاده نمود. اما در اینجا همانطور که در نتایج نشان داده می‌شود بدون انجام این تبدیل نیز قادریم با ویژگیهای نهفته در فضای پیچش زمانی کار کنیم. تفاوت عمده دیگری که روش ما با روشهای مرسوم کرنلی دارد این است که بجای محاسبه تابع کرنل برای دودودی نمونه‌ها، از هر کلاس کنش یکی از نمونه های آموزشی به عنوان مرجع برای آن کلاس انتخاب می‌شود و سپس تابع کرنل فقط نسبت به نمونه های مرجع محاسبه می‌شود. به همین دلیل نام "شبه کرنل" را برای توابع معرفی شده انتخاب کرده ایم. در مورد روش پیدا کردن نمونه های مرجع در ادامه توضیح داده می‌شود. مزیت مهم رویکرد پیشنهادی، کاهش قابل توجه محاسبات پیچش زمانی پویا می‌باشد. تفاوت مهم دیگر آن است که به جای جایگذاری مقادیر تابع کرنل در ماتریس کواریانس، مستقیماً از آنها به عنوان ویژگی استفاده می‌کنیم و سپس مراحل آموزش و تست طبقه‌بند را با این ویژگیها انجام می‌دهیم. با این کار در واقع یک تبدیل غیر خطی برای ویژگیهای اولیه ایجاد می‌کنیم.

اگر حداقل هزینه در معادله (۲) را به عنوان فاصله پیچش زمانی پویا برای دو سری زمانی  $X$  و  $Y$  با نماد  $DTW(X, Y)$  در نظر بگیریم، ویژگی شبه کرنلی بین کنشی را برای نمونه  $X$  به صورت زیر تعریف می‌کنیم:

$$\Phi_1(X) = [DTW(X, \theta^1), \dots, DTW(X, \theta^g)]^T \quad (13)$$

که در آن  $\Phi_1$  تابع تولید ویژگیهای شبه کرنلی بین کنشی،  $\theta^g$  نمونه مرجع برای کلاس  $g$ -ام و  $G$  تعداد کلاسها می‌باشد. از فاصله اقلیدسی نیز برای تابع فاصله  $d$  در (۳) استفاده می‌کنیم.

ویژگیهای شبه کرنلی بین کنشی تعریف شده در بالا، مجموع فاصله بین تمام مفاصل را بین دو نمونه کنش در طول زمان مدل می‌کند. اما وضعیت تک تک مفاصل متناظر نسبت به هم را به خوبی بیان نمی‌کند. برای آنکه ویژگیهای دقیق تری بدست بیاوریم، فاصله بین تک تک مفاصل در طول زمان را نیز به عنوان

قرار می‌گیرند و سپس میانگین نرخ بازشناسی در تمام اجراها به عنوان نتیجه نهایی اعلام می‌شود. در هر اجرا، پارامترهای موجود در روش که شامل پارامتر انقباض طبقه‌بند و ضرایب دسته‌های ویژگی می‌باشند، از طریق اعتبارسنجی روی مجموعه آموزشی بدست می‌آیند. برای از بین بردن اثر مخربی که جهت حرکت افراد یا زاویه بدن آنها با سنسور در ویژگیهای کنش‌ها دارند، لازم است اسکلت‌ها هم تراز<sup>۲۹</sup> شوند. برای هم تراز کردن اسکلت‌ها در تمام فریم‌ها دو تبدیل انجام می‌شود. یک تبدیل انتقال مفصل لگن به مرکز مختصات است. تبدیل دیگر دوران دادن اسکلت حول محور  $y$  است به گونه‌ای که تصویر خط متصل‌کننده مفصل شانه‌ها در صفحه  $xz$  موازی محور  $x$  باشد. این تبدیل‌ها در شکل ۲ نشان داده شده‌اند.

### ۵-۱- دیتاست TST

دیتاست TST Fall [۱۸] شامل دو دسته کنش‌های روزمره و کنش‌های غیر ارادی افتادن می‌باشد. دسته کنش‌های روزمره شامل نشستن، برداشتن شیء، راه رفتن و دراز کشیدن، و دسته کنش‌های افتادن شامل کنش‌های متفاوت افتادن مانند افتادن رو به جلو، افتادن به پشت، افتادن به پهلو و افتادن منتهی به نشستن می‌باشد. هر کنش سه دفعه توسط ۱۱ فرد انجام شده است و در نتیجه این دیتاست شامل ۲۶۴ نمونه کنش می‌باشد. علاوه بر داده‌های عمق ثبت شده بوسیله سنسور کینکت، داده‌های سنسورهای پوشیدنی نیز در این دیتاست موجود است که اطلاعات ارزشمندی را فراهم می‌آورد. اما در اینجا با توجه به هدف این مقاله، فقط از اطلاعات عمق برای بازشناسی کنش‌ها استفاده شده است.

مفاصل موجود در داده‌های اسکلت این دیتاست در شکل ۲ نشان داده شده است. مفاصل ۲۲ تا ۲۵ به علت عدم دخالت در کنش‌های این دیتاست در روش ارائه شده مورد استفاده قرار نگرفته‌اند. همچنین مفاصلی که برای هم‌ترازی اسکلتها در فریم‌های متوالی یک کنش استفاده شده‌اند، در این شکل مشخص شده‌اند.

کنش‌های غیر ارادی موجود در این دیتاست آن را به یکی از چالش‌برانگیزترین دیتاست‌ها تبدیل کرده است که روشهایی که بتوانند بازشناسی موفقی روی این دیتاست داشته باشند، علی‌الخصوص برای کاربردهای مراقبت از افراد بیمار و سالخورده بسیار مؤثر خواهند بود.

