

تشخیص حرکات دست برای تعامل انسان و رایانه با استفاده از کدگذاری خط سیر حرکت و شبکه عصبی مصنوعی

فرزانه آزادی پورقهستانی^۱ و عصمت راشدی^۲

چکیده

حرکات دست انسان حالتی از تعامل غیرکلامی است که در ارتباط بین انسان و رایانه مورد استفاده قرار می‌گیرد. شهودی و طبیعی بودن حرکات دست مهمترین عامل ایجاد انگیزه در محققان است که دست‌ها را برای بهبود تعامل بین انسان و رایانه مورد استفاده قرار دهند. در این مقاله جهت تشخیص حرکت دست، با استفاده از اختلاف فریم‌ها، دست به عنوان تنها شی متحرک در تصاویر جدا شده است. پس از آن، بردار ویژگی حرکت دست استخراج شده است. این بردار ویژگی به عنوان ورودی شبکه عصبی مصنوعی جهت طبقه‌بندی حرکات انتقالی دست استفاده شده است. برای استخراج بردار ویژگی دو روش ارائه شده است. بردار ویژگی اول حاصل از کدگذاری خط سیر انتهایی‌ترین پیکسل دست در طی فریم‌ها است. بردار ویژگی دوم، از دو هیستوگرام زاویه‌ای برای کدگذاری دست استفاده می‌کند. شناسایی شش حرکت مختلف دست در داده‌های آزمایشی با نرخ بازشناسی ۹۵/۵۴ درصد با استفاده از بردار ویژگی اول و تشخیص این حرکات با نرخ بازشناسی ۹۱/۵۳ درصد توسط بردار ویژگی دوم، کارایی بردارهای ویژگی پیشنهادی در تشخیص حرکات دست را نشان می‌دهد. همچنین مقایسه‌ی بردارهای ویژگی پیشنهادی با بردار ویژگی یک روش متداول، برتری روش‌های پیشنهادی را از نظر دقت، تعداد ویژگی‌ها و زمان آموزش طبقه‌بند نشان می‌دهد.

کلیدواژه‌ها

تشخیص حرکات انتقالی دست، تشخیص دست، کدگذاری خط سیر حرکت دست، هیستوگرام زاویه‌ای، شبکه عصبی مصنوعی

۱ مقدمه

پیشرفت تکنولوژی‌های جدید در زمینه‌ی تعامل انسان و رایانه بسیار چشم‌گیر است. مطالعات بسیاری در این زمینه با هدف بهبود رابطه‌ی انسان و رایانه انجام شده است. تکنولوژی‌های اخیر به سمت تعامل انسان با استفاده از حرکات بدن گرایش دارند. یکی از راه‌هایی که کاربر می‌تواند برای ارتباط با رایانه مورد استفاده قرار دهد، استفاده از حرکات دست است.

تشخیص حرکات دست^۱، امکان ارتباط شهودی و طبیعی بین انسان و رایانه را فراهم می‌کند. حرکات دست انواع مختلفی دارد. این حرکات می‌توانند حرکات انتقالی دست در طی فریم‌ها [۱] تا [۶] یا توالی حالات دست [۷] تا [۱۰] باشند. تشخیص حرکت دست بر پایه‌ی بینایی ماشین به دلیل محاسبات پیچیده و درجه‌ی زیاد آزادی در حرکت دست، مسئله‌ای چالش برانگیز است.

در زندگی روزمره، از دیدن و شنیدن به عنوان منابع اصلی دریافت اطلاعات از محیط پیرامون استفاده می‌شود. بنابراین، با گسترش ابزار رایانه‌ای، نیاز به ابزارهای جدیدی برای ارتباط انسان و رایانه هست که به‌طور مشابه از ورودی‌های بصری یا شنیداری استفاده کنند.

این مقاله در بهمن‌ماه سال ۱۳۹۴ دریافت، در تیرماه سال ۹۵ بازنگری و در مردادماه همان سال پذیرفته شد.

^۱ دانشکده برق و کامپیوتر، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته کرمان. رایانامه: f.azadi@student.kgut.ac.ir

^۲ دانشکده برق و کامپیوتر، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته کرمان. رایانامه: e.rashedi@kgut.ac.ir

نویسنده مسئول: عصمت راشدی

¹Hand gesture recognition

روش‌های تشخیص حرکات‌های انتقالی دست به دو دسته روش‌های بر پایه‌ی مدل و روش‌های بر پایه‌ی نمونه تقسیم‌بندی می‌شوند [۱۲]. روش‌های بر پایه‌ی مدل شامل مدل مخفی مارکوف^۴ [۱۳] تا [۱۷]، ماشین‌های حالت متناهی [۱۸] و [۱۹] و روش‌های بر پایه‌ی نمونه شامل توصیف‌کننده‌های مسیر حرکت [۲۰] هستند.

روش پیشنهادی این مقاله از جمله روش‌های بر پایه‌ی نمونه است. هدف این مقاله، شناسایی حرکات انتقالی دست در ویدیو-های دریافتی است. تعدادی از الگوریتم‌های پیشنهادی جهت تشخیص حرکت دست در مقالات اخیر به گونه‌ای طراحی شده‌اند که قادر به شناسایی انواع مختلف حرکات انتقالی نیستند [۴] و [۱۱]. همچنین بعضی از روش‌های پیشنهادی به منظور شناسایی حرکات دست [۲] در مواردی مانند: تغییر در ابعاد حرکت، تفاوت در ابعاد تصویر، تغییر در نقطه شروع حرکت و جابه‌جایی موقعیت مکانی اجرای حرکت در تصویر، توانایی تشخیص صحیح حرکات دست را ندارند.

الگوریتم پیشنهادی این مقاله به گونه‌ای طراحی شده است که قادر به شناسایی حرکات انتقالی دست به طرفین در محیط‌های بسته و در شرایط ثابت بودن دوربین می‌باشد. در این روش، کدگذاری خط سیر حرکت دست و شبکه عصبی مصنوعی^۵ (ANN) به منظور تشخیص حرکات مختلف دست مورد استفاده قرار گرفته‌اند. دو بردار ویژگی جهت کدگذاری حرکت پیشنهاد شده است. روشاویل بر مبنای کدگذاری خط سیر حرکت است که تاکنون جهت تشخیص حرکات دست استفاده نشده است. در این روش، تغییر در ابعاد حرکت، تفاوت در ابعاد تصویر و جابه‌جایی موقعیت مکانی اجرای حرکت در تصویر، خللی در تشخیص صحیح حرکت دست ایجاد نمی‌کند. روش پیشنهادی دوم مبتنی بر هیستوگرام‌های زاویه‌ای است و علاوه بر تغییرناپذیری بردار ویژگی نسبت به ابعاد تصویر، ابعاد حرکت، و جابه‌جایی موقعیت مکانی اجرای حرکت در تصویر، نسبت به نقطه شروع و جهت حرکت نیز تغییرناپذیر است.

