



دانشگاه صنعتی سهند

نشریه علمی-فصلنامه‌ای در زمینه مهندسی برق

دوره ۷ - شماره ۱

بهار و تابستان ۱۳۹۹

صفحات ۴ الی ۳۲

ISSN: 2322-3146

<http://jnsee.sut.ac.ir>

بازشناسی زبان اشاره با ترکیب اطلاعات بدست آمده از سنسور کنترل کننده حرکت و تصویر دست

خدیجه مهدیخانلو^۱ و حسین ابراهیم نژاد^۲

^۱ دانشجوی دکترای مهندسی برق مخابرات، آزمایشگاه تحقیقاتی بینایی کامپیوتر، دانشکده مهندسی برق، دانشگاه صنعتی سهند، تبریز، ایران، kh_mahdikhanlou@sut.ac.ir

^۲ نویسنده مسئول، استاد مخابرات، آزمایشگاه تحقیقاتی بینایی کامپیوتر، دانشکده مهندسی برق، دانشگاه صنعتی سهند، تبریز، ایران، ebrahimnezhad@sut.ac.ir

(تاریخ دریافت مقاله: ۱۳۹۸/۰۶/۰۴ تاریخ پذیرش مقاله: ۱۳۹۸/۱۲/۱۱)

چکیده

واژه‌های کلیدی

بازشناسی حالت دست،

سنسور کنترل کننده حرکت،

هیستوگرام گرادیان‌های جهت‌دار،

جنگل تصادفی.

سیستم‌های بازشناسی زبان اشاره کمک می‌کنند تا افراد ناشنوا به رسانه‌های مختلف دسترسی داشته باشند. در این مقاله، سنسور کنترل کننده حرکت و تصویر دست برای بازشناسی زبان اشاره بکار گرفته شده‌اند. سنسور کنترل کننده حرکت اطلاعات مربوط به مکان سه بُعدی مفاصل دست را فراهم می‌کند. سری اول ویژگی‌ها از اطلاعاتی که این سنسور فراهم می‌کند، استخراج می‌شود. مواقعی که دست عمود بر سنسور کنترل کننده حرکت قرار نگرفته یا در حالت‌هایی که دست مشت می‌شود محل دقیق مفاصل دست قابل شناسایی نیست. سری دوم ویژگی‌ها که از تصویر دست استخراج می‌شود کمک می‌کنند تا اکثر حالت‌های دست دقیق‌تر بازشناسی شوند. سری دوم ویژگی‌ها شامل هیستوگرام گرادیان‌های جهت‌دار و فاصله‌ی کانتور از مرکز تصویر است. همچنین یک پایگاه داده‌ی متنوع از حالت‌های دست زبان اشاره‌ی آمریکایی ساخته شده است که شامل ۶۴ هزار نمونه است. در مرحله‌ی بازشناسی از طبقه‌بند جنگل تصادفی استفاده می‌شود که برای پایگاه داده‌های بزرگ انتخاب خوبی است. نتایج آزمایش‌ها نشان می‌دهند که روش پیشنهادی نسبت به روش‌های مشابه نتایج بهتری ارائه می‌دهد.



Sahand University
of Technology

Journal of Nonlinear
Systems in Electrical
Engineering

Vol.7, No.1

Spring and Summer 2020

ISSN: 2322 – 3146

<http://jnsee.sut.ac.ir>

Sign Language Recognition by Combining Leap Motion Controller and Hand Image Information

Khadijeh Mahdikhanelou¹ and Hossein Ebrahimnezhad²

¹Ph.D. student, Computer Vision Res. Lab, Electrical Engineering Department, Sahand University of Technology, kh_mahdikhanelou@sut.ac.ir

²**Corresponding Author**, Professor, Computer Vision Res. Lab, Electrical Engineering Faculty, Sahand University of Technology, ebrahimnezhad@sut.ac.ir

ABSTRACT

Keywords

hand gesture recognition,
Leap Motion Controller,
histogram of oriented
gradients,
random forest.

Sign language recognition systems help deaf people to access various media. In this paper, the Leap Motion Controller (LMC) and the image of the hand are exploited for sign language recognition. The LMC provides 3D position of the hand joints. The first set of features are extracted from the data provided by the LMC. When the hand is not located in vertical view of the LMC or when the hand is posed like a fist, the precise position of the hand joints are not recognizable. The second feature extracted from hand image helps most hand gestures be recognized precisely. The second feature includes histogram of oriented gradients and the distance of the hand contour from the center of the hand. Also, a dataset composed of variant American Sign Language gestures is created which includes 64000 samples. In the recognition stage, the random forest is applied which is a good option for large datasets. The experimental results show that the proposed method performs better than similar methods.

۱- مقدمه

در زبان اشاره به جای استفاده از صوت و نوشتار برای انتقال پیام، از پیام‌های دیداری مانند حرکات بدن، حالت‌های صورت و حرکات دست استفاده می‌شود. زبان اشاره اصلی‌ترین راه ارتباطی بین افراد ناشنواست که در دو نوع ایستا و پویا وجود دارد و می‌تواند برای بیان حروف، اعداد، کلمات و جملات استفاده شود. زبان اشاره را فقط افراد ناشنوا و حداکثر خانواده و دوستان آنها یاد می‌گیرند. از آنجایی که اکثر افراد جامعه با این زبان آشنا نیستند، جامعه ناشنوایان در انزوا قرار می‌گیرند. محققان برای حل این مشکل، سیستم‌های بازشناسی زبان اشاره را طراحی کرده‌اند تا علائم زبان اشاره به صورت خودکار، از تصویر و ویدئو به متن و گفتار تبدیل شوند. بنابراین سیستم‌های بازشناسی زبان اشاره امکان دسترسی افراد ناشنوا به رسانه‌های بیشتری را فراهم کرده‌اند.

می‌توان سیستم‌های بازشناسی زبان اشاره را براساس نوع ورودی به سه گروه دسته‌بندی کرد. تعدادی از سیستم‌ها از تصاویر رنگی یا خاکستری استفاده می‌کنند. این سیستم‌ها در صورت وجود پس‌زمینه پیچیده در تصویر ممکن است دچار مشکل شوند. لبه‌ها در تصاویر رنگی و خاکستری بسیار شفاف هستند. بیشتر سیستم‌های بازشناسی مطرح شده در سال‌های اخیر از تصاویر عمق استفاده می‌کنند. تصاویر عمق اطلاعات مربوط به فاصله دست از دوربین را فراهم می‌کنند و به راحتی می‌توانند دست را از ناحیه پس‌زمینه جدا کنند. ویژگی‌هایی که برای تصاویر رنگی تعریف می‌شوند در تصاویر عمق به دلیل بافت ضعیف، نویز بالا و رزولوشن پایین کارکرد ضعیفی دارند. علاوه بر این، چرخش‌های خارج از صفحه^۱ باعث می‌شوند دو نمونه از حالت دست از یک کلاس، بسیار متفاوت بنظر برسند. تعداد بسیار کمی از سیستم‌های اخیر از سنسور کنترل کننده حرکت^۲ (LMC) برای بازشناسی زبان اشاره استفاده کرده‌اند. سیستم‌های پیشنهاد شده حداکثر نیمی از حروف زبان اشاره آمریکایی را طبقه‌بندی کرده‌اند و تعدادی از سیستم‌ها زبان‌های اشاره غیر آمریکایی را مورد بررسی قرار داده‌اند.

با ظهور سنسور LMC می‌توان موقعیت سه بُعدی مفاصل دست را بدست آورد. دو دوربین تک رنگ و سه LED مادون قرمز داخل LMC تعبیه شده است و با کابل USB به کامپیوتر وصل می‌شود. LED ها حرکت را تشخیص می‌دهند و دو دوربین عمق را محاسبه می‌کنند. LMC در هر ثانیه ۲۰۰ فریم تولید می‌کند و حداکثر تا یک متر را می‌بیند. زمانی که دست حالت مشت به خود می‌گیرد، LMC نمی‌تواند موقعیت دقیق انگشت شست را شناسایی کند. همچنین در حالت‌هایی که دست در راستای عمودی سنسور قرار نمی‌گیرد دقت محل مفاصل کاهش می‌یابد.