که در آن  $\beta$ ها ضرایبی هستند که برای جبران اسکیل متفاوت ویژگیها نسبت به یکدیگر به کار می‌روند تا آنها را از نظر اندازه در محدوده‌های نزدیک به هم قرار دهند. این ضرایب با استفاده از اعتبارسنجی متقابل چندلایه قابل تنظیم می‌باشد. در نهایت طبقه‌بندی کنش‌ها با استفاده از یک طبقه‌بند مناسب قابل انجام است. در اینجا ما از روش تنظیم شونده فیشر<sup>۲۷</sup> به علت اینکه در عین به صرفه بودن از نظر محاسباتی، از قدرت طبقه‌بندی بسیار بالایی برخوردار است، استفاده کرده ایم. با توجه به اینکه معمولا تعداد نمونه‌ها نسبت به تعداد ویژگیها قابل ملاحظه نیست، خطر بیش‌برازش وجود دارد و استفاده از پارامتر انقباض/ تنظیم<sup>۲۸</sup> تا حد خوبی این خطر را کاهش داده و منجر به یک طبقه‌بند با پارامترهای نسبتا پایدار می‌شود. برای بدست آوردن مقدار مناسب پارامتر تنظیم نیز از روش اعتبارسنجی متقابل چند لایه استفاده می‌شود. لازم به ذکر است که پرداختن به انتخاب طبقه‌بند مناسب در حیطه تمرکز این مقاله نبوده است و در عوض استخراج ویژگیهای مناسبی که نسبت به تغییرات داخل کلاسی مقاومت خوبی داشته باشد، هدف اصلی این مقاله بوده است.

### ۵ آزمایشات

در این قسمت روش پیشنهادی را بر روی سه دیتاست معروف کنش ارزیابی و نتایج را با آخرین روشهای موجود مقایسه می‌کنیم. دیتاست‌های انتخاب شده دیتاست TST Fall، دیتاست UTKinect و دیتاست UCFKinect می‌باشند. این دیتاست‌ها شامل تنوع خوبی از دسته‌های متفاوت کنش هستند. دیتاست TST Fall شامل کنش‌های غیر ارادی افتادن است که طبیعتا شامل تفاوت‌های زیاد بین فردی می‌باشد. دیتاست UTKinect شامل کنش‌های روزمره است که دارای سطح متوسطی از تفاوت‌های بین فردی می‌باشند. دیتاست UCFKinect نیز شامل کنش‌های نسبتا نمایشی می‌باشد که مناسب برای بازیها یا تعامل‌های انسان و رایانه هستند. هر کدام از این دیتاست‌ها در ادامه با جزئیات بیشتری شرح داده شده‌اند. اما نکته‌ای که حائز اهمیت است این است که روشهای موجود غالبا بر روی یک دیتاست تمرکز کرده و روش خود را بر روی آن دیتاست خاص خیره کرده‌اند. این در حالی است که روشی که بتواند روی دیتاست‌های متفاوت عملکرد قابل ملاحظه‌ای داشته باشد، از جامعیت و کاربرد بهتری برخوردار است.

روش ارزیابی در این تحقیق روی تمام دیتاست‌ها به این صورت است که در هر بار اجرای الگوریتم، کنش‌های مربوط به یک فرد به عنوان نمونه‌های تست و کنش‌های افراد دیگر به عنوان نمونه‌های آموزش در نظر گرفته می‌شود و نرخ بازشناسی محاسبه می‌گردد. الگوریتم به تعداد افراد موجود در دیتاست تکرار شده و هر بار یکی از افراد به عنوان فرد تست و مابقی به عنوان افراد آموزش

<sup>27</sup>Regularized Fisher Linear Discriminant Analysis

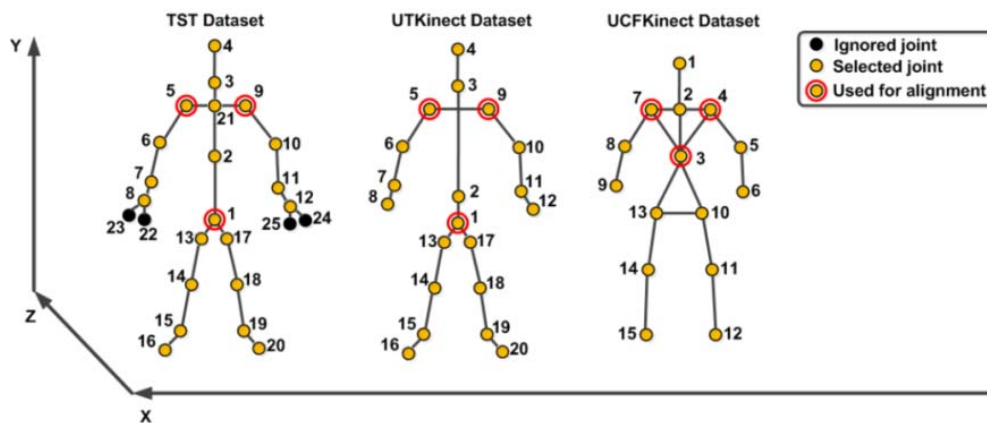
<sup>28</sup>Shrinkage



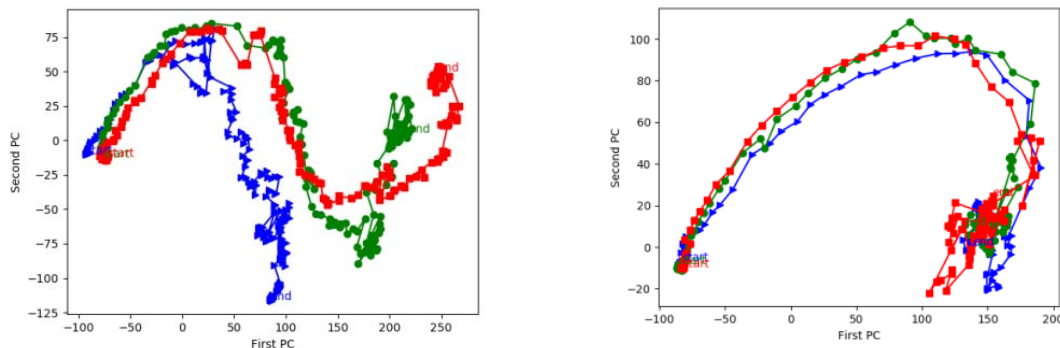
۲-۵- دیتاست UTKinect

این دیتاست برای ارزیابی استفاده می‌کنند، انجام می‌شود. لذا دیتاست UTKinect مورد استفاده، شامل ۱۹۹ نمونه کنش می‌باشد. مفاصل موجود در داده های اسکلت این دیتاست در شکل ۲ نشان داده شده است.