ادامه‌ی این مقاله به این نحو ساماندهی شده است: در بخش دوم روش پیشنهادی جهت شناسایی حرکات مختلف دست آورده شده است. این بخش شامل چگونگی جداسازی دست در هر فریم، چگونگی استخراج بردارهای ویژگی و طبقه‌بندی حرکات مختلف دست با استفاده از شبکه عصبی مصنوعی است. در بخش سوم، نتیجه‌ی اعمال بردارهای ویژگی پیشنهادی بر روی ویدیوهای مورد نظر و مقایسه‌ی این ویژگی‌ها با ویژگی‌های پیشنهادی در مرجع [۶] آورده شده است. در نهایت بخش چهارم، جمع‌بندی مقاله را در بر می‌گیرد.

در [۱] به منظور تعامل شهودی و طبیعی بین انسان و رایانه، از ضبط حرکت دست به‌وسیله‌ی دوربین به جای استفاده از موشواره استفاده شده است. این سیستم از تشخیص رنگ پوست برای جداسازی دست بهره برده است. پس از آن، الگوریتم خطوط و بدنه محدب^۱، برای تشخیص تعداد انگشتان دست به کار رفته است. از آنجاکه در این سیستم نشانه‌گر موشواره حرکت دست کاربر را دنبال می‌کند، نیاز به آشکارسازی حرکت دست است. به این منظور با دنبال کردن مکان نقطه‌ی مرکزی مستطیل احاطه‌کننده دست، موقعیت نشانه‌گر موشواره تعیین می‌شود.

در سیستم پیشنهادی [۲]، اسکلت دست و ساعد کاربر با استفاده از اطلاعات عمق و رنگ تشخیص داده شده است. پس از آن جابه‌جایی دست و ساعد با مجموعه جابه‌جایی‌های پایگاه داده مقایسه و حرکت دست شناسایی شده است. در [۳]، تشخیص دست و حرکت آن تنها با استفاده از محدودیت‌های هندسی و رنگ انجام شده است. در این سیستم از فیلتر کالمن برای ردیابی دست استفاده شده است. کدگذاری حرکات دست براساس مجموعه‌ای از حرکات پایه در سیستم [۴] انجام شده است. این سیستم تنها برای تشخیص حرکاتی که ترکیبی از حرکات پایه‌ی تعریف شده باشند، مناسب است و قابلیت تشخیص هر نوع حرکت انتقالی را ندارد.

[۵] طبقه‌بند آبخاری^۲ را برای تعیین موقعیت دست و طبقه‌بندی حالات دست مورد استفاده قرار داده است. این سیستم از روش camshift^۳ به منظور ردیابی دست بهره برده است. لین [۶] پس از آشکارسازی دست با استفاده از ویژگی‌های HOG^۳ و ماشین بردار پشتیبان، به شناسایی حرکت دست پرداخته است. در این سیستم، موقعیت دست و سرعت و زاویه‌ی حرکت، برداری چهاربعدی را برای ارائه‌ی حرکت دست تشکیل می‌دهند.

در [۱۰]، ابتدا حالت دست با استفاده از ویژگی‌های هندسی و یک مدل ویژگی حالت دست فازی تشخیص داده شده است و سپس براساس توالی حالت‌های دست، حرکت دست کاربر شناسایی شده است. در سیستم [۱۱]، دستور زبانی ارائه شده است که ۱۶۰ حرکت دست را تفسیر می‌کند. در این سیستم راهکارهایی برای تشخیص توالی حالات دست و شش حرکت انتقالی دست پیشنهاد شده است. دستورات حرکت از سه طریق ایجاد شده‌اند. اولین راه ایجاد حرکت با تغییر حالت دست به حالتی دیگر است. راه دوم دنبال کردن مسیر جابه‌جایی هر حالت است که شامل بالا، پایین، چپ و راست می‌شود. آخرین راه، مشاهده‌ی اندازه یا مقیاس حالت دست است شامل مواردی که دست از دوربین دور می‌شود (کوچک‌نمایی) و یا هنگامی که به دوربین نزدیک می‌شود (بزرگ‌نمایی).

⁴Hidden Markov

⁵Artificial Neural Network

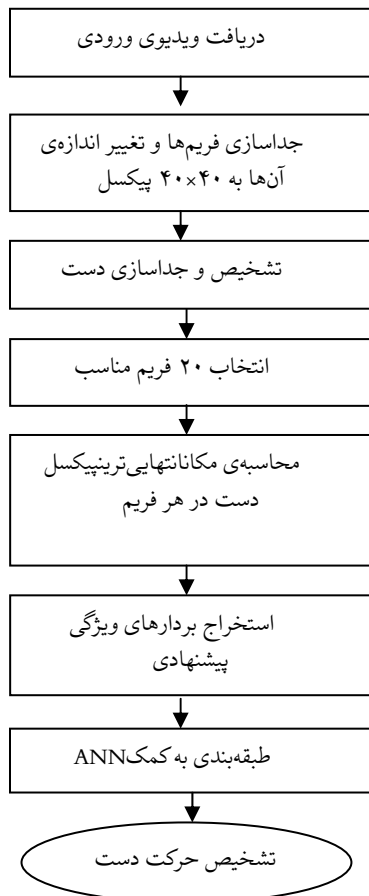
⁶Motion trajectory coding

¹contours and convex hull

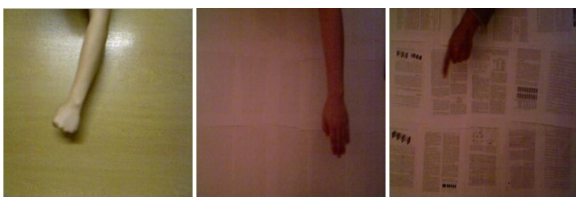
²Haar

³Histogram of Oriented Gradients

باشند. در این مقاله، روشی کم‌هزینه و با محاسبات کم به منظور جداسازی دست در هر فریم مورد استفاده قرار گرفته است.



شکل ۱ نمودار بلوکی روش پیشنهادی به منظور تشخیص حرکات انتقالی دست.



شکل ۲ نمونه‌هایی از تصاویری که روش‌های تشخیص رنگ پوست قادر به شناسایی دست در آن‌ها نمی‌باشند.

تعدادی از نمونه‌های مورد بررسی، دارای شرایط نوری نامناسب می‌باشند اما با توجه به عدم تغییر شرایط نوری در حین هر حرکت و نیز ثابت بودن دوربین در زمان فیلم‌برداری و با توجه به این‌که در این نمونه‌ها دست تنها شی متحرک در تصویر است، از اختلاف فریم‌ها برای تشخیص و جداسازی دست استفاده شده است [۲۴]. شکل ۳ چگونگی انجام این فرآیند را نشان می‌دهد. به این منظور، هر فریم به یک تصویر خاکستری تبدیل شده است. سپس اختلاف هر فریم با دو فریم بعد از آن محاسبه شده است. پس از آن دو مقدر آستانه‌ی سطوح خاکستری در دو تصویر

۲ سیستم پیشنهادی

نمودار بلوکی روش پیشنهادی به منظور تشخیص حرکات انتقالی دست در شکل ۱ نشان داده شده است. در این روش پس از جداسازی فریم‌ها و تغییر اندازه‌ی فریم‌ها به 40×40 پیکسل، تشخیص و جداسازی دست در هر فریم انجام می‌شود. در بسیاری از کاربردهای تشخیص حرکات دست در تعامل با رایانه، دوربین روی تصویر دست تنظیم می‌شود. بنابراین در اکثر مواقع دست تنها شی در ویدیوهای دریافتی است. با توجه به این نکته که دست تنها شی متحرک در تصاویر است، از اختلاف فریم‌ها به منظور جداسازی دست استفاده شده است.