دست انسان با ۱۶ قسمت (۱۵ مفصل انگشتان و کف دست) یکی از پیچیده‌ترین اعضای بدن اوست. هریک از مفاصل انگشت و مچ دست می‌توانند زوایای مختلفی داشته باشند. همچنین هر یک از قسمت‌های دست با حرکات خود می‌توانند باعث انسداد قسمت‌های دیگر شوند. دست انسان با ۲۷ درجه آزادی می‌تواند در یک ثانیه ۵ متر انتقال و ۳۰۰ درجه چرخش داشته باشد [۱، ۲]. این عضو از بدن آناتومی بسیار پیچیده‌ای دارد و دو تصویر با دو زاویه مختلف از یک حالت دست می‌تواند بسیار متفاوت با یکدیگر باشند. بنابراین تخمین حالت دست نیاز به پایگاه داده بسیار بزرگ و متنوع دارد. تصویر رنگی اطلاعات مربوط به رنگ

^۱ Out-of-plane

^۲ Leap Motion Controller

و بافت دست را فراهم می کند. اما این ورودی دو بعدی است و برای تحلیل سه بعدی حالت دست کافی نیست. بنظر می رسد ترکیب ورودی های تصویر رنگی و سنسور LMC اطلاعات کامل تری برای تخمین حالت دست فراهم کند.

در این مقاله با ترکیب اطلاعات سنسور LMC و تصویر دست می توان به نتایج بهتری دست یافت. با استفاده از اطلاعات سنسور LMC زاویه بین مفاصل دست، جهت کف دست و جهت مچ دست مشخص می شود. از تصویر دست هم هیستوگرام گرایان های جهت دار [۳] و فاصله ی کانتور از مرکز دست محاسبه می شود.

در این مقاله برای اولین بار تمام حروف و اعداد تک رقمی زبان اشاره آمریکایی با استفاده از ترکیب تصویر و اطلاعات حاصل از سنسور LMC برای بازشناسی زبان اشاره مورد مطالعه قرار می گیرد. یک پایگاه داده ی بزرگ و چالشی شامل ۶۴ هزار نمونه با استفاده از سنسور LMC و یک دوربین ساخته شده است که شامل ۲۴ حرف انگلیسی و ۱۰ عدد است. در تمام مقالات دیگر که از سنسور LMC استفاده کرده اند، پایگاه داده فقط از تعداد محدودی حروف انگلیسی تشکیل شده است. حالت دست در زبان اشاره آمریکایی برای ۷ حرف و یک عدد به صورت مشت است (شکل ۱). روش پیشنهادی نسبت به روش هایی که از تصاویر عمق استفاده می کنند در بازشناسی ژست های دست در حالت مشت بهتر عمل می کند. شکل ۲ مراحل کلی روش پیشنهادی را نشان می دهد.



شکل ۱. کلاس های زبان اشاره آمریکایی که حالت مشت دارند.

در این مقاله، برای دسته بندی داده ها از جنگل تصادفی استفاده می شود. جنگل تصادفی یک الگوریتم قوی یادگیری ماشین است که در کاربردهای بسیاری اعمال می شود. یک جنگل تصادفی از تعداد بسیاری درخت تصمیم گیری تشکیل شده است. در سال ۱۹۹۰، شفیره [۴] ایده ی یادگیری گروهی^۱ را برای موارد مشابه معرفی کرد تا پایداری و صحت الگوریتم های یادگیری ماشین را بهبود بخشد. در الگوریتم یادگیری گروهی، چند طبقه بند^۲ وجود دارد که هر یک روی یک زیرمجموعه از پایگاه داده آموزش داده می شود و نتیجه ی نهایی با استفاده از اکثریت آرای طبقه بندها حاصل می شود. ترکیب نتایج طبقه بندهای مختلف کمک شایانی به بهبود صحت خروجی الگوریتم یادگیری کرده است. جنگل تصادفی هم یک الگوریتم یادگیری گروهی است.

با تغییر پارامترهای درخت تصمیم گیری می توان آن را به انواع داده های پیوسته، گسسته، برچسب خورده و بدون برچسب اعمال کرد. جنگل تصادفی نسبت به نویز، داده های پرت و مقادیر از دست رفته مقاوم است. مکانیزم جنگل تصادفی طوری طراحی شده است که بتواند مفیدترین ویژگی ها را انتخاب کند. در یک جنگل تصادفی، درخت های تصمیم گیری مستقل از یکدیگر هستند. بنابراین می توانند به صورت موازی پردازش شوند که این امر سرعت پردازش را بسیار بالا می برد. امروزه معماری کامپیوتر به صورت موازی است تا هر جز بتواند مستقل از دیگر اجزاء، پردازش را انجام دهد. علاوه بر این، ماهیت تصادفی بودن پارامترهای جنگل تصادفی فاکتور تعمیم را افزایش می دهد [۵-۸].

¹ Ensemble learning

² Classifier

مطالعه‌ای که سال ۲۰۱۳ روی پایگاه داده‌های مختلف صورت گرفت [۹] نشان داد که جنگل تصادفی در مقایسه با شبکه‌های عصبی پس انتشار^۱ و ماشین‌های بردار پشتیبان^۲ عملکرد بهتری داشته است. در شبکه‌های عصبی پس انتشار طراح باید با تجربه قبلی پارامترها را با دقت زیادی تنظیم کند تا تعادل معقولی بین دقت و قدرت تعمیم^۳ طبقه‌بند ایجاد کند. بنابراین کار با این شبکه‌ها برای افراد غیر متخصص و مبتدی بسیار سخت خواهد بود. ماشین‌های بردار پشتیبان نیز مشکلات مشابهی ایجاد می‌کنند. این دسته از طبقه‌بندها پیچیده هستند و سرعت یادگیری آنها پایین است. این طبقه‌بند اصولاً برای طبقه‌بندی بین دو کلاس طراحی شده است و وقتی تعداد کلاس‌ها افزایش می‌یابد پیچیدگی آن بسیار افزایش می‌یابد. تعداد پارامترهایی که در جنگل تصادفی باید تنظیم شود بسیار کم است. نسبت به بیش برآزش^۴ مقاوم است. حتی زمانی که بعضی مولفه‌های بردار ویژگی در دسترس نباشد می‌تواند آن بردار ویژگی را طبقه‌بندی کند. داده‌ی نامتوازن برای این طبقه‌بند مشکلی ایجاد نمی‌کند. در [۱۰] کار مقایسه‌ای بین جنگل تصادفی، ماشین‌های بردار پشتیبان، الگوریتم کی-نزدیک‌ترین همسایگی^۵ و آنالیز افتراقی خطی^۶ انجام گرفت که در آن جنگل تصادفی بالاترین درصد بازشناسی را بدست آورد. لی و همکاران [۱۱] در مطالعه‌ای که برای مقایسه دو الگوریتم جنگل تصادفی و ماشین‌های بردار پشتیبان داشتند مشاهده کردند که صرف نظر از اینکه تحلیل مولفه‌های اصلی^۷ را برای کاهش ابعاد داده بکار ببرند یا نه، جنگل تصادفی نسبت به تغییر پارامترهای خود بسیار پایدار است. در حالی که ماشین‌های بردار پشتیبان حساسیت بسیار زیادی نسبت به پارامترهای خود دارد. آنها برای بهبود نتایج خود از روش بهینه‌سازی ازدحام ذرات^۸ استفاده کردند. هرچند بدون این روش بهینه‌سازی نیز جنگل تصادفی با پارامترهای پیش‌فرض خود به نتایج خوب و با قابلیت تعمیم دست یافت. اما پارامترهای ماشین‌های بردار پشتیبان پس از بهینه‌سازی بهبود چشم‌گیری داشت.

ساختار ادامه‌ی مقاله به این صورت است: بخش ۲ معرفی مختصری راجع به کارهای مربوطه و مطالعات گذشته دارد. ویژگی‌های استخراج شده، فرایند ساخت پایگاه داده و جنگل تصادفی در بخش ۳ توصیف شده‌اند. نتایج آزمایش در بخش ۴ گزارش شده‌اند. تفسیر نتایج در بخش ۵ تفسیری از جدول‌ها و نمودارها ارائه می‌دهد. بخش ۶ نتیجه‌گیری از مقاله را بیان می‌کند.