دیتاست UTKinect [۱۹] شامل ده کنش راه رفتن، نشستن، برخاستن، برداشتن شیء، حمل کردن، پرتاب کردن، هل دادن، کشیدن، دست تکان دادن و دست زدن می‌باشد. هر کنش دو مرتبه توسط ده فرد انجام شده است. در این دیتاست، یکی از نمونه های کلاس حمل کردن به درستی انجام نشده است و کنش ثبت شده شباهتی به حمل کردن ندارد. در نتیجه این نمونه را از دیتاست حذف کرده‌ایم. این حذف در مقالات دیگری نیز که از



شکل ۲: مفاصل موجود در داده های اسکلت برای هر دیتاست به همراه راستاهای محورهای مختصات.



شکل ۳: دنباله کنشهای دراز کشیدن (چپ) و افتادن منتهی به نشستن در دیتاست TST که به ترتیب توسط افراد شماره ۳ و ۱۱ انجام شده است. هر کنش سه اجرای متفاوت دارد. ابتدا و انتهای هر دنباله با کلمات start و end مشخص شده است.

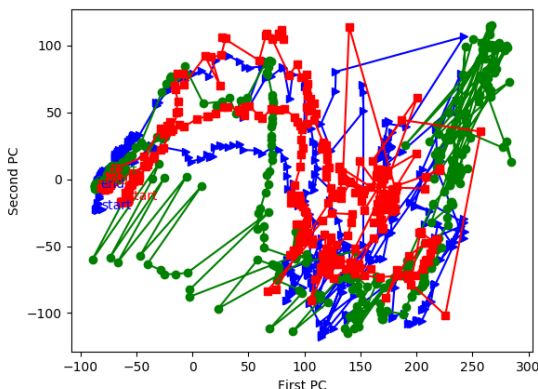
دیتاست UCFKinect [۲۰] شامل ۱۶ کنش نسبتاً کوتاه می‌باشد که هر کنش پنج مرتبه توسط ۱۶ فرد انجام شده است. در مجموع تعداد ۱۲۸۰ کنش در این دیتاست وجود دارد. مفاصل موجود در داده های اسکلت این دیتاست در شکل ۲ نشان داده شده است.

این دیتاست بیشتر برای هدف شبیه سازی کنشهایی که در بازیها کاربرد دارند جمع آوری شده است. کنشهای موجود در این دیتاست شامل بالا رفتن از نردبان، پریدن، بالا پریدن، پریدن

واریانس نسبتاً زیاد در نحوه انجام دادن هر کنش توسط افراد مختلف از چالشهای این دیتاست به شمار می‌آید. نکته دیگری که در مورد این دیتاست وجود دارد نویز اندازه گیری موقعیت مفاصل است، به طوری که پرشهای غیر ارادی در موقعیت مفاصل را ایجاد کرده است. این نویز یکی دیگر از چالشهای این دیتاست به شمار می‌رود.

۳-۵- دیتاست UCFKinect

دنباله دراز کشیدن در دیتاست TST که توسط افراد شماره ۲ و ۳ انجام شده است را نشان می‌دهد. در این شکل تناظر بین حالت‌های اسکلت در دو دنباله بوسیله خطوط آبی نشان داده شده است.



شکل ۴: دنباله سه اجرای متفاوت کنش دراز کشیدن در دیتاست TST که توسط فرد شماره ۱ انجام شده است. این فرد بعد از دراز کشیدن به موقعیت اولیه اش بر می‌گردد.

تناظر حاصل از پیچش زمانی بین دو دنباله در شکل ۵ سمت راست نشان داده شده است که برتری تناظرهای حاصل از پیچش زمانی به خوبی نمایان می‌کند.

#### ۵-۵- ارزیابی روش ارائه شده

با توجه به این که شماره گذاری مفاصل در این سه دیتاست یکسان نیستند، در جدول ۱، مفاصلی که در دسته دوم ویژگی‌های بین کنشی و دسته ویژگی‌های بین اسکلتی برای هر کدام از سه دیتاست مورد استفاده قرار گرفته‌اند، مشخص شده است.

نرخ بازشناسی روش ارائه شده به همراه نرخ بازشناسی بهترین روش‌های موجود بر روی دیتاست TST در جدول ۲ نشان داده شده است. ماتریس ابهام بازشناسی کنش‌ها با استفاده از روش ارائه شده بر روی این دیتاست در شکل ۶ نشان داده شده است.

لازم به ذکر است که برخی نرخ‌هایی که در ادبیات برای این دیتاست گزارش شده، با استفاده از داده‌های سنسورهای پوشیدنی بدست آمده است و در نتیجه بخاطر عدم ارتباط با موضوع این تحقیق در این جدول ذکر نشده است.

نرخ بازشناسی روش پیشنهادی و بهترین روش‌های موجود بر روی دیتاست UTKinect در جدول ۳ نشان داده شده است. لازم به ذکر است که همان‌طور که در جدول ۱ آمده است، تنها مفاصل ستون فقرات، گردن و سر در دسته دوم ویژگی‌های شبه کرنلی بین کنشی در این دیتاست استفاده شده است. استفاده از مفاصل بیشتر در این دسته اندکی نرخ بازشناسی را کاهش می‌دهد. علت این امر می‌تواند نويز موجود در تخمین موقعیت مفاصل در دیتاست باشد که وقتی مفاصل به صورت مجزا استفاده می‌شوند

کوتاه، پریدن با جفت پا رو به جلو، پریدن از روی مانع، تعادل با دست‌های باز، لگد زدن، مشت زدن، دویدن، یک قدم به عقب، جلو، چپ و راست رفتن، چرخیدن به سمت چپ و چرخیدن به سمت راست.