در این روش، ۲۰ فریم مناسب در دنباله ویدیویی انتخاب می‌شوند. به این ترتیب که اگر تعداد فریم‌های حاصل از فرآیند-های فوق برابر با n باشد، مقدار $t = \lfloor n/20 \rfloor$ را محاسبه و از t فریم یک فریم به عنوان نماینده انتخاب می‌شود.

پس از آن، از کدگذاری حرکت به منظور ایجاد بردارهای ویژگی جهت تشخیص حرکات انتقالی دست استفاده شده است. به این منظور در ابتدا مکان انتهایی‌ترین پیکسل دست شناسایی شده است. از آنجا که در کاربردهای مورد مطالعه این مقاله، اشکال توسط انگشتان دست رسم می‌شود، این نقطه نماینده مناسبی جهت دنبال کردن شکل رسم شده توسط دست می‌باشد. در این مقاله دو روش جهت کدگذاری حرکت پیشنهاد شده است. در روش پیشنهادی اول، جابه‌جایی انتهایی‌ترین پیکسل دست در هر دو تصویر متوالی کدگذاری شده است. کد بدست آمده، به منظور آموزش شبکه عصبی مصنوعی برای شناسایی حرکت دست مورد استفاده قرار گرفته است.

در روش پیشنهادی دوم، زاویه‌ی خط‌گذرنده از هر دو نقطه متوالی از مکان انتهایی‌ترین پیکسل دست نسبت به محور افقی و زاویه‌ی خط‌گذرنده از این نقاط و مرکز ثقل مسیر حرکت محاسبه شده است. هیستوگرام این زوایا به عنوان ورودی شبکه عصبی مصنوعی جهت تشخیص حرکات انتقالی دست، به کار گرفته شده است. جزییات هر قسمت در ادامه آورده شده است.

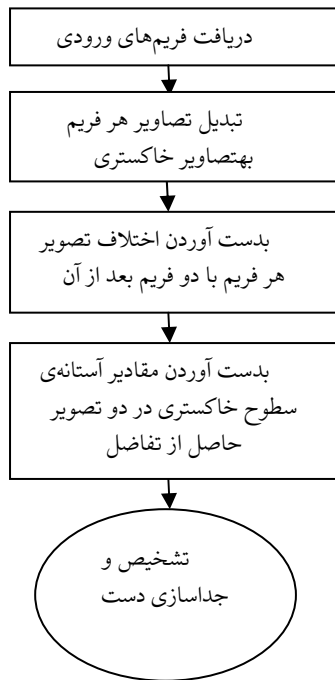
۲-۱ تشخیص و جداسازی دست

تشخیص رنگ پوست می‌تواند به عنوان یک ویژگی مهم در تشخیص و ردیابی دست انسان مورد استفاده قرار بگیرد [۲۱] تا [۲۳]. اما به دلیل شرایط نوری نامناسب و نزدیک بودن رنگ زمینه به رنگ پیکسل‌های ناحیه‌ی دست در تعدادی از نمونه‌های داده‌ی استفاده شده، به منظور جداسازی دست در این گونه نمونه‌ها روش تشخیص رنگ پوست کارآمد نیست. تعدادی از این تصاویر در شکل ۲ آورده شده‌اند.

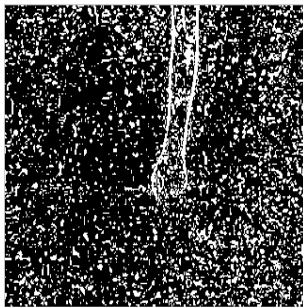
برای اشیای متحرک روش‌های دنبال‌سازی^۱ قابل استفاده هستند [۳] و [۵]. اما این روش‌ها بسیار پیچیده و هزینه‌بر می-

^۱Tracking

انگشتان دست) به عنوان نقطه مرجع انتخاب شده است. در صورتی که مکان ورود دست در تصویر نامشخص باشد، با توجه به اینکه ساعد دست در حاشیه تصویر قرار می‌گیرد، می‌توان مکان ورود دست را تعیین کرد و با چرخش تصویر، ساعد دست را در حاشیه‌ی بالای تصویر قرار داد. در این حالت، انتهایی‌ترین پیکسل دست را می‌توان به عنوان نقطه‌ی مرجع به منظور تعیین حرکت دست انتخاب کرد (شکل ۶). توجه به این نکته حائز اهمیت است که حالت انگشتان در طول حرکت انتقالی دست باید ثابت باشد. در غیر این صورت با دو نوع حرکت دست به صورت هم‌زمان مواجه خواهیم بود که موضوع مورد بررسی این مقاله نمی‌باشد. از طرفی در ارتباط با رایانه، کاربر حرکات موردنظر دست را با حرکت انگشتان دست ترسیم می‌کند. بنابراین انتخاب انتهایی‌ترین پیکسل دست برای تعیین حرکات انتقالی دست مناسب است.



شکل ۳ چگونگی جداسازی دست به عنوان تنها شی متحرک در تصویر.



شکل ۴ نمونه‌ای از تصاویر با نویز شدید حاصل از تفاضل فریم‌ها.

تفاضلی، با هدف تشخیص محدوده‌ی دست به عنوان شی متحرک، مورد استفاده قرار گرفته‌اند. این مقادیر آستانه با استفاده از روش اتسوا^۱ محاسبه شده‌اند. رابطه‌ی ۱، نحوه‌ی جداسازی دست را نشان می‌دهد.

$$\text{IF } [(I^t(i,j) - I^{t+1}(i,j)) \geq \text{Th1 or } (I^t(i,j) - I^{t+2}(i,j)) \geq \text{Th2}] \text{ then } (i,j) \in \text{hand} \quad (1)$$

که در آن، $I^t(i,j)$ مقدار سطح خاکستری در سطر i ام و ستون j ام در فریم مربوطه، $I^{t+1}(i,j)$ مقدار سطح خاکستری در سطر i ام و ستون j ام در فریم بعد از فریم مربوطه و $I^{t+2}(i,j)$ مقدار سطح خاکستری در سطر i ام و ستون j ام در دو فریم بعد از فریم مربوطه است. Th1 مقدار آستانه‌ی سطوح خاکستری در تصویر حاصل از تفاضل فریم مربوطه و فریم بعد از آن و Th2 مقدار آستانه‌ی سطوح خاکستری در تصویر حاصل از تفاضل فریم مربوطه و دو فریم بعد از آن است.