^۱ Back propagation neural network

^۲ Support vector machines

^۳ Generalization

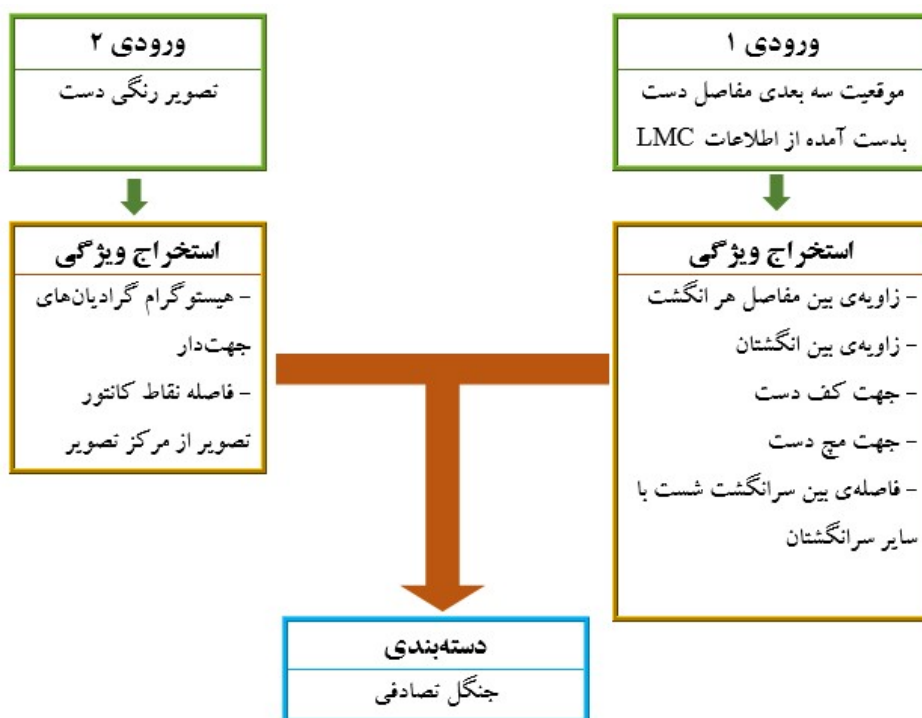
^۴ Overfitting

^۵ k-nearest neighbors

^۶ Linear discriminant analysis

^۷ Principal component analysis

^۸ Particle swarm optimization



شکل ۲. مراحل کلی روش پیشنهادی

۲- کارهای گذشته

در سال‌های گذشته الگوریتم‌های زیادی برای بازشناسی حالت دست و زبان اشاره به صورت ایستا و پویا ارائه شده است. این بخش مرور کوتاهی بر این الگوریتم‌ها دارد. روش‌های بازشناسی دست را براساس نوع ورودی که سیستم می‌گیرد، دسته‌بندی می‌کنیم. سیستم‌های معرفی شده برای بازشناسی حالت دست ورودی‌ها را به صورت تصویر عمق یا تصویر رنگی یا اطلاعات LMC یا ترکیبی از این‌ها دریافت می‌کنند.

در روش خطوط تصادفی [۱۲]، تعدادی پاره‌خط به صورت تصادفی روی تصویر عمق قرار می‌گیرند. سپس، از این پاره‌خط‌ها تعدادی ویژگی ساده استخراج می‌شود مانند: درصد پیکسل‌های پاره‌خط که روی پیش‌زمینه تصویر قرار گرفته‌اند، میانگین پیکسل‌های پاره‌خط که روی پیش‌زمینه تصویر قرار گرفته‌اند، مکان پیکسل با کم‌ترین مقدار که روی پیش‌زمینه تصویر قرار گرفته است و درجه نوسانات در طول پاره‌خط. روشی براساس سوپر پیکسل‌ها در مرجع [۱۳] معرفی شده است. یک سوپر پیکسل خوشه‌ای از پیکسل‌هاست که فاصله‌ی کمی باهم دارند و رنگ و مقدار عمق آنها مشابه است. در کارهای [۱۴، ۱۵] فیلترهای گابور برای استخراج ویژگی‌های بافتی در تصاویر عمق بکار گرفته شده‌اند. ژانگ و همکاران [۱۶] پیکسل‌های تصویر عمق را به صورت توده‌ای از نقاط سه بُعدی در نظر گرفته‌اند و اطلاعات تصویر عمق را به صورت هیستوگرامی از این توده‌ها بیان کرده‌اند. در کار مرجع [۱۷]، یک جنگل تصادفی چند لایه‌ای برای طبقه‌بندی حالت دست از تصاویر عمق ارائه شده است. فنگ و همکاران [۱۸] مختصات پیکسل‌های تصویر عمق را در سه جهت نگاشت^۱ نموده و از آنها یک مجموعه ویژگی استخراج می‌کنند.

^۱ Project

در روش هایی که از تصاویر عمق استفاده می کنند، ویژگی هایی که برای تصاویر عمق تعریف شده اند روی ژست های مشت بسیار ضعیف عمل می کنند. روش خطوط تصادفی [۱۲] روی حروف $X, T, S, Q, P, N, M, K, G, E$ بسیار ضعیف عمل می کند. بیشتر نمونه های کلاس P به عنوان نمونه های کلاس H و Q و حدود نیمی از نمونه های کلاس Q به عنوان کلاس A دسته بندی شده اند. همچنین نمونه های کلاس M و N به جای یکدیگر طبقه بندی شده اند. در روش مرجع [۱۷] که از تصاویر عمق ویژگی استخراج می کند، بیشترین خطای بازشناسی مربوط به کلاس های M, N و T است. در روش سوپر پیکسل [۱۳] که از ویژگی های بافت، رنگ و عمق برای بازشناسی استفاده شده است، عملکرد ضعیفی روی حروف E, K, M, N, P, Q و T دارد. فیلترهای گابور تعریف شده روی تصاویر عمق [۱۴] نیز برای بازشناسی و تمایز بین حروف M, N و T کارآمد نیست. روش هیستوگرام وجه های سه بُعدی [۱۶] نیز در شناسایی حروف M, N, P و Q ضعیف عمل می کند. در مرجع [۱۹] که از یادگیری عمیق استفاده شده، درصد بازشناسی حروف حالت های مشت نسبت به سایر حالت های دست پایین است.

مارین و همکاران [۲۰] از دو سنسور LMC و کینکت برای بازشناسی حالت دست استفاده کرده اند. آنها نسخه ی اولیه ی LMC را بکار بردند که فقط مرکز دست و سرانگشتان و جهت مچ دست را مشخص می کرد. آنها الگوریتم خود را روی پایگاه داده ای شامل ۱۰ حالت دست و ۱۴۰۰ نمونه آزمایش کردند و تنها یکی از حالت های دست ژست مشت را دارد. در مرجع [۲۱]، از LMC برای بازشناسی زبان اشاره عربی استفاده شده است. کومار و همکاران [۲۲] از LMC برای گرفتن حرکات دست و از دوربین کینکت برای گرفتن حالات صورت استفاده کرده اند تا ۵۱ کلمه را با زبان اشاره پویا بازشناسی کنند. در مرجع [۲۳]، اسکلت دست برای ۱۱ حرف انگلیسی و ۷ عدد جمع آوری شده است که تنها شامل یک حالت دست به صورت مشت است. پایگاه داده ای آنها به صورت فریم های متوالی است. میتال و همکاران [۲۴] LMC برای شناسایی جملات هندی استفاده کرده اند که در آن هر جمله از چند ژست متوالی دست تشکیل شده است. لی و همکاران [۲۵] از LMC برای شناسایی ۱۰ ژست ایستا (اعداد یک تا ده) و چهار ژست پویا استفاده کرده اند.