#### ۵-۴- مصور سازی دنباله کنش‌ها در فضای حالت اسکلت

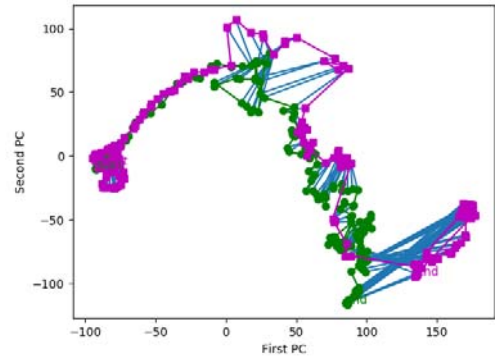
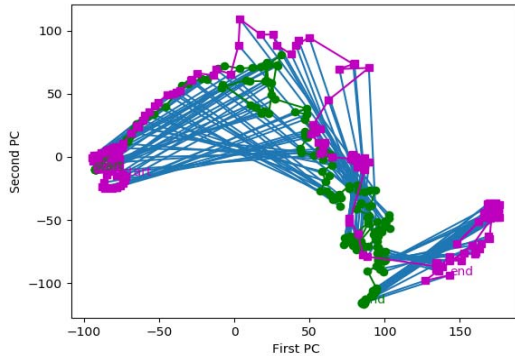
در این قسمت ما برای توضیح مؤثرتر روش پیشنهادی، راه حلی را برای مصور سازی دنباله کنش‌ها ارائه می‌دهیم. این راه حل دارای مزایای متعددی دیگری نیز هست، از جمله اینکه به ما کمک می‌کند بدون بررسی تک تک فریم‌های یک کنش، در آن واحد بتوانیم دنباله یک کنش را مشاهده کنیم. از آنجا که صفحه دو بعدی برای مشاهده مناسب تر است، دنباله کنش‌ها را در صفحه دو بعدی ترسیم می‌کنیم. برای انجام این کار ابتدا موقعیت مفاصل در هر فریم را در یک بردار قرار می‌دهیم. به عنوان مثال اگر داده‌های اسکلت دارای موقعیت ۲۰ مفصل باشند، از آنجا که موقعیت هر مفصل سه بعدی است، یک بردار ۶۰ بعدی برای هر فریم خواهیم داشت. این بردارها را بردار اسکلت می‌نامیم. از بردارهای اسکلت در تمام فریم‌های مربوط به نمونه‌های کنش یکی از افراد دیتاست، برای آموزش PCA استفاده می‌کنیم. فضای PCA بدست آمده با اینکه تنها توسط نمونه‌های یک فرد آموزش داده شده است اما به خوبی می‌تواند توصیف کننده بردارهای اسکلت مربوط به کنش‌های افراد دیگر نیز باشد. دو مولفه اول PCA را برای تصویر کردن بردارهای اسکلت در تمام فریم‌ها برای هر نمونه کنش استفاده می‌کنیم و بدین ترتیب دنباله کنش را در یک فضای دو بعدی ترسیم می‌کنیم. این فضا را فضای حالت اسکلت می‌نامیم. شکل ۳ دنباله‌های مربوط به سه نمونه کنش دراز کشیدن که توسط فرد شماره ۳، و سه نمونه کنش افتادن منتهی به نشستن که توسط فرد شماره ۱۱ در دیتاست TST انجام شده است را نشان می‌دهد.

یکی از مزایای مهم این مصور سازی، یافتن نمونه‌های غیر عادی در یک دیتاست است. به عنوان مثال در شکل ۴ دنباله‌های سه اجرای کنش دراز کشیدن که توسط فرد شماره ۱ در دیتاست TST انجام شده است، ترسیم شده است. بعد از مشاهده این دنباله‌ها متوجه شدیم که در دنباله ترسیم شده، نقاط ابتدا و انتهای دنباله بسیار نزدیک به هم قرار دارند و نتیجه گرفتیم که این فرد برخلاف ۱۰ فرد دیگر این دیتاست (نمونه‌های آن در سمت راست شکل ۳ نشان داده است)، بعد از انجام دراز کشیدن مجدداً به موقعیت اولیه اش برمی‌گردد. در نهایت نیز با بررسی دنباله فریم‌های اصلی این فرد صحت این فرضیه تأیید شد. با این وجود برای تست پایداری روش نسبت به نمونه‌های دورافتاده این نمونه‌ها را از دیتاست حذف نکردیم.

مصور سازی دنباله‌ها توانایی روش پیچش زمانی پویا در ایجاد تناظر بین فریم‌ها در دو دنباله را نیز می‌تواند به خوبی نشان دهد. شکل ۵ سمت چپ، تناظر معمولی بدون پیچش زمانی بین دو

تست انتخاب شده است. در نتیجه نمونه هایی از فرد تست شونده در مجموعه آموزش قرار می گیرد و به یادگیری طبقه بند کمک می کند.

نمود بیشتری پیدا می کند. شایان ذکر است که نرخهای مانند ۹۸٫۵٪ [۳۷] و ۹۸٫۸٪ [۳۸] و ۹۹٫۱۹٪ [۳۹] نیز در مقالات دیگر برای این دیتاست گزارش شده است که در جدول نیامده است. علت عدم گزارش این نرخها این است که در آزمایشات انجام شده در این مقالات، نمونه های تمام افراد با هم مخلوط شده و سپس درصدی از آنها به صورت تصادفی برای



شکل ۵: تناظر بین فریمها در دو دنباله دراز کشیدن که توسط افراد شماره ۲ و ۳ در دیتاست TST انجام شده است. تناظرها بوسیله خطوط آبی نشان داده شده است. چپ: تناظر بوسیله اندیسهای فریمی یکسان، راست: تناظر حاصل از پیش زمانی پویا.

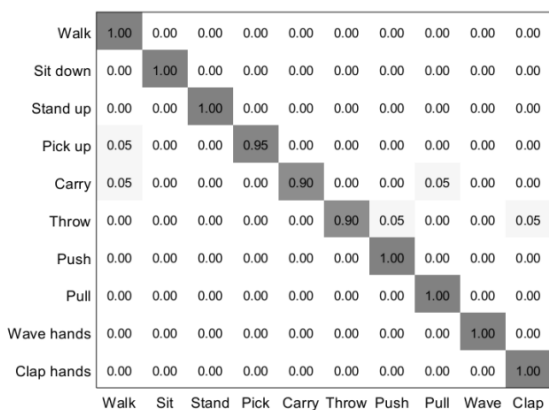
جدول ۱: مفصل مورد استفاده در دسته دوم ویژگی های بین کنشی و دسته ویژگیهای بین اسکلتی برای هر کدام از سه دیتاست

دیتاست	مفاصل دسته دوم ویژگیهای بین کنشی	جفت مفاصل دسته ویژگیهای بین اسکلتی
TST	2, 3, 4, 7, 11, 15, 19	(7,11), (7,15), (11,19), (15,19), (4,7), (4,11), (4,15), (4,19), (5,7), (7,19), (11,15), (1,15), (1,19), (1,7), (1,11)
UTKinect	2, 3, 4	(7,11), (7,15), (11,19), (15,19), (4,7), (4,11), (4,15), (4,19), (5,7), (7,19), (11,15), (1,15), (1,19), (1,7), (1,11)
UCFKinect	1, 2, 6, 9, 12, 15	(6,9), (12,15), (6,15), (9,15), (6,12), (9,12), (1,6), (1,9), (1,12), (1,15), (3,6), (3,9), (3,12), (3,15)