تا این مرحله، دست به عنوان تنها شی متحرک در تصویر، جدا شده است. فریم‌های ابتدا و انتهای ویدیوهای مورد استفاده، به دلیل قطع و وصل دوربین دارای کیفیت مناسبی نیستند. بنابراین بعضی از تصاویر حاصل از تفاضل فریم‌ها (نظیر شکل ۴) دچار نویز شدید هستند. با توجه به این که برای تشخیص حرکت دست از دنباله تصاویر نمونه‌برداری شده است، حذف این تصاویر خلی در جهت تشخیص حرکت دست ایجاد نمی‌کند. بنابراین تصاویر با نویز شدید، با این شرط که اگر فاصله‌ی بین چپ‌ترین پیکسل و راست‌ترین پیکسل در تصاویر حاصله بیشتر از نصف عرض تصویر باشد، حذف شده‌اند. این شرط با توجه به نسبت پهنا‌ی دست به عرض تصویر در نظر گرفته شده است. با حذف فریم‌های نامناسب، از هر ویدیو شامل حرکت کامل دست، ۲۰ فریم برای کدگذاری انتخاب می‌شوند.

در هر فریم شامل پیکسل‌های متحرک، با حذف مجموعه پیکسل‌های دارای مساحت کمتر از ۱۰ پیکسل، نویزهای جزئی نشان داده شده در تصویر شکل ۵ حذف می‌شوند. به این ترتیب که مساحت هر مجموعه پیکسل آشکار شده به عنوان شی متحرک در تصویر محاسبه و پس از آن، مجموعه پیکسل‌های با مساحت کمتر از ۱۰ پیکسل حذف شده‌اند. با این روش، نویزهای موجود در تصویر که باعث خطا در انتخاب انتهایی‌ترین پیکسل ناحیه‌ی دست می‌شوند، حذف می‌گردند.

۲-۲ استخراج بردارهای ویژگی پیشنهادی

برای تشخیص حرکت دست، به نقطه‌ی مرجعی نیاز است که جا-به‌جایی آن نقطه را بتوان به عنوان جابه‌جایی کل ناحیه‌ی دست در نظر گرفت. در این مقاله، انتهایی‌ترین پیکسل دست (مکان

¹Otsu

مقاله پیشنهاد می‌شود از اطلاعات کد جابه‌جایی استفاده شود. مقادیر واقعی جابه‌جایی در یک حرکت خاص، با تغییر در ابعاد حرکت، تغییر ابعاد تصویر و جابه‌جایی موقعیت مکانی اجرای حرکت در تصویر، متفاوت خواهند شد. در نتیجه با استفاده از خود مقادیر جابه‌جایی، تشخیص حرکت دست در شرایط ذکر شده به درستی انجام نخواهد شد. اما با استفاده از مقادیر جا-به‌جایی گذشته، می‌توان حرکت دست را در شرایط ذکر شده نیز تشخیص داد.

۲-۲-۲ بردار ویژگی دوم- هیستوگرام زاویه‌ای

بردار ویژگی اول به‌گونه‌ای است که تغییر در نقطه‌ی شروع حرکت یا تغییر در جهت حرکت (ساعت‌گرد یا پادساعت‌گرد) باعث تغییر در بردار ویژگی استخراجی از یک حرکت خاص و در نتیجه تشخیص نادرست حرکت می‌شود. بنابراین بردار ویژگی دیگری ارائه می‌شود که علاوه بر اینکه نسبت به تغییر در ابعاد تصویر و تغییر در ابعاد و مکان حرکت تغییرناپذیر است، نسبت به نقطه شروع و جهت حرکت نیز تغییرناپذیر خواهد بود.

با توجه به اینکه نقطه‌ی شروع و جهت حرکت در توالی نقاط استخراجی تاثیری نخواهد داشت، زاویه‌ی خط‌گذرنده از هر دو نقطه‌ی متوالی از مکان انتهایی‌ترین پیکسل دست نسبت به محور افقی، محاسبه شده است. به دلیل اینکه ترتیب این زوایا با تغییر در نقطه‌ی شروع و جهت حرکت متفاوت خواهد شد، هیستوگرام این زوایا به عنوان ویژگی پیشنهاد شده است. علاوه بر این هیستوگرام، هیستوگرام زاویه‌ی خط‌گذرنده از مکان انتهایی‌ترین پیکسل در هر تصویر و مرکز ثقل حرکت نیز در نظر گرفته شده است. مقادیر حاصل از این دو هیستوگرام در کنار هم، بردار ویژگی پیشنهادی را تشکیل می‌دهند. مرکز ثقل حرکت با استفاده از رابطه‌ی ۲ محاسبه شده است.

$$x_m = 1/20 \sum_{i=1}^{20} x_i(2)$$

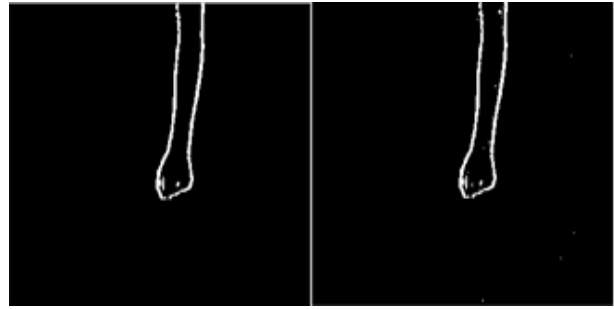
$$y_m = 1/20 \sum_{i=1}^{20} y_i$$

که در این رابطه، x_m و y_m مختصات مرکز ثقل حرکت x_i و y_i مختصات انتهایی‌ترین پیکسل دست در تصویر i ام است. نمایشی از این زوایا برای یک حرکت دلخواه در شکل ۸ نشان داده شده است. با توجه به اینکه برای هر هیستوگرام هشت بخش در نظر گرفته شده است، طول این بردار ویژگی برای هر نمونه برابر ۱۶ خواهد شد.

۲-۳ طبقه‌بندی حرکت دست به کمک شبکه عصبی

مصنوعی

در این قسمت، از شبکه عصبی مصنوعی برای تشخیص حرکت دست استفاده شده است. این طبقه‌بند برای آموزش شش دسته‌ی



(الف) (ب)

شکل ۵ (الف) تصویر نشان‌دهنده اشیاء متحرک همراه با نویز (ب) حذف نویز با حذف مجموعه پیکسل‌های با مساحت کمتر از ۱۰ پیکسل.



شکل ۶ نمایش انتهایی‌ترین پیکسل مربوط به ناحیه‌ی دست.

۱-۲-۲ بردار ویژگی اول - کدگذاری خط سیر حرکت دست

با استفاده از جابه‌جایی انتهایی‌ترین پیکسل دست در تصاویر، کدی بدست می‌آید که از این کد برای آموزش شبکه عصبی مصنوعی به عنوان طبقه‌بند حرکت‌های مختلف دست، استفاده شده است. با توجه به مؤلفه‌های x و y مکان این پیکسل در دو تصویر متوالی، کدگذاری انجام می‌شود.