رن و همکاران [۲۶] یک روش کانتوری برای بازشناسی حالت دست از تصاویر رنگی ارائه داده اند. آنها ابتدا جهت دست را تعیین کرده و در ادامه فاصله ی نقاط کانتور دست را از مرکز دست به عنوان ویژگی در نظر گرفته اند. برای تعیین دست، آنها مستطیل محیطی شکل دست را در مقیاس های مختلف به مستطیل های کوچک تر تقسیم می کنند و در هر مستطیل جهت قطعات کانتور را محاسبه می کنند و سپس، هیستوگرام جهت ها در مستطیل ها را محاسبه می کنند. همچنین آنها به هر مستطیل وزنی اختصاص می دهند. برای مثال برای مستطیل هایی که گوشه های کانتور را شامل می شوند وزن های کمتری اختصاص می دهند و بنابراین جهت دست در حالتی که دست مشت شده است دقیق مشخص نمی شود. آنها روش خود را روی پایگاه داده ای مشکل از ۱۴ حالت دست که شامل هیچ ژست مشت نمی شود، آزمایش کرده اند. در مرجع [۲۷]، برای بازشناسی حالت دست یک روش کانتوری ارائه شده است. آنها شیب را در تمام نقاط کانتور محاسبه کرده اند و نقاطی که در آن علامت شیب عوض می شود به عنوان نقاط شاخص کانتور در نظر گرفته اند. این روش بسیار نوپز پذیر است زیرا شیب را به صورت محلی در هر نقطه محاسبه می کند و ممکن است تعداد نقاط شاخص برای تمام نمونه ها یکسان نباشد. نویسندگان روشی برای یکسان سازی تعداد نقاط شاخص ارائه کرده اند. آنها در مکان هایی که چگالی نقاط شاخص بالاست تعداد نقاط شاخص را کاهش داده اند. هرچند این کار مشکل را به طور کامل حل نکرده است.

۳- روش پیشنهادی

در این قسمت جزئیات سیستم پیشنهادی برای بازشناسی سه بعدی زبان اشاره آمریکایی ارائه می شود. در بخش ۱-۱-۳ و ۲-۳ بترتیب استخراج ویژگی از داده های LMC و روش محاسبه هیستوگرام گرادیان های جهت دار فاصله کانتر دست از مرکز دست توضیح داده شده است. در ادامه نیز فرایند ساخت پایگاه داده و طبقه بندی با استفاده از جنگل تصادفی ارائه شده است.

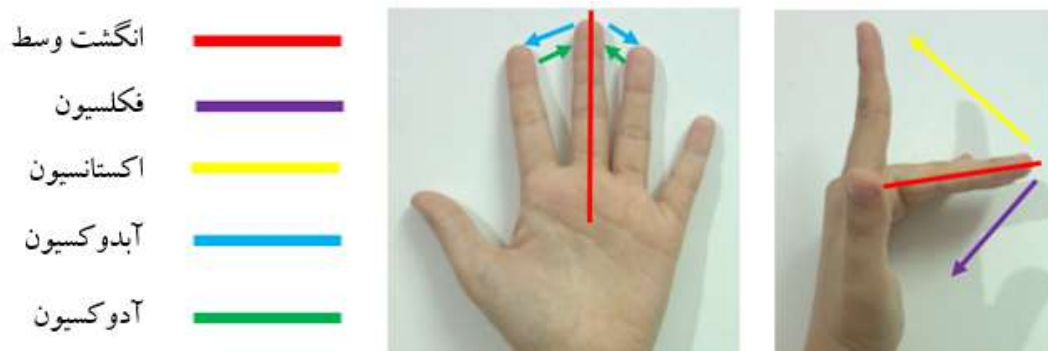
۳-۱- استخراج ویژگی

دو سری ویژگی از داده های LMC و تصویر بدست آمده از دوربین استخراج می شود. در نتایج آزمایش ها خواهیم دید که این دو مجموعه ویژگی مکمل یکدیگر هستند و می توان با ترکیب آنها به دقت بالایی در بازشناسی زبان اشاره رسید. ویژگی های تعریف شده ساده هستند و به راحتی محاسبه می شوند.

۳-۱-۱- استخراج ویژگی از داده های سنسور LMC

دست انسان می تواند حالت های بسیاری را به خود بگیرد. با این وجود، حرکات دست را می توان در چند دسته قرار داد و تمام حالت های دست را براساس همین حرکات دسته بندی کرد. در ادامه حرکاتی که دست انسان می تواند انجام دهد بیان شده است و شکل ۳ این حرکات را نشان می دهد.

- فلکسیون^۱: حرکات انگشتان به طرف داخل دست؛
- اکستانسیون^۲: حرکات انگشتان به طرف خارج دست؛
- آبدوکسیون^۳: دور کردن انگشتان از انگشت وسط؛
- آدوکسیون^۴: نزدیک کردن انگشتان به انگشت وسط.



شکل ۳. حرکات دست

¹ Flexion
² Extension
³ Abduction
⁴ Adduction

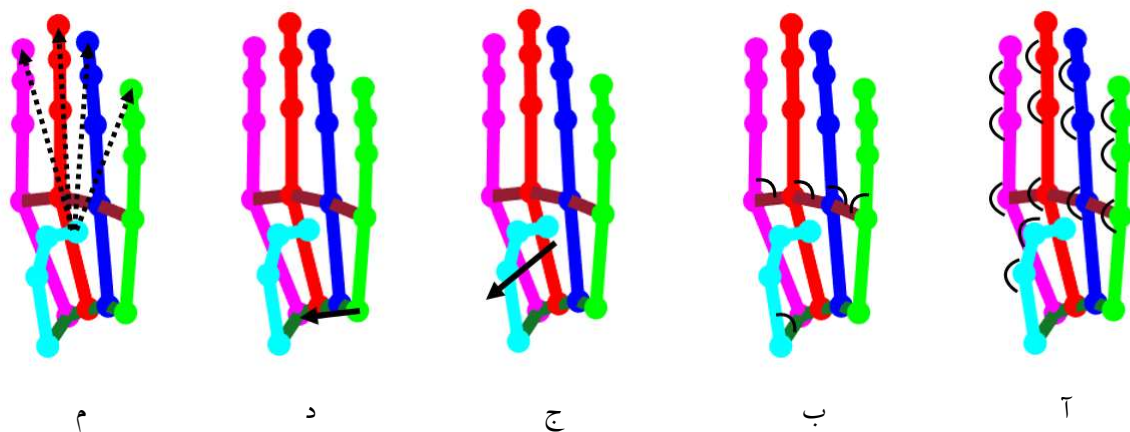
با این حرکات زاویه‌های مفاصل دست تغییر می‌کنند. بنابراین با محاسبه زاویه‌ی بین مفاصل می‌توان حالت‌های دست را از یکدیگر تفکیک کرد. LMC موقعیت سه بُعدی مفاصل و جهت کف دست را محاسبه می‌کند. با استفاده از این اطلاعات می‌توان جهت میچ دست و زاویه‌ی مفاصل را بدست آورد. مجموعه اول ویژگی‌های استخراج شده شامل موارد زیر است:

- زاویه بین مفاصل هر انگشت که میزان حرکت فلکسیون و اکستنسیون را مشخص می‌کند (شکل ۴-آ)؛
- زاویه بین انگشتان که میزان حرکت آبدوکسیون و آدوکیسیون را مشخص می‌کند (شکل ۴-ب)؛
- جهت کف دست (شکل ۴-ج)؛
- جهت میچ دست (شکل ۴-د)؛
- فاصله‌ی بین سرانگشت شست با سایر سرانگشتان (شکل ۴-م).