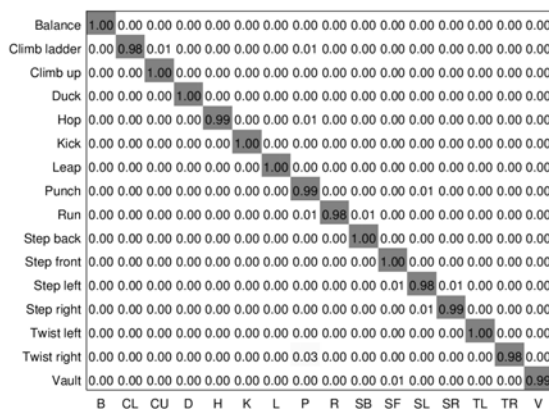
### ۵-۶- اثر ویژگیهای مختلف

در شکل ۹، دقت تشخیص در هر دیتاست بر حسب دسته ویژگیهای مختلف (به ترتیب در روابط (۱۳) تا (۱۵)) و ترکیبهای متفاوت آنها رسم شده است. همان طور که مشاهده می شود، بهترین دقت زمانی حاصل می شود که هر سه دسته ویژگی باهم استفاده می شوند. همچنین هنگامی که هر کدام از آنها به تنهایی بکار روند، از بین سه دسته ویژگی معرفی شده، دسته ویژگی دوم بهترین عملکرد و دسته ویژگی سوم بدترین عملکرد را نتیجه می دهد.

این در حالیست که چالش اصلی در بازشناسی کنش آن است که کنشهای افرادی که قبلا توسط سیستم دیده نشده اند، بازشناسی شوند. ماتریس ابهام بازشناسی کنشها با استفاده از روش ارائه شده بر روی این دیتاست در شکل ۷ نشان داده شده است. نرخ بازشناسی روش پیشنهادی و بهترین روشهای موجود بر روی دیتاست UCFKinect در جدول ۴ نشان داده شده است. مجددا نرخهایی مانند ۹۷٫۹٪ [۲۹] و ۹۸٫۷٪ [۳۰] برای این دیتاست که در مقالات دیگر با تقسیم بندی نمونه ها برای آموزش و تست به جای تقسیم بندی افراد، گزارش شده است، در جدول نیامده است. ماتریس ابهام بازشناسی کنشها با استفاده از روش ارائه شده بر روی این دیتاست در شکل ۸ نشان داده شده است.



شکل ۷: ماتریس ابهام طبقه‌بندی کنشها در دیتاست UTKinect



شکل ۸: ماتریس ابهام طبقه‌بندی کنشها در دیتاست UCFKinect

جدول ۴: نرخ شناسایی روش ارائه شده در مقایسه با سایر روش‌های ارائه شده بر روی دیتاست UCFKinect

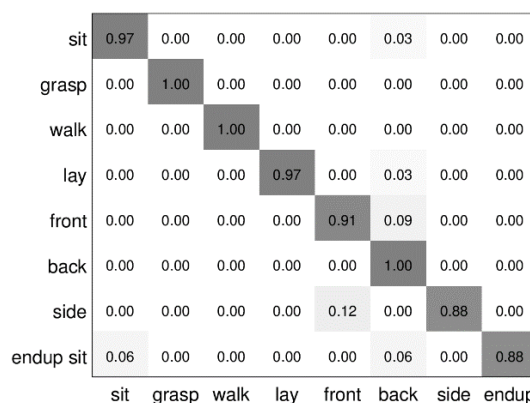
روش	نرخ شناسایی	سال ارائه
Zanfir et al. [23]	98.5%	2013
Kerola et al. [24]	98.8%	2014
Yang et al. [13]	97.1%	2014
Beh et al. [25]	98.9%	2014
Ding et al. [26]	98.0%	2015
Lu et al. [4]	97.6%	2016
Ghodsi et al. [22]	97.9%	2018
This method	98.7%	-

جدول ۲: نرخ شناسایی روش ارائه شده در مقایسه با سایر روش‌های ارائه شده بر روی دیتاست TST

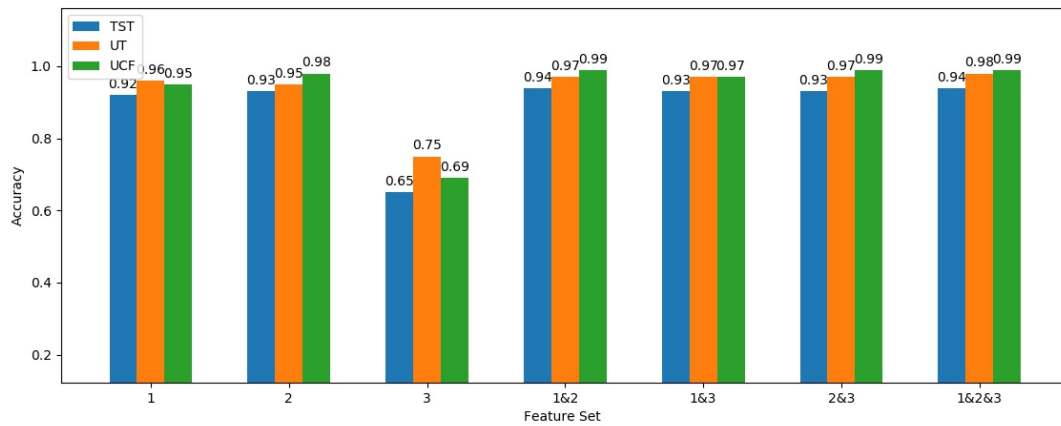
روش	نرخ شناسایی	سال
Ghojogh et al. [21]	88.6%	2018
Ghodsi et al. [22]	92.3%	2018
Tabejamaatetal. [42]	91.7%	2020
Our method	94.3%	-

جدول ۳: نرخ شناسایی روش ارائه شده در مقایسه با سایر روش‌های ارائه شده بر روی دیتاست UTKinect

روش	نرخ شناسایی	سال
Vemulapalli et al. [12]	97.0%	2014
Antunes et al. [27]	95.1%	2016
Gupta and Bhavsar [28]	96.0%	2016
Ghodsi et al. [22]	96.8%	2018
Zhang et al. [40]	95.9%	2018
Naveenkumar et al. [41]	98.9%	2020
Tabejamaat et al. [42]	96.1%	2020
Our method	97.5%	-