به این ترتیب که در هر تصویر اگر مؤلفه‌ی x انتهایی‌ترین پیکسل دست کمتر از مؤلفه‌ی x این پیکسل در تصویر قبل باشد، عدد -1 ، اگر بیشتر باشد عدد 1 و اگر برابر باشد عدد 0 را به عنوان کد در نظر می‌گیریم. در صورتی‌که این کار را برای تمام تصاویر انجام دهیم، با توجه به اینکه تعداد تصاویر را به 20 تصویر محدود خواهیم کرد، برداری به طول 19 بدست می‌آید که کدگذاری جابه‌جایی دست در راستای مؤلفه‌ی x محسوب می‌شود.

با انجام همین روند برای مؤلفه‌ی y ، برداری به طول 19 به عنوان کد جابه‌جایی دست در راستای مؤلفه‌ی عمودی بدست می‌آید. با قرار دادن این دو بردار در کنار هم، برداری به طول 38 بدست خواهد آمد که به عنوان ورودی شبکه عصبی مصنوعی برای تشخیص حرکت دست مورد استفاده قرار می‌گیرد. نمونه‌ای از کدگذاری جابه‌جایی برای حرکتی دلخواه در شکل ۷ نشان داده شده است. در روش‌های پیشین، از خود مقادیر جابه‌جایی به منظور شناسایی حرکت دست استفاده شده است [۲]. در این

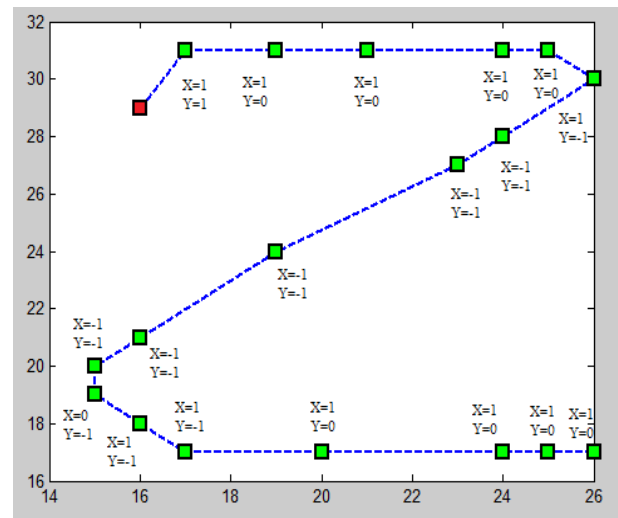
حرکت چپ و راست، حرکت موجی و حرکت z دست، جهت بررسی کارایی بردارهای ویژگی پیشنهادی مورد استفاده قرار گرفته‌اند. سایر حرکات این پایگاه شامل نوازش، بیا-اینجا، بالا-پایین و چرخش مچ، مربوط به دور و نزدیک شدن دست به دوربین می‌باشند. به عنوان مثال در حرکت بالا-پایین به دلیل عمود بودن دوربین به حرکت دست، مسیر حرکت توسط دوربین قابل مشاهده نیست. جهت شناسایی این حرکات نیاز به اطلاعات فاصله دست از دوربین است که توسط سنسورهای عمق سنج محاسبه می‌شود و موضوع بحث این مقاله نیست. روش دیگر استفاده از دوربین دوم جهت تصویربرداری از پهلو است. در این مقاله هدف بررسی کارایی بردارهای کدگذاری حرکت پیشنهادی در شناسایی حرکات به طرفین دوربین است. بنابراین تنها شش حرکت انتخاب شده است.

نمایی از این شش حرکت در شکل ۹ نشان داده شده است. این داده‌ها از شش فرد جمع‌آوری شده‌اند و دارای دو شرایط نوری مختلف، سه پس‌زمینه و شامل سه حالت باز، بسته و اشاره دست می‌باشند. ۶۴۸ فیلم از این داده، مورد استفاده قرار گرفته است که برای هر حرکت ۱۰۸ فیلم نمونه در نظر گرفته شده است. ۸۰٪ این نمونه‌ها برای آموزش شبکه‌ی عصبی مصنوعی و ۲۰٪ به عنوان داده‌ی آزمایشی مورد استفاده قرار گرفته‌اند.

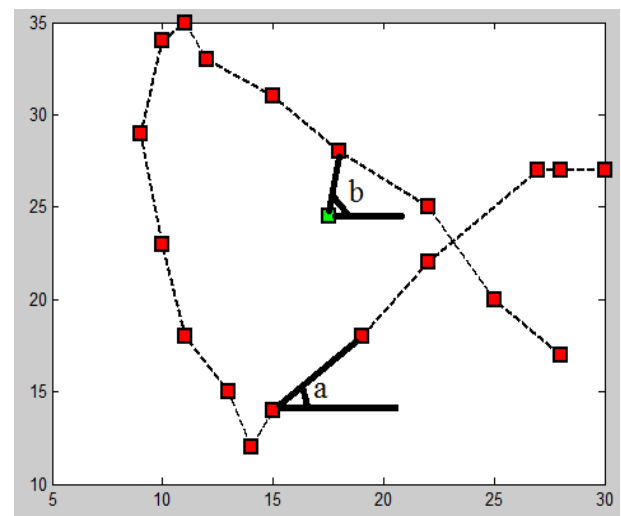
بر اساس روش شرح داده شده در شکل ۱، پس از انتخاب ۲۰ فریم از ویدیوی دریافتی و جداسازی دست در هر فریم، بردارهای ویژگی پیشنهادی از تصاویر دست استخراج شده و به عنوان ورودی شبکه عصبی مصنوعی مورد استفاده قرار گرفته‌اند. تعداد فریم‌های انتخابی از ویدیوی دریافتی، به صورت تجربی در نظر گرفته شده است. در صورتی که تعداد فریم‌های انتخابی کمتر از ۲۰ فریم باشد، دقت سیستم پیشنهادی کاهش می‌یابد و اگر تعداد فریم‌های نماینده را بیش از ۲۰ فریم انتخاب کنیم، دقت سیستم تغییر نمی‌کند. این در حالی است که انتخاب بیش از ۲۰ فریم به عنوان فریم‌های نماینده، سرعت سیستم را کاهش می‌دهد. از طرفی اگر تعداد فریم‌های انتخابی از حدی بیشتر شوند، به دلیل تغییرات اندک جابه‌جایی در نمونه‌های مختلف از یک حرکت خاص، دقت سیستم کاهش می‌یابد.

در بردار ویژگی اول، طول بردار کد استخراج شده برای هر حرکت ۳۸ واحد است که از کدگذاری جابه‌جایی انتهایی‌ترین پیکسل دست در دو راستای افقی و عمودی بدست آمده است. با توجه به اینکه برای هر هیستوگرام هشت قسمت در نظر گرفته شده است، دومین بردار ویژگی پیشنهادی شامل ۱۶ واحد است. این بردارهای ویژگی به عنوان ورودی شبکه عصبی مصنوعی با ۲۵ نرون در لایه میانی و ۶ نرون در لایه خروجی مورد استفاده قرار می‌گیرند. بعد از آموزش ANN، دسته‌ی داده‌های آزمایشی با توجه به خروجی ANN مشخص می‌شود. نرون با خروجی بیشتر، مشخص‌کننده دسته‌ی داده‌ی آزمایشی است.