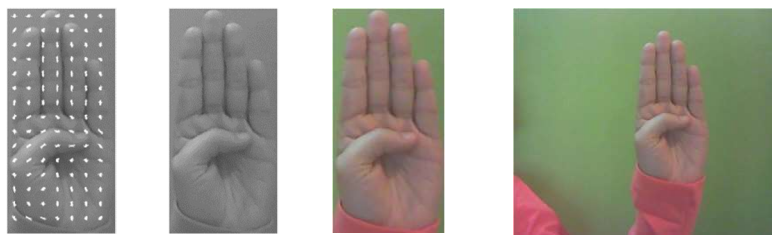
تنها تفاوت حالت دست برای حرف K و حرف P جهت کف دست است. حروف U, V و R فقط در زاویه بین انگشتان متفاوت هستند. تفاوت بین حرف O و عدد 0 در جهت میچ دست و جهت کف دست است. حرکت انگشت شست با بقیه انگشتان متفاوت است. انگشت شست درجه آزادی زیادی دارد و ساختار آناتومی آن با بقیه انگشتان متفاوت است. بنابراین، تعیین موقعیت دقیق انگشت شست مشکل است. فاصله بین انگشت شست با سایر انگشتان را به عنوان ویژگی دیگر برای حالت دست در نظر می‌گیریم تا موقعیت تقریبی انگشت شست مشخص شود.

۳-۱-۲- استخراج ویژگی از تصویر دست

هیستوگرام گرادیان‌های جهت‌دار را می‌توان برای تک تک کانال‌های رنگی (قرمز- آبی- سبز) یا برای تصویر خاکستری تعریف کرد. در این مقاله ابتدا تصویر دست از کل تصویر جدا می‌شود. پس از تبدیل تصویر رنگی دست به تصویر خاکستری ویژگی هیستوگرام گرادیان‌های جهت‌دار استخراج می‌شود (شکل ۵).



شکل ۴. ویژگی‌های استخراج شده از داده‌های سنسور LMC



شکل ۵. (آ) کل تصویر رنگی، (ب) تصویر رنگی ناحیه ی دست، (ج) تصویر خاکستری دست، (د) هیستوگرام گرادیان های جهت دار در نواحی مختلف دست

برای محاسبه هیستوگرام گرادیان های جهت دار [۳] گرادیان هر پیکسل از تصویر در جهت های افقی و عمودی بدست می آید. گرادیان در جهت افقی اختلاف مقدار خاکستری پیکسل های چپ و راست است. به طور مشابه گرادیان در جهت عمودی با محاسبه اختلاف مقدار خاکستری پیکسل های بالا و پایین بدست می آید. برای مثال در شکل ۶ برای پیکسل مرکزی گرادیان در جهت افقی و عمودی، مقدار گرادیان و جهت آن به صورت زیر محاسبه می شود.

	90	
100	Target Pixel	60
	40	

شکل ۶. مقادیر خاکستری پیکسل های اطراف پیکسل مرکزی

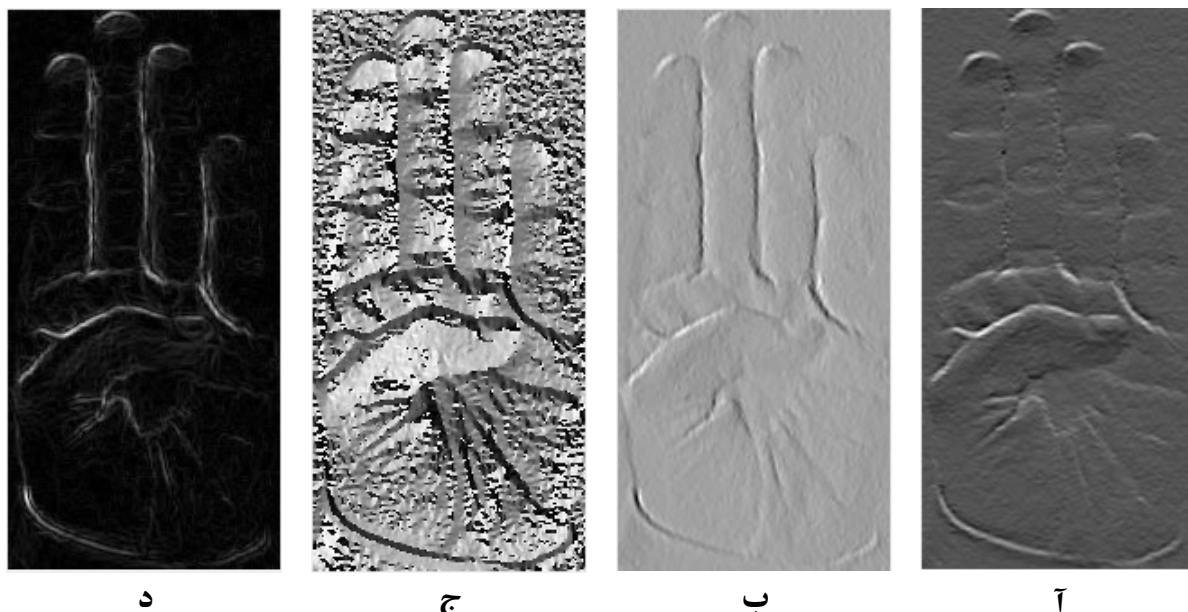
$$I_x = 60 - 100 = -40 \quad (۱)$$

$$I_y = 40 - 90 = -50 \quad (۲)$$

$$A = \sqrt{(-40)^2 + (-50)^2} = 64.03 \quad (۳)$$

$$\theta = \tan^{-1}\left(\frac{-50}{-40}\right) = 51.34 \quad (۴)$$

شکل ۷ (آ و ب) تصاویر گرادیان در جهت های افقی و عمودی را نشان می دهد. با استفاده از این دو تصویر می توان مقدار گرادیان و جهت آن را بدست آورد (شکل ۷ ج و د).



شکل ۷. (آ) تصویر گرادیان در جهت افقی، (ب) تصویر گرادیان در جهت عمودی، (ج) تصویر جهت گرادیان، (د) تصویر مقدار گرادیان.

در تساوی‌های (۵) تا (۶)، I_x و I_y به ترتیب تصویر خاکستری، تصویر گرادیان در جهت افقی و تصویر گرادیان در جهت عمودی است. تصویر جهت گرادیان و تصویر مقدار گرادیان به ترتیب با استفاده از روابط (۷) و (۸) بدست می‌آید.

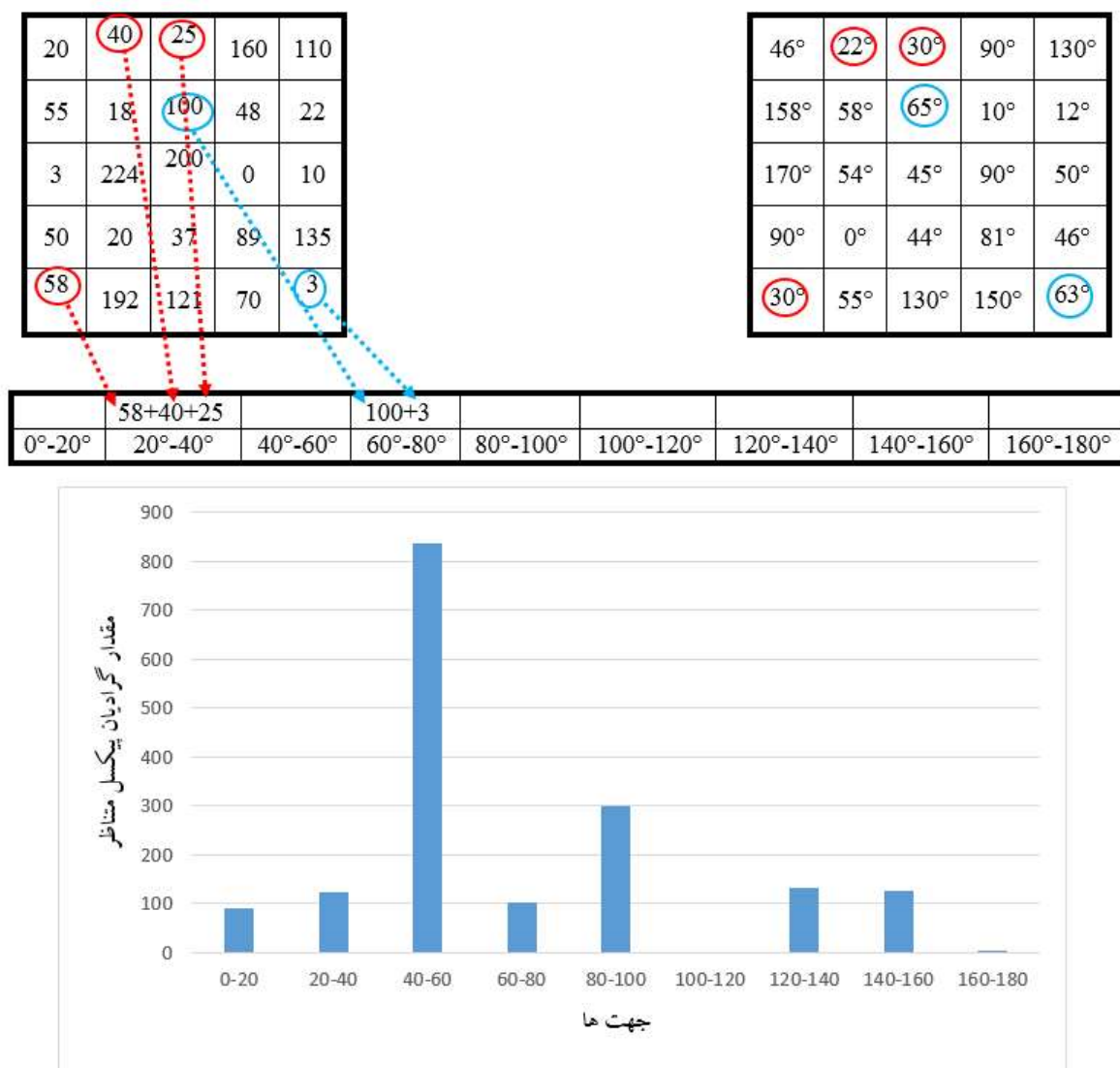
$$\frac{dl}{dx} = [-1 \ 0 \ 1] \times I = I_x \quad (5)$$