شکل ۶: ماتریس ابهام طبقه‌بندی کنشها در دیتاست TST



شکل ۹: دقت تشخیص برحسب دسته ویژگیها مختلف و ترکیبهای متفاوت آنها

۵-۷- انتخاب نمونه‌ی مرجع

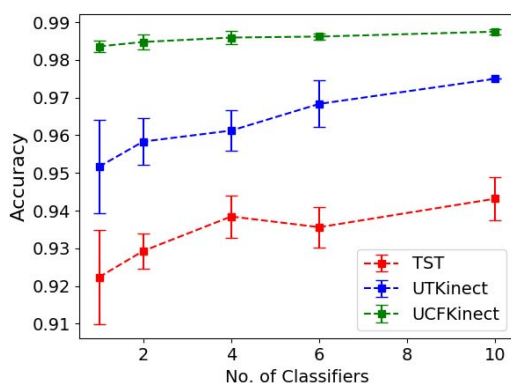
همان طور که در بخش ۳ بیان شد، در روش پیشنهادی، از بین نمونه های آموزشی هر کلاس کنش، یک نمونه به عنوان نمونه مرجع انتخاب می شود و سپس ویژگیهای شبه کرنلی نسبت به نمونه های مرجع استخراج می گردد. در این تحقیق، روشهای مختلفی برای انتخاب نمونه مرجع بهینه، بررسی و آزمایش گردید.

یکی از روشهایی که آزمایش شد، جستجوی نمونه‌ای بود که با معیار فاصله پیچش زمانی، میانگین نمونه‌های یک کلاس باشد (در این تعریف نمونه‌ای که کمینه مجموع مربعات فاصله از تمام نمونه‌های دیگر یک کلاس را دارا باشد، نمونه میانگین یک کلاس در نظر گرفته شد). روش دیگر، انتخاب نمونه‌ای بود که با معیاری شبیه معیار فیشر، بهترین نمونه باشد؛ بدین معنی که نسبت مجموع مربعات فاصله‌اش از نمونه‌های کلاسهای دیگر، به مجموع مربعات فاصله‌اش از نمونه‌های کلاس خودش، بیشینه باشد. در روشی دیگر، نمونه‌ی مرجع بهینه با روش اعتبارسنجی برای هر مجموعه آموزش جستجو شد. اما نهایتاً روشی که بهترین نتیجه را در برداشت، انتخاب تصادفی نمونه مرجع از بین نمونه‌های یک کلاس بود. با توجه به اینکه این روش کمی خاصیت تصادفی نیز به طبقه‌بند می‌بخشد، باعث می‌شود بتوان چند طبقه‌بند را که هر کدام نمونه های مرجع متفاوتی دارد، با هم ترکیب کرد و دقت نهایی را بهبود بخشید. برای ترکیب طبقه‌بندهای متفاوت نیز از روش ساده جمع امتیازها استفاده کردیم: ابتدا امتیازی که طبقه‌بندهای مختلف برای هر کلاس کنش را بدست می‌دهند را با هم جمع کردیم و سپس بین کلاسهای مختلف ماکزیمم گیری انجام دادیم. در جدول ۵، دقت طبقه بندی بر روی دیتاست TST با استفاده از چهار روش مختلف ذکر شده

کلاس کنش ایجاد می‌کنند، چالشی ترین دیتاست است. نکته دیگری که قابل مشاهده است این است که در مورد دیتاست UCFKinect، به علت وجود نمونه های آموزشی به تعداد کافی، حتی یک طبقه‌بند هم نتیجه بسیار خوبی را می‌دهد.

جدول ۵: دقت طبقه بندی بر روی دیتاست TST با استفاده از چهار روش مختلف برای انتخاب نمونه‌ی مرجع

روش انتخاب نمونه‌ی مرجع	دقت طبقه بندی
نمونه‌ی میانگین	89.8%
معیار فیشر	90.5%
اعتبارسنجی متقابل	92.0%
انتخاب تصادفی و تجمیع طبقه بندهای پایه	94.3%



شکل ۱۰: دقت طبقه‌بند نهایی برحسب تعداد طبقه‌بندهای پایه

۵-۸ بار محاسباتی روش ارائه شده

همان طور که در بخش ۳ توضیح داده شد، استفاده مستقیم از فاصله پیچش زمانی پویا برای طبقه بندی کنش ها، نیازمند اجرای تابع پیچش زمانی پویا بین هر کدام از نمونه‌های تست با تمام نمونه‌های آموزشی می‌باشد که در نتیجه این روش را از نظر

برای انتخاب نمونه‌ی مرجع، باهم مقایسه شده است. در شکل ۱۰ میانگین و انحراف معیار دقت طبقه بندی برای هر سه دیتاست نسبت به تعداد طبقه‌بندهای ترکیب شده رسم شده است. همانطور که مشاهده می‌شود دیتاست TST به علت دارا بودن کنش‌های غیرارادی که واریانس بسیار زیادی را در یک

های مبتنی بر حقه‌ی کرنل، این روش می‌تواند همزمان به دو مسئله‌ی توزیع غیر خطی دنباله‌های حرکتی و چالش شیفت زمانی موجود در آنها پرداخته و سبب حذف اثر آنها در فضای کرنل ایجاد شده گردد. از آنجا که تابع شبه کرنل ارائه شده تنها نسبت به نمونه‌های مرجع هر کلاس محاسبه می‌گردد، پیچیدگی محاسباتی آن نسبت به سایر روش‌های مبتنی بر پیچش زمانی پویا به طور قابل توجهی کاهش می‌یابد. مجموعه ارزیابی‌های انجام شده بر سه پایگاه داده‌ی TST، UTKinect و UCFKinect حاکی از قابل مقایسه بودن عملکرد روش ارائه شده با سایر روش‌های موجود در این حوزه می‌باشد.