مختلف با استفاده از بردارهای ویژگی پیشنهادی به عنوان ورودی طراحی شده است. جهت آموزش طبقه‌بند از ویدیوهای آموزشی استفاده می‌شود. از روش یادگیری پس انتشار خطا برای آموزش شبکه استفاده شده است.



شکل ۷ نحوه‌ی کدگذاری حرکت Z دست در روش اول. نقطه قرمز رنگ نقطه شروع حرکت است.



شکل ۸ نمایش زوایای مورد استفاده در بردار ویژگی دوم در نمونه‌ای از حرکت ضربدری دست. نقطه سبزرنگ نمایش مرکز نقل حرکت است. زاویه‌ی a نمونه‌ای از زاویه‌ی خط گذرنده از دو نقطه متوالی از مکان انتهایی‌ترین پیکسل دست و زاویه‌ی b نمونه‌ای از زاویه‌ی خط گذرنده از مکان انتهایی‌ترین پیکسل در هر تصویر و مرکز نقل مسیر حرکت است.

۳ آزمایش‌ها و نتایج

برای بررسی کارایی بردارهای ویژگی پیشنهادی در تشخیص حرکات انتقالی دست، از ویدیوهای پایگاه داده [۲۵] استفاده شده است. این داده شامل ده حرکت مختلف دست است که شش حرکت شامل حرکت دایره‌ای، حرکت مثلثی، حرکت ضربدری،

استفاده از ویژگی‌های ارائه شده در [۶] است. از طرفی بردارهای ویژگی پیشنهادی تعداد ویژگی کمتری نسبت به ویژگی‌های [۶] دارند که این امر موجب کاهش فضای اشغالی بر روی مدارات مجتمع خواهد شد. علاوه بر این موارد، روش‌های پیشنهادی این مقاله بر خلاف روش پیشنهادی [۴]، قابلیت شناسایی انواع مختلفی از حرکات انتقالی را دارد.

بردارهای ویژگی پیشنهادی این مقاله، علاوه بر افزایش دقت و کاهش پیچیدگی، قابلیت تشخیص صحیح حرکت دست با وجود تغییر در ابعاد حرکت، تفاوت در ابعاد تصاویر و جابه‌جایی موقعیت مکانی اجرای حرکت در تصویر را دارند.

بردار ویژگی دوم نسبت به بردار ویژگی اول نرخ بازشناسی کمتری را دارا است، اما این بردار مزایایی نسبت به بردار ویژگی پیشنهادی اول دارد که موجب می‌شود در بعضی از کاربردها، استفاده از آن مناسب‌تر باشد. از جمله برتری‌های بردار ویژگی دوم نسبت به بردار ویژگی اول، قابلیت تشخیص حرکت با تغییر در جهت و نقطه شروع حرکت، تعداد ویژگی‌های کمتر و در نتیجه کاهش زمان آموزش است.

بنابراین در کاربردهایی که جهت و نقطه شروع حرکت قابل تنظیم است، بردار ویژگی اول و در صورت نامشخص بودن جهت و نقطه شروع حرکت، بردار ویژگی دوم کارآمدتر خواهد بود.

۴ جمع‌بندی

در این مقاله تشخیص و جداسازی دست به عنوان تنها شی متحرک در تصویر با استفاده از اختلاف فریم‌ها انجام شد. از کدگذاری حرکات دست به منظور استخراج بردار ویژگی و تشخیص حرکت توسط آموزش شبکه عصبی استفاده شده است. دو بردار ویژگی به منظور تشخیص حرکات انتقالی دست ارائه شد. در روش اول از کدگذاری انتهایی‌ترین پیکسل دست استفاده شده است. ایده این روش، استفاده از کدگذاری حرکات دست به جای موقعیت مکانی دست است. به‌کارگیری این بردار ویژگی به عنوان ورودی شبکه عصبی مصنوعی به منظور شناسایی شش حرکت مختلف دست، نرخ بازشناسی ۹۷/۹۵٪ برای داده‌های آموزشی و ۹۵/۵۴٪ برای داده‌های آزمایشی را نتیجه داده است. این نتیجه کارایی روش پیشنهادی در تشخیص حرکات مختلف دست را نشان می‌دهد. این بردار ویژگی نسبت به بردار ویژگی شامل موقعیت مکانی دست و سرعت و زاویه حرکت، علاوه بر اینکه طول کمتری دارد، نرخ بازشناسی بهتری را نیز ارائه می‌دهد. کاهش تعداد ویژگی‌ها باعث کاهش زمان لازم جهت استخراج ویژگی و کاهش زمان آموزش طبقه‌بند می‌شود. از طرفی بردار ویژگی پیشنهادی به‌گونه‌ای است که با وجود تغییر در ابعاد حرکت، تفاوت در ابعاد تصاویر و جابه‌جایی موقعیت مکانی اجرای حرکت در تصویر، قادر به تشخیص صحیح حرکت دست خواهد بود.

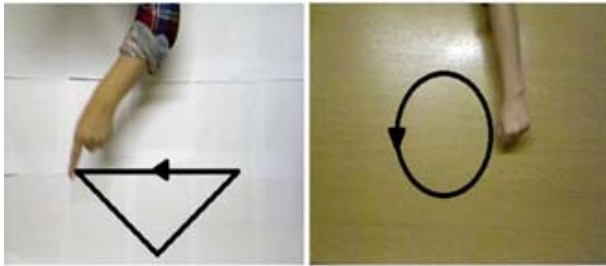
در بردار ویژگی دوم، دو هیس‌توگرام زاویه‌ای جهت استخراج ویژگی مورد استفاده قرار گرفت. این بردار ویژگی اگرچه نرخ

لازم است شبکه عصبی مصنوعی روی داده‌های آموزشی، آموزش داده شود. به منظور نشان دادن کارایی بردارهای ویژگی پیشنهادی در تشخیص حرکت دست، شبکه پنج بار به صورت مستقل آموزش داده شده و نتایج بازشناسی داده‌های آزمایشی پس از هر بار آموزش محاسبه شده است. متوسط این نتایج جهت بررسی و مقایسه محاسبه شده است که بردار ویژگی اول، متوسط نرخ بازشناسی ۹۷/۹۵٪ برای داده‌های آموزشی و ۹۵/۵۴٪ برای داده‌های آزمایشی و بردار ویژگی دوم، متوسط نرخ بازشناسی ۹۶/۹۱٪ برای داده‌های آموزشی و ۹۱/۵۳٪ را برای داده‌های آزمایشی نتیجه می‌دهند. این نتایج، کارایی بردارهای ویژگی پیشنهادی را در تشخیص حرکات انتقالی دست نشان می‌دهند.