$$\frac{dl}{dy} = [-1 \ 0 \ 1]^T \times I = I_y \quad (6)$$

$$\theta = \tan^{-1}\left(\frac{I_y}{I_x}\right) \quad (7)$$

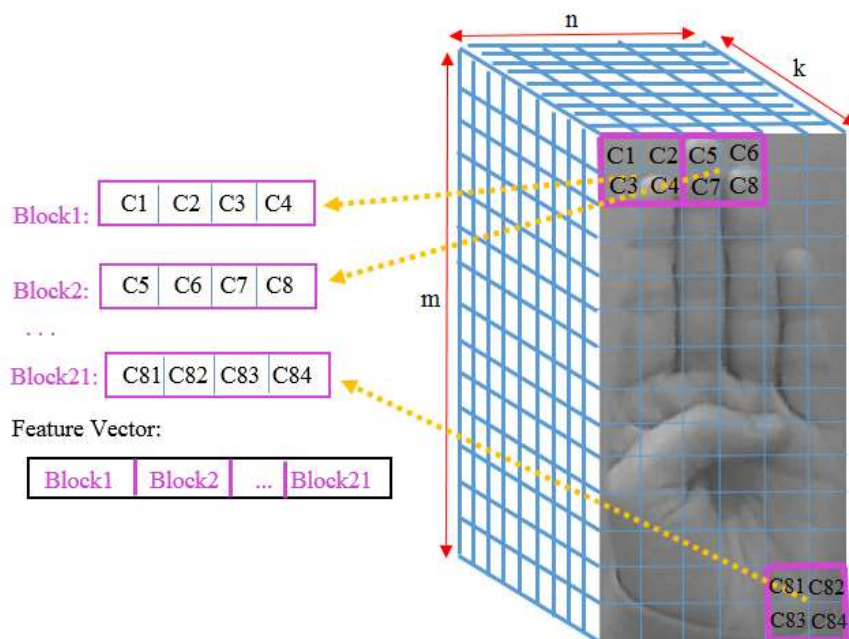
$$A = \sqrt{I_x^2 + I_y^2} \quad (8)$$

شکل ۸ طرز محاسبه هیستوگرام گرادیان‌های جهت‌دار را با یک مثال ساده نشان می‌دهد. ماتریس سمت چپ و راست به ترتیب مقادیر گرادیان‌ها و جهت آنها را برای ۲۵ پیکسل نشان می‌دهند. ماتریس پایین شکل ۸ مقادیر میله‌های هیستوگرام را نشان می‌دهد.



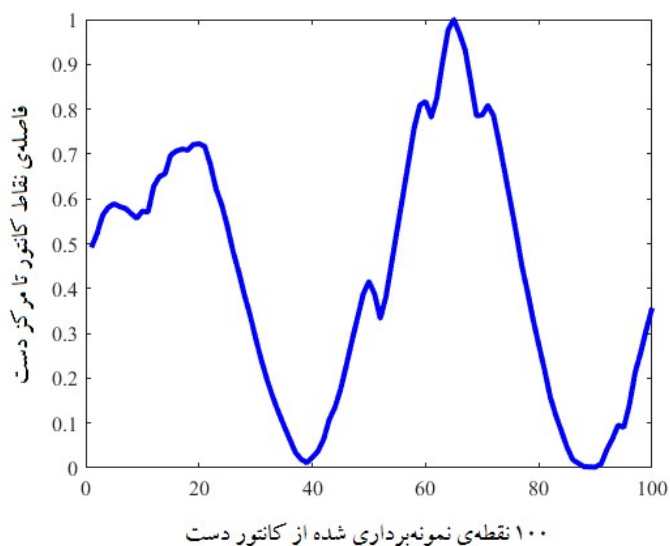
شکل ۸. محاسبه‌ی هیستوگرام گرادین‌های جهت‌دار

تصویر به نواحی کوچکتري به نام سلول تقسیم می‌شود. تصویر دست در شکل ۹ شامل 308×150 پیکسل است. اگر اندازه هر سلول را 22×25 در نظر بگیریم کل تصویر به ۸۴ سلول تقسیم می‌شود ($m \times n = 84$). هر مربع آبی روی تصویر شکل ۹ یک سلول را نشان می‌دهد. در هر سلول هیستوگرام گرادین‌ها در k میله جمع آوری می‌شود. در شکل ۹ و شکل ۸، ۹ میله برای هیستوگرام وجود دارد ($k=9$). برای نرمالیزه کردن بردار ویژگی، یک سری بلوک تعریف می‌شود. هر بلوک از تعدادی سلول تشکیل شده است. بلوک‌ها می‌توانند با یکدیگر همپوشانی داشته باشند که در این صورت اطلاعات مضاعف ایجاد ولی عملکرد بهتری حاصل می‌شود. بردار ویژگی از کنار هم قرار گرفتن بلوک‌ها تشکیل می‌شود (شکل ۹). در شکل ۹، کادرهای بنفش بلوک‌ها را نشان می‌دهد که با یکدیگر همپوشانی ندارند. در این شکل ۲۱ بلوک وجود دارد که هر بلوک شامل ۴ سلول و هر سلول شامل ۹ میله است. بنابراین بُعد بردار ویژگی $21 \times 4 \times 9 = 756$ خواهد بود.



شکل ۹. (آ) پنجره‌های آبی و بنفش روی تصویر به ترتیب سلول‌ها و بلوک‌ها را نشان می‌دهند. k تعداد میله‌های هیستوگرام را نشان می‌دهد. هر بلوک از تعدادی سلول تشکیل شده است. بردار ویژگی از کنار هم قرار گرفتن بلوک‌ها تشکیل می‌شود.

ویژگی بعدی که از تصویر دست استخراج می‌شود فاصله نقاط کانتور دست از مرکز دست است (شکل ۱۰). برای این کار ابتدا با تعیین کانال‌های رنگی (قرمز-سبز-آبی) پوست دست از پیش زمینه تصویر جدا می‌شود. سپس با استفاده از الگوریتم مور^۱ دنباله نقاط کانتور (مرزی) دست استخراج می‌شود [۲۸]. با شروع از اولین نقطه‌ی سمت چپ و پایین و در جهت پادساعتگرد، فاصله اقلیدسی نقاط کانتور دست از مرکز دست حاصل می‌شود. مرکز دست میانگین نقاط دست است.