## مراجع

- [1] Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M. and Moore, R., "Real-time human pose recognition in parts from single depth images". Communications of the ACM, 56(1), pp.116-124, 2013.
- [2] Lv, F. and Nevatia, R., May. Recognition and segmentation of 3-d human action using hmm and multi-class adaboost. In European conference on computer vision (pp. 359-372). Springer, Berlin, Heidelberg, 2006.
- [3] Guo, Y., Li, Y. and Shao, Z., "A spatiotemporal descriptor for rigid body motion recognition", IEEE transactions on cybernetics, 48(5), pp.1513-1525, 2018.
- [4] Hussein, M.E., Torki, M., Gowayyed, M.A. and El-Saban, M., "Human Action Recognition Using a Temporal Hierarchy of Covariance Descriptors on 3D Joint Locations", In IJCAI (Vol. 13, pp. 2466-2472), 2013.
- [5] Lu, G., Zhou, Y., Li, X. and Kudo, M., "Efficient action recognition via local position offset of 3D skeletal body joints", Multimedia Tools and Applications, 75(6), pp.3479-3494, 2016.
- [6] Gowayyed, M.A., Torki, M., Hussein, M.E. and El-Saban, M., "Histogram of Oriented Displacements (HOD): Describing Trajectories of Human Joints for Action Recognition", In IJCAI (pp. 1351-1357), 2013.
- [7] Guo, Y., Li, Y. and Shao, Z., "A flexible trajectory descriptor for articulated human action recognition", Pattern Recognition, 76, pp.137-148, 2018.
- [8] Wang, C., Wang, Y. and Yuille, A.L., 2013, June. An approach to pose-based action recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (pp. 915-922), 2013.
- [9] Ofli, F., Chaudhry, R., Kurillo, G., Vidal, R. and Bajcsy, R., "Sequence of the most informative joints (smij): A new representation for human skeletal action recognition", Journal of Visual Communication and Image Representation, 25(1), pp.24-38, 2014.
- [10] Weng, J., Weng, C., Yuan, J. and Liu, Z., "Discriminative Spatio-Temporal Pattern Discovery for 3D Action

محاسباتی سنگین می‌سازد. از طرف دیگر، در روش ارائه شده در این مقاله، با استفاده از ایده‌ی استفاده از نمونه‌های مرجع، حتی با وجود ارائه ویژگی‌های جدیدتر نیز، محاسبات به طرز قابل ملاحظه‌ای کاهش می‌یابد. در جدول ۶، زمان مورد نیاز برای روش نزدیکترین همسایه با استفاده از تابع پیچش زمانی پویا و روش ارائه شده در این مقاله، بر روی دیتاست TST نشان داده شده است. کامپیوتر مورد استفاده دارای پردازنده -core i7 4790 با سرعت 3.60 GHz بوده است. روش نزدیکترین همسایه فاقد مرحله آموزش است اما در روش ارائه شده از مجموعه‌ی آموزش برای استخراج ویژگی‌های مؤثر استفاده می‌شود و این نیازمند صرف زمان است. از آنجا که مرحله آموزش معمولاً تنها یک بار و آن هم به صورت برون‌خط انجام می‌شود، زمان آموزش نسبت به زمان تست از اهمیت کمتری برخوردار است. اما همان‌طور که مشاهده می‌شود، زمان تست روش ارائه شده به صورت قابل ملاحظه‌ای (۳۴ برابر) کمتر از زمان تست روش نزدیکترین همسایه است. دقت شود که زمان تست مربوط به ۴۸ نمونه‌ی تست است. نکته‌ی مهم دیگر آن است که کاهش زمان، منجر به کاهش نرخ شناسایی نشده است و با آن که نمونه‌ی تست با تمام نمونه‌های آموزش مقایسه نمی‌شود، اما به خاطر ویژگی‌های تمایزی استخراج شده، نرخ شناسایی به طور قابل ملاحظه‌ای افزایش یافته است.

جدول ۶: زمان آموزش و تست با استفاده از روش نزدیکترین همسایه و روش ارائه شده بر روی دیتاست TST در یک لایه از اعتبارسنجی

### مقابل

روش	زمان آموزش برحسب ثانیه (۲۱۶ نمونه‌ی آموزش)	زمان تست بر حسب ثانیه (۴۸ نمونه‌ی تست)	نرخ شناسایی
نزدیکترین همسایه با استفاده از تابع پیچش زمانی پویا	0	2453.3	85.6%
روش ارائه شده	409.1	72.4	94.3%

## ۶ نتیجه گیری

در این مقاله، یک روش بازشناسی کنش با استفاده از اطلاعات اسکلتی استخراج شده از تصاویر عمقی بدن انسانها ارائه گردیده است. روش ارائه شده دنباله‌های حرکتی را به صورت سری‌های زمانی چند بعدی مدل سازی نموده و از متریک DTW به منظور افکنش سری‌ها، به یک فضای جدید که قابلیت تمایز کلاسهای متفاوت از هم را دارد، استفاده می‌نماید. برخلاف سایر روش-