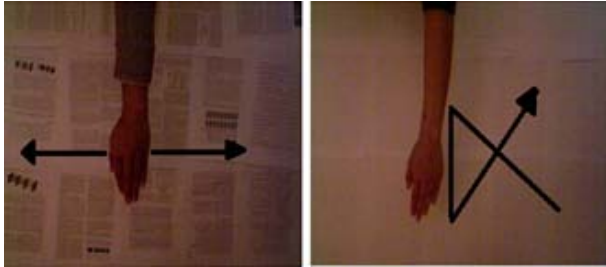
تحلیل ماتریس اغتشاش برای نرخ بازشناسی داده‌های آزمایشی با استفاده از شبکه عصبی مصنوعی و بردار ویژگی پیشنهادی اول در جدول ۱ نشان داده شده است. بر اساس این جدول، بیشترین نرخ طبقه‌بندی نادرست بین دو حرکت موجی دست و چپ و راست دست، به دلیل شباهت این دو حرکت به یکدیگر است. این امر موجب شده است که کمترین نرخ طبقه‌بندی درست مربوط به حرکت موجی دست باشد. همان‌طور که در جدول نشان داده شده است، سه حرکت z، مثلثی و ضربدری دست به دلیل تفاوت با دیگر حرکات‌ها، بالاترین نرخ بازشناسی صحیح را دارا می‌باشند.

تحلیل ماتریس اغتشاش برای نرخ بازشناسی داده‌های آزمایشی با استفاده از شبکه عصبی مصنوعی و بردار ویژگی پیشنهادی دوم در جدول ۲ نشان داده شده است. بر اساس این جدول، بیشترین نرخ طبقه‌بندی نادرست بین حرکت موجی دست و حرکت چپ و راست دست است. به همین دلیل کمترین نرخ طبقه‌بندی درست مربوط به حرکت چپ و راست دست است. طبق جدول، بردار ویژگی دوم تا حدی حرکت z، حرکت ضربدری و حرکت موجی دست را بهتر از بردار ویژگی اول شناسایی کرده است، اما در حرکات دیگر، نرخ بازشناسی کمتری را نتیجه داده است. بیشترین نرخ بازشناسی صحیح با استفاده از بردار ویژگی دوم، مربوط به حرکت ضربدری دست است.

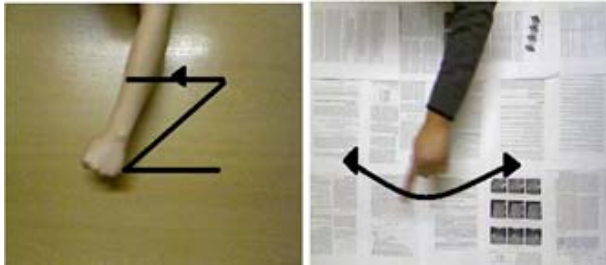
به منظور نشان دادن کارایی بردارهای ویژگی پیشنهادی، مقایسه‌ای بین این ویژگی‌ها و ویژگی‌های پیشنهاد شده در مرجع [۶] انجام شده است. ویژگی‌های پیشنهادی مرجع [۶] جهت شناسایی حرکات مختلف دست شامل موقعیت مکانی دست در دو راستای افقی و عمودی، سرعت و زاویه حرکت دست می‌باشند. جهت بهبود نتایج این مرجع، در روش پیشنهادی اول از کدگذاری حرکت دست به جای موقعیت مکانی استفاده شده است. همان‌طور که در جدول ۳ نشان داده شده است، متوسط نرخ بازشناسی شبکه عصبی مصنوعی پس از پنج آموزش مستقل و با به‌کارگیری بردارهای ویژگی پیشنهادی، در هر دو مورد بیش از متوسط نرخ بازشناسی ANN با استفاده از ویژگی‌های [۶] است. همچنین با به‌کارگیری بردارهای ویژگی پیشنهادی، زمان مصرفی جهت آموزش ANN بسیار کمتر از زمان آموزش ANN با



(الف) (ب)



(پ) (ت)



(ث) (ج)

شکل ۹ نمایی از شش حرکت دست (الف) حرکت دایره‌ای (ب) حرکت مثلثی (پ) حرکت ضربدری (ت) حرکت چپ و راست (ث) حرکت موجی و (ج) حرکت Z دست.

جدول ۳ مقایسه بردار ویژگی پیشنهادی با بردار ویژگی ارائه شده در مرجع [۶] (نتایج متوسط پنج آموزش مستقل شبکه هستند).

نرخ بازشناسی داده‌های آزمایشی ANN (%)	زمان آموزش (sec) ANN	تعداد ویژگی‌ها	زمان استخراج ویژگی از هر نمونه (ms)
۹۵/۵۴	۱۴/۴۳	۳۸	۰/۱۱
۹۱/۵۳	۶/۷۰	۱۶	۰/۴۰
۸۷/۸۳ [۶]	۴۸/۰۸	۷۶	۰/۱۸

مراجع

[1] Hartanto, R., Susanto, A., Santosa P.I., "Real Time Hand Gesture Movements Tracking and Recognizing System", Electrical Power, Electronics, Communications, Controls, and Informatics Seminar (EECCIS), pp. 137-141, 2014.

بازشناسی کمتری را نسبت به بردار ویژگی اول دارد، اما مزیت‌های این بردار ویژگی شامل قابلیت تشخیص حرکت دست با تغییر در نقطه شروع و جهت حرکت و همچنین تعداد ویژگی‌های کمتر، باعث می‌شود که در بعضی از کاربردها، استفاده از این بردار ویژگی مناسب‌تر از بردار ویژگی اول باشد. علاوه بر این، بردار ویژگی دوم همانند بردار ویژگی اول به‌گونه‌ای است که با وجود تغییر در ابعاد و مکان حرکت و تغییر در ابعاد تصویر، قادر به تشخیص انواع حرکات انتقالی خواهد بود.

در ادامه تحقیق، می‌توان ویژگی‌های بیشتری جهت شناسایی حرکات بیشتر به بردارهای فوق اضافه نمود. به این منظور می‌توان از اطلاعات فاصله دوربین‌های دارای سنسور عمق سنج مانند دوربین کینکت، یا از تصاویر چند دوربین جهت استخراج ویژگی‌های مناسب استفاده نمود.

جدول ۱ ماتریس اغتشاش برای نرخ بازشناسی داده‌های آزمایشی با استفاده از بردار ویژگی اول و شبکه عصبی مصنوعی (نتایج متوسط پنج آموزش مستقل شبکه هستند). مقادیر بزرگ‌شده، مقدار بیشینه در هر ستون را نشان می‌دهند.

Z	موجی	چپ و راست	ضربدری	مثلثی	دایره‌ای
۰/۹۰	۰/۹۵	۰/۹۵	۱/۸۶	۰	۹۵/۳۲
۰	۰	۰/۹۰	۰/۹۵	۹۸/۱۳	۰
۲/۷۶	۰	۰	۹۶/۳۱	۰	۰/۹۰
۰	۴/۶۲	۸۹/۹۰	۰	۰/۹۰	۴/۵۴
۰	۸۷/۰۹	۹/۲۶	۰	۰	۳/۶۳
۹۵/۹۹	۰/۸۳	۰	۰/۷۶	۱/۶۳	۰/۷۶

جدول ۲ ماتریس اغتشاش برای نرخ بازشناسی داده‌های آزمایشی با استفاده از بردار ویژگی دوم و شبکه عصبی مصنوعی (نتایج متوسط پنج آموزش مستقل شبکه هستند). مقادیر بزرگ‌شده، مقدار بیشینه در هر ستون را نشان می‌دهند.