ب

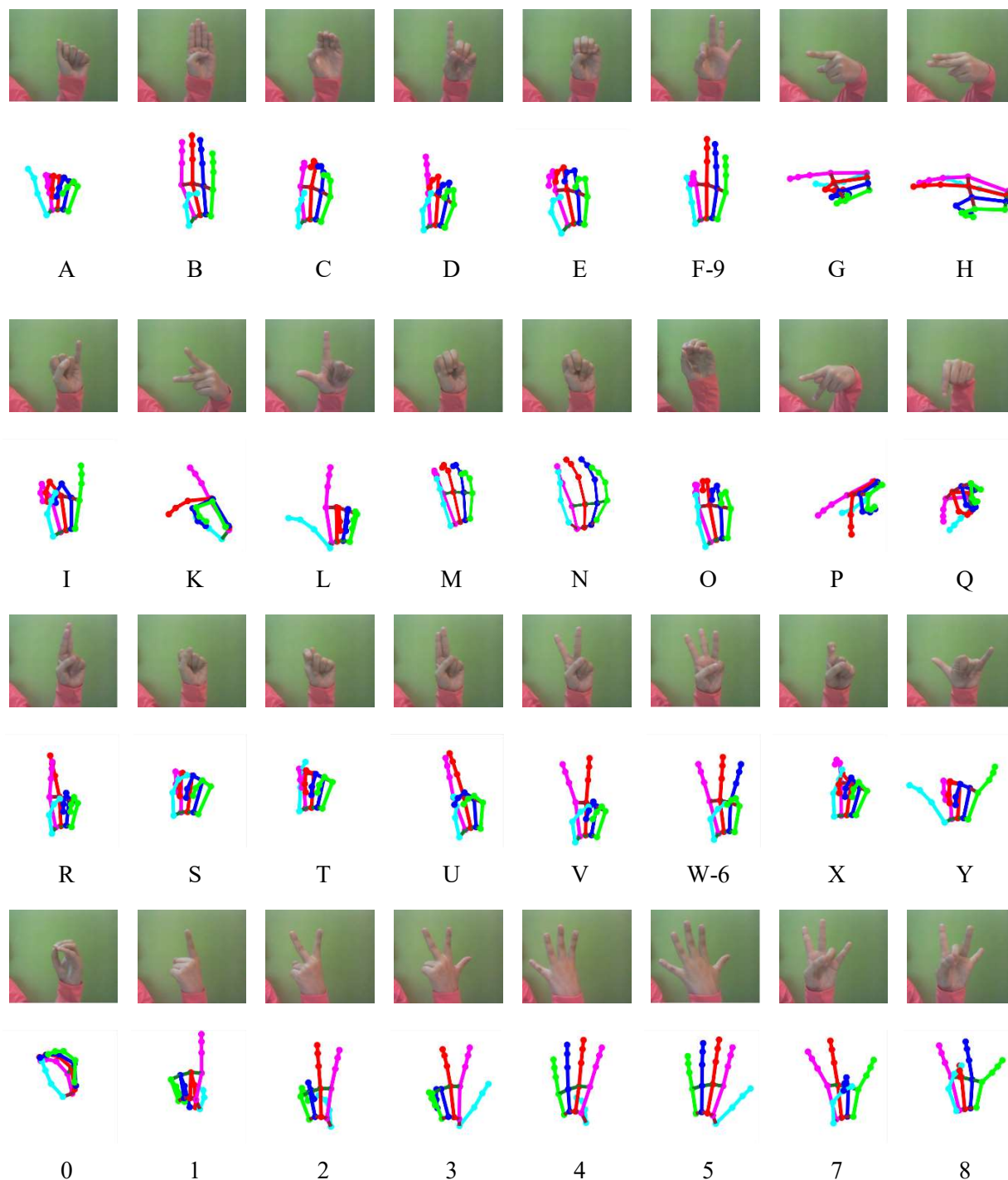
آ

شکل ۱۰. (آ) نقاط کانتور دست و مرکز دست (ب) نمودار فاصله‌ی نقاط کانتور تا مرکز دست

¹ Moore

۳-۲- پایگاه داده

برای انجام این پژوهش یک پایگاه داده متشکل از ۳۲ کلاس ساخته شد. پایگاه داده شامل حروف انگلیسی (به جز حروف J و Z) و اعداد (0 تا 9) است. حالت دست برای 6 و W همچنین برای 9 و F یکسان است. زبان اشاره برای حروف J و Z پویاست و با حرکت دست همراه است. هر کلاس شامل ۲۰۰۰ نمونه است. بنابراین، کل پایگاه داده شامل ۶۴ هزار نمونه از تصویر رنگی و اسکلت دست است. یک فریم از هر ۱۸ فریم متوالی انتخاب شده است تا شباهت فریم‌ها کمتر باشد. شکل ۱۱ نمونه‌هایی از این پایگاه داده را نشان می‌دهد. در جدول ۱ پارامترهای پایگاه داده و مشخصات دوربین و سنسور بیان شده است.



شکل ۱۱. نمونه‌هایی از پایگاه داده

جدول ۱. پارامترهای پایگاه داده، دوربین و سنسور.

۳۲	تعداد کلاس‌ها
۲۰۰۰	تعداد نمونه در هر کلاس
۵۴۰×۴۶۰	رزولوشن تصاویر
Logitech Webcam C170	مدل دوربین تصویربرداری
۵ مگا پیکسل	رزولوشن دوربین
۳۰ فریم بر ثانیه	سرعت فیلم‌برداری دوربین
۷۰ درجه	محدوده زاویه دید لنز دوربین
V ₂	مدل سنسور LMC
Winsows_2.2.7	مدل سنسور SDK LMC

۳-۳- طبقه‌بندی با استفاده از جنگل تصادفی

جنگل تصادفی مجموعه‌ای از چند درخت تصمیم‌گیری است که هر یک از درخت‌ها روی زیرمجموعه‌ای از پایگاه داده آموزش داده می‌شود. این زیرمجموعه از پایگاه داده برای هر درخت به صورت تصادفی انتخاب می‌شود. می‌توان برخی یا تمام پارامترهای جنگل تصادفی را به صورت تصادفی انتخاب کرد. در تئوری گراف، یک درخت گرافی است که شامل هیچ حلقه‌ای نباشد و بین گره‌های آن فقط یک مسیر وجود داشته باشد. در هر نود^۱ (گره) درخت تصمیم‌گیری یک آزمایش (تصمیم) روی یک ویژگی انجام می‌شود. لبه‌های درخت نتایج تصمیم‌گیری را نشان می‌دهند و به گره‌های پایانی هر درخت، گره‌های هدف گفته می‌شود. وقتی درخت تصمیم‌گیری یک بردار ویژگی دریافت می‌کند، یک سری آزمایش متوالی انجام می‌دهد تا به هدف (برچسب کلاس بردار ویژگی) برسد. پاسخ آزمایش فعلی مشخص می‌کند که در نود بعدی چه آزمایشی انجام شود. میزان اعتماد^۳ به درخت تصمیم‌گیری بستگی به تعداد آزمایش‌هایی است که در طول درخت انجام می‌شود.

ساختار و شکل کلی یک درخت به چند مفهوم وابسته است. مفهوم اول نوع آزمایش‌هایی است که در طول درخت باید انجام شود. این آزمایش‌ها توابع انشعاب^۴ یا توابع تست یا یادگیرنده‌های ضعیف^۵ نامیده می‌شوند. شکل ۱۲ سه تابع رایج برای تست در هر نود را نشان می‌دهد. ممکن است بُعد ویژگی بسیار بالا باشد که بکارگیری کل بُعد ویژگی در برخی مواقع عملی نیست و همچنین بعضی از آنها زائد^۶ هستند. مفهوم دوم مربوط به انتخاب زیرمجموعه‌ای از بُعد ویژگی است. نمونه‌های آموزشی در هر نود به چند شاخه تقسیم می‌شوند (در این مقاله فرم درخت به صورت باینری است یعنی در هر گره از درخت دو شاخه خروجی داریم).

^۱ Node

^۲ Edge

^۳ Confidence

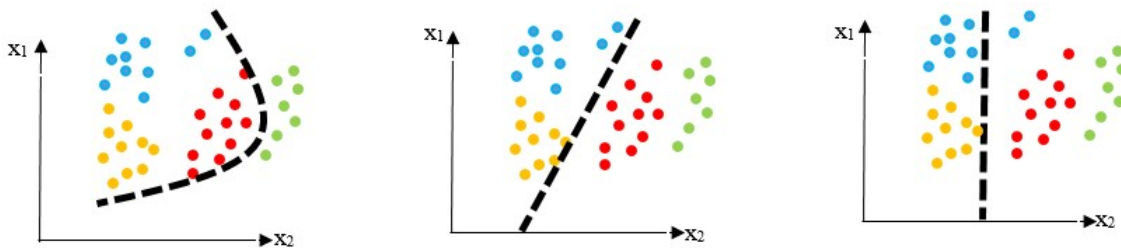
^۴ Split function

^۵ Weak learner

^۶ Redundant

مفهوم سوم تعیین می کند که چگونه هر دوره نمونه های آموزشی به دو قسمت تقسیم شوند. به عبارت دیگر معیار جداسازی داده های آموزشی در هر نود مفهوم سوم را مشخص می کند. این مفهوم الویت آزمایش ها را در طول درخت تعیین می کند. هر نود متناظر با یک تابع انشعاب است. در این مقاله توابع انشعاب باینری هستند. یعنی هر نود به دو بخش تقسیم می شود. در هر نود مجموعه $\theta = (\phi, \psi, \tau)$ پارامترهای نود را مشخص می کند. زیرمجموعه ای از ویژگی ها را انتخاب می کند که در شکل ۱۲ نیز ویژگی های x_1 و x_2 را برگزیده است. ψ نوع تابع تست را از بین توابع سهمی، تابع ثابت و تابع خطی و τ مقدار حد آستانه را مشخص می کند. برای مثال در شکل ۱۲ به ازای $\phi(x_1, x_2) > \tau$ داده های آبی و نارنجی بالای تابع خطی ψ قرار می گیرند.

$$\theta = (\phi, \psi, \tau) \quad (9)$$



شکل ۱۲. سه تابع رایج برای جداسازی داده ها (الف) تابع ثابت، (ب) تابع خطی، (ج) تابع سهمی.

در هر نود باید یک تابع هدف تعریف شود به طوری که به بهترین شکل داده های آموزشی را به دو قسمت تقسیم کند. تابع هدف نوع تابع انشعاب و پارامترهای آن را نیز مشخص می کند. تابع هدف باید طوری داده های آموزشی را تقسیم کند که هر قسمت بیشترین خلوص یا کمترین انتروپی ممکن را داشته باشد. برای محاسبه ای انتروپی در هر نود فرض می شوند که توزیع داده ها به صورت گاوسی است. بنابراین، انتروپی مجموعه داده ای S_j در نود j -ام و با بُعد d به صورت زیر محاسبه می شود:

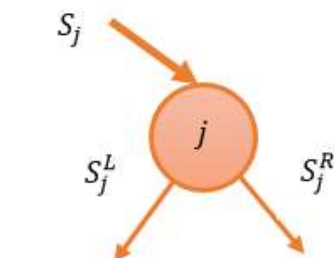
$$H(S_j) = \frac{1}{2} \log((2\pi e)^d |\Lambda(S_j)|) \quad (10)$$

Λ ماتریس کواریانس با بُعد $d \times d$ است و بهره ای اطلاعات به صورت زیر محاسبه می شود:

$$I(S_j, \theta_j) = H(S_j) - \sum_{i \in \{L, R\}} \frac{|S_j^i|}{|S_j|} H(S_j^i) \quad (11)$$

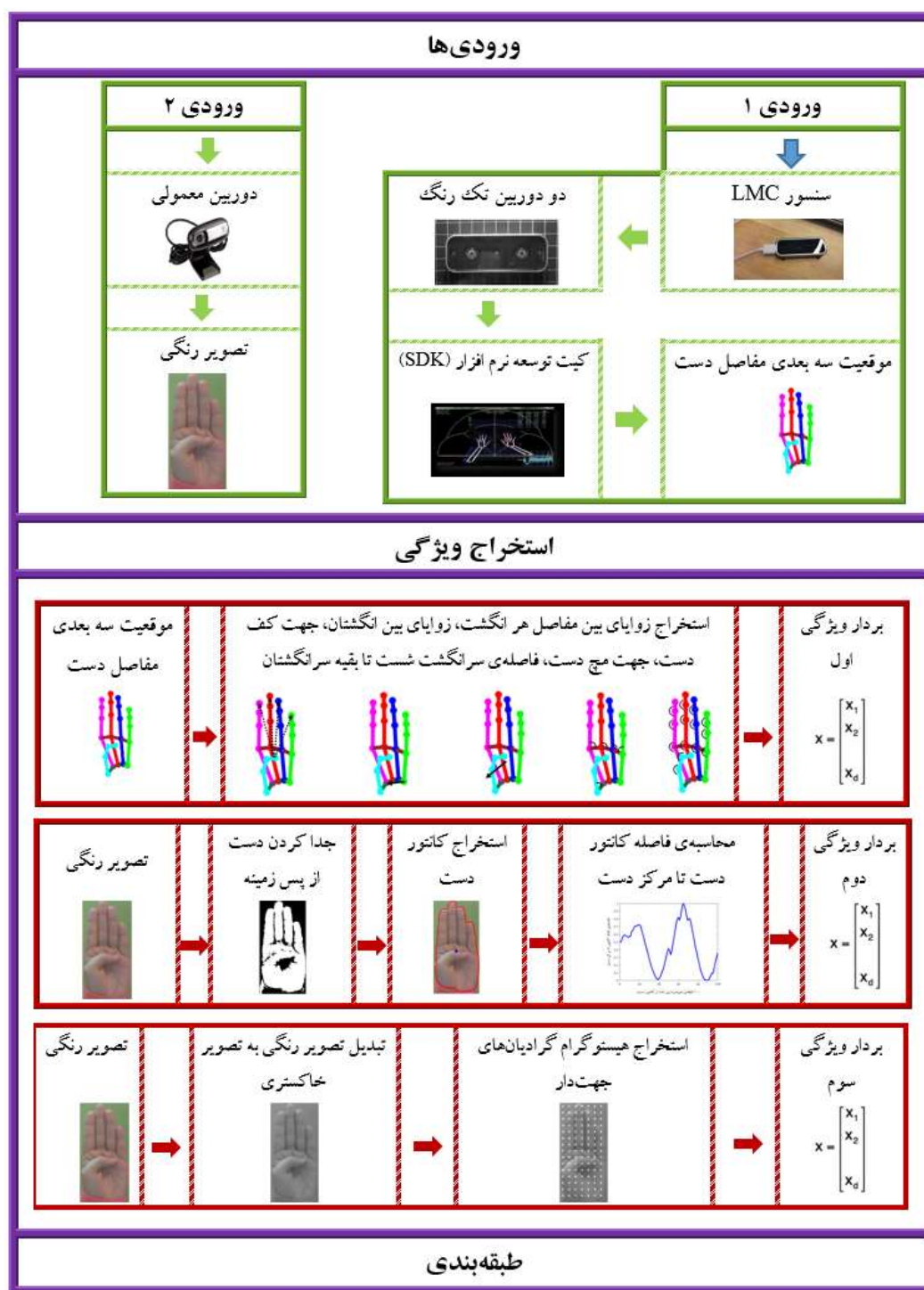
$$\theta_j^* = \operatorname{argmax}_{\theta_j} I(S_j, \theta_j) \quad (12)$$

S_j^i زیرمجموعه ای از داده ها بعد از تقسیم S_j به زیرمجموعه های چپ و راست است (شکل ۱۳). در هر نود از درخت باید انتروپی حداقل یا بهره ای اطلاعات حداکثر شود.



شکل ۱۳. یک نود از درخت تصمیم گیری

جزئیات روش پیشنهادی در شکل ۱۴ نشان داده شده است.



شکل ۱۴. جزئیات مراحل روش پیشنهادی

۴- نتایج آزمایش

کارایی سیستم پیشنهادی در این پژوهش برای سناریوهای مختلفی بررسی شده است که نتایج بدست آمده در این بخش