- [26] Ding, W., Liu, K., Cheng, F. and Zhang, J., "STFC: spatio-temporal feature chain for skeleton-based human action recognition", *Journal of Visual Communication and Image Representation*, 26, pp.329-337, 2015.
- [27] Antunes, M., Aouada, D. and Ottersten, B., "A revisit to human action recognition from depth sequences: Guided svm-sampling for joint selection", In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016 (pp. 1-8). IEEE, 2016.
- [28] Gupta, K. and Bhavsar, A., "Scale invariant human action detection from depth cameras using class templates", In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 38-45), 2016.
- [29] Slama, R., Wannous, H., Daoudi, M. and Srivastava, A., "Accurate 3D action recognition using learning on the Grassmann manifold", *Pattern Recognition*, 48(2), pp.556-567, 2015.
- [30] Jiang, X., Zhong, F., Peng, Q. and Qin, X., "Robust action recognition based on a hierarchical model", In *International Conference on Cyberworlds (CW)*, (pp. 191-198). IEEE, 2013.
- [31] E. Keogh, C. Ratanamahatana, "Exact indexing of time dynamic warping", *Knowledge and Information Systems*, vol. 7, pp. 358, 2005.
- [32] S. Riofrío, D. Pozo, J. Rosero, and J. Vázquez, "Gesture Recognition using Dynamic Time Warping and Kinect: A Practical Approach," *Proc. Intl. Conf. Information Systems and Computer Science*, 2017.
- [33] A. C̃alin, "Gesture Recognition on Kinect Time Series Data Using Dynamic Time Warping and Hidden Markov Models," *Proc. Intl. Symp. Symbolic and Numeric Algorithms for Scientific Computing*, 2016.
- [34] S. Sempena, N. Maulidevi, P. Aryan, "Human Action Recognition Using Dynamic Time Warping," *Proc. Intl. Conf. Electrical Engineering and Informatics*, 2011.
- [35] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 37, no. 1, pp. 2-86, 1991.
- [36] B. Schölkopf, A. Smola, K. Müller, "Kernel principal component analysis," *Proc. Intl. Conf. Artificial Neural Network*, 2005.
- [37] Liu, J., Wang, G., Hu, P., Duan, L.Y. and Kot, A.C., "Global context-aware attention lstm networks for 3d action recognition", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [38] Yang, Y., Deng, C., Tao, D., Zhang, S., Liu, W. and Gao, X., "Latent max-margin multitask learning with skeletons for 3-D action recognition", *IEEE Transactions on cybernetics*, 47(2), pp.439-448. 137, 2017.
- [39] Yang, Y., Deng, C., Gao, S., Liu, W., Tao, D. and Gao, X., "Discriminative multi-instance multitask learning for 3D action recognition", *IEEE Transactions on Multimedia*, 19(3), pp.519-529, 2017.
- [40] S. Zhang, Y. Yang, J. Xiao, X. Liu, Y. Yang, D. Xie, Y. Zhuang, "Fusing geometric features for skeleton-based action Recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.
- [11] Li, X., Zhang, Y. and Liao, D., "Mining key skeleton poses with latent svm for action recognition", *Applied Computational Intelligence and Soft Computing*, 2017.
- [12] Evangelidis, G., Singh, G. and Horaud, R., "Skeletal quads: Human action recognition using joint quadruples", In *IEEE 22nd International Conference on Pattern Recognition (ICPR)*, (pp. 4513-4518), 2014.
- [13] Vemulapalli, R., Arrate, F. and Chellappa, R., "Human action recognition by representing 3d skeletons as points in a lie group", In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 588-595), 2014.
- [14] Yang, X. and Tian, Y., "Effective 3d action recognition using eigenjoints", *Journal of Visual Communication and Image Representation*, 25(1), pp.2-11, 2014.
- [15] Xia, L., Chen, C.C. and Aggarwal, J.K. "View invariant human action recognition using histograms of 3d joints", In *IEEE Computer vision and pattern recognition workshops (CVPRW)*, (pp. 20-27), 2012.
- [16] Du, Y., Wang, W. and Wang, L., "Hierarchical recurrent neural network for skeleton based action recognition", In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1110-1118), 2015.
- [17] Li, C., Zhong, Q., Xie, D. and Pu, S., "Co-occurrence Feature Learning from Skeleton Data for Action Recognition and Detection with Hierarchical Aggregation", *arXiv preprint arXiv:1804.06055*, 2018.
- [18] <https://ieee-dataport.org/documents/tst-fall-detection-dataset-v2>
- [19] <http://cvrc.ece.utexas.edu/KinectDatasets/HOJ3D.html>
- [20] <http://www.cs.ucf.edu/~smasood/datasets/>
- [21] Ghojogh, B., Mohammadzade, H. and Mokari, M., "Fisherposes for Human Action Recognition Using Kinect Sensor Data", *IEEE Sensors Journal*, vol. 18, no. 4, pp. 1612-1627, 2018.
- [22] Ghodsi, S., Mohammadzade, H. and Korke, E., "Simultaneous Joint and Object Trajectory Templates for Human Activity Recognition from 3-D Data" *Journal of Visual Communication and Image Representation*, vol. 55, pp. 729-741, 2018.
- [23] Zafir, M., Leordeanu, M. and Sminchisescu, C., "The moving pose: An efficient 3d kinematics descriptor for low-latency action recognition and detection", In *Proceedings of the IEEE international conference on computer vision* (pp. 2752-2759), 2013.
- [24] Kerola, T., Inoue, N. and Shinoda, K., "Spectral graph skeletons for 3D action recognition", In *Asian Conference on Computer Vision* (pp. 417-432). Springer, Cham, 2014.
- [25] Beh, J., Han, D.K., Durasiwami, R. and Ko, H., "Hidden Markov model on a unit hypersphere space for gesture trajectory recognition", *Pattern Recognition Letters*, 36, pp.144-153, 2014.

نرجس الهدی محمدزاده مدرک کارشناسی خود را از دانشگاه صنعتی امیرکبیر، کارشناسی ارشد را از دانشگاه کلگری، کلگری، کانادا و مدرک دکترا را از دانشگاه تورنتو، تورنتو، کانادا همه در رشته مهندسی برق دریافت نمود. او از سال ۱۳۹۳ تاکنون استادیار دانشکده مهندسی برق دانشگاه صنعتی شریف بوده است. زمینه‌های پژوهشی مورد علاقه وی یادگیری ماشین، یادگیری عمیق، بینایی ماشین و پردازش سیگنال آماری می‌باشد.

محسن تابع جماعت مدرک کارشناسی خود را از دانشگاه آزاد اسلامی دریافت نمود. زمینه‌های پژوهشی مورد علاقه او هندسه محاسباتی، شبیه‌سازی عددی، آنالیز معنایی و بیوانفورماتیک می‌باشد.

recognition using multilayer lstm networks,” IEEE Transactions on Multimedia, 20 (9), pp. 2330–2343, 2018.

- [41] Naveenkumar M., Domnic S., “Deep Ensemble Network Using Distance Maps and Body Part Features for Skeleton based Action Recognition,” Pattern Recognition, vol. 100, 2020.
- [42] Tabejamaat M. and Mohammadzade H., “Contributive Representation based Reconstruction for Online 3D Action Recognition,” *International Journal of Pattern Recognition and Artificial Intelligence*, 2020.
- [43] Q. Nie, J. Wang, X. Wang, Y. Liu, “View-invariant Human Action Recognition Based on A 3D Bio-constrained Skeleton Model,” IEEE Transactions on Image Processing, 28 (8), pp. 3959–3972, 2019.
- [44] K. Papadopoulos, G. Demisse, E. Ghorbel, M. Antunes, D. Aouada, B. Ottersten, “Localized Trajectories for 2D and 3D Action Recognition,” Sensors, vol. 19, 3503, 2019.
- [45] S. Cho, M.H. Maqbool, F. Liu, H. Foroosh, “Self-Attention Network for Skeleton-based Human Action Recognition,” IEEE Winter Conference on Applications of Computer Vision, 2020.
- [46] J. Liu, N. Akhtar, A. Mian, “Skepxels: Spatio-temporal Image Representation of Human Skeleton Joints for Action Recognition,” IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.