Z	موجی	چپ و راست	ضربدری	مثلثی	دایره‌ای
۰/۹۰	۰/۹۵	۲/۸۱	۱/۸۶	۳/۷۲	۸۹/۷۳
۰/۹۰	۰	۰/۹۰	۰/۹۵	۹۱/۷۶	۵/۴۵
۰	۰	۰	۹۷/۲۲	۰/۹۰	۱/۸۶
۱/۸۶	۱۲/۰۷	۸۳/۳۳	۰/۹۰	۰	۱/۸۱
۰/۹۰	۸۷/۱۷	۱۰/۱۲	۰	۰	۱/۸۶
۹۶/۲۳	۰/۹۵	۰	۰	۱/۸۶	۰/۹۵

- Conference on Automatic Face and Gesture Recognition, pp.456-461, 2000.
- [15] Rajko, S., Qian, G., Ingalls, T., James, J., "Real-Time Gesture Recognition with Minimal Training Requirements and on-Line Learning", IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.
- [16] Elgammal, A., Shet, V., Yacoob, Y., Davis, L. S., "Learning Dynamics for Exemplar-Based Gesture Recognition", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 571-578, 2003.
- [17] Wang, S., Quattoni, A., Morency, L. P., Demirdjian, D., Darrell, T., "Hidden Conditional Random Fields for Gesture Recognition", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 1521-1527, 2006.
- [18] Davis, J., Shah, M., "Recognizing Hand Gestures", Computer Vision-ECCV, pp. 331-340, 1994.
- [19] Hong, P., Turk, M., Huang, T. S., "Gesture Modeling and Recognition Using Finite State Machines", Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 410-415, 2000.
- [20] Yang, M. H., Ahuja, N., Tabb, M., "Extraction of 2D Motion Trajectories and Its Application to Hand Gesture Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, pp.1061-1074, 2002.
- [21] Reale, M. J., Canavan, S., Yin, L., Hu, K., Hung, T., "A Multi-Gesture Interaction System Using a 3-D Iris Disk Model for Gaze Estimation and an Active Appearance Model for 3-D Hand Pointing", IEEE Transactions on Multimedia, Vol. 13, pp. 474-486, 2011.
- [22] Huang, D. Y., Hu, W. C., Chang, S. H., "Gabor Filter-Based Hand-Pose Angle Estimation for Hand Gesture Recognition Under Varying Illumination", Expert Systems with Applications, Vol. 38, pp. 6031-6042, 2011.
- [23] Erden, F., Çetin, A. E., "Hand Gesture Based Remote Control System Using Infrared Sensors and a Camera", IEEE Transactions on Consumer Electronics, Vol. 60, pp. 675-680, 2014.
- [24] Zhan, C., Duan, X., Xu, S., Song, Z., Luo, M., "An Improved Moving Object Detection Algorithm Based on Frame Difference and Edge Detection", Fourth International Conference on Image and Graphics, pp. 519-523, 2007.
- [25] Liu, L., Shao, L., "Learning Discriminative Representations from RGB-D Video Data", IJCAI '13 Proceedings of the Twenty Third international joint conference on Artificial Intelligence, Vol. 1, pp. 1493-1500, 2013.
- [2] Yusoff, Y. A., Basori, A. H., Mohamed, F., "Interactive Hand and Arm Gesture Control for 2D Medical Image and 3D Volumetric Medical Visualization", Procedia-Social and Behavioral Sciences. Vol. 97, pp. 723-729, 2013.
- [3] Dondi, P., Lombardi, L., Porta, M., "Development of Gesture-Based Human-Computer Interaction Applications by Fusion of Depth and Colour Video Streams", IET Computer Vision, Vol. 8, pp. 568-578, 2014.
- [4] Xie, R., Sun, X., Xia, X., Cao, J., "Similarity Matching Based Extensible Hand Gesture Recognition", IEEE sensors journal, Vol. 15, pp. 3475-3483, 2015.
- [5] Rautaray, S. S., Agrawal, A., "Real Time Gesture Recognition System for Interaction in Dynamic Environment", Procedia Technology, Vol. 4, pp. 595-599, 2012.
- [6] Lin, J., Ding Y., "A Temporal Hand Gesture Recognition System Based on Hog and Motion Trajectory", Optik - International Journal for Light and Electron Optics, Vol. 124, pp. 6795-6798, 2013.
- [7] Premaratne, P., Ajaz, S., Premaratne, M., "Hand Gesture Tracking and Recognition System Using Lucas-Kanade Algorithms for Control of Consumer Electronics", Neurocomputing, Vol. 116, pp. 242-249, 2013.
- [8] Nolker, C., Ritter, H., "Visual Recognition of Continuous Hand Postures", IEEE Transactions on Neural Networks, Vol. 13, pp. 983-994, 2002.
- [9] Yi, B., Harris Jr, F. C., Wang, L., Yan, Y., "Real-Time Natural Hand Gestures", Computing in Science & Engineering, Vol. 7, pp. 92-96, 2005.
- [10] Varkonyi-Koczy, A. R., Tusor, B., "Human-Computer Interaction for Smart Environment Applications Using Fuzzy Hand Posture and Gesture Models", IEEE Transactions on Instrumentation and Measurement, Vol. 60, pp. 1505-1514, 2011.
- [11] Dardas, N. H., Georganas, N. D., "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques", IEEE Transactions on Instrumentation and Measurement, Vol. 60, pp. 3592-3607, 2011.
- [12] Shen, X., Hua, G., Williams, L., Wu, Y., "Dynamic Hand Gesture Recognition: An Exemplar-Based Approach from Motion Divergence Fields", Image and Vision Computing, Vol. 30, pp. 227-235, 2012.
- [13] Yamato, J., Ohya, J., Ishii, K., "Recognizing Human Action in Time-Sequential Images Using Hidden Markov Model", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 379-385, 1992.
- [14] Marcel, S., Bernier, O., Viallet, J. E., Collobert, D., "Hand Gesture Recognition Using Input-Output Hidden Markov Models", Fourth IEEE International



عصمت راشدی مدرک دکتری خود را از دانشگاه شهید باهنر کرمان در سال ۱۳۹۲ در رشته برق مخابرات دریافت کرد. وی هم اکنون به عنوان استادیار در دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته کرمان فعالیت می‌کند. زمینه‌های تحقیقاتی مورد علاقه‌ی او بازشناسی الگو، پردازش تصویر و الگوریتم‌های فراابتکاری است.



فرزانه آزادی‌پور قهستانی مدرک کارشناسی خود را از دانشگاه شهید باهنر کرمان و کارشناسی ارشد خود را از دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته کرمان در رشته برق الکترونیک به ترتیب در سال‌های ۱۳۹۲ و ۱۳۹۴ دریافت کرد. زمینه‌های تحقیقاتی مورد علاقه‌ی او پردازش تصویر و الگوریتم‌های فراابتکاری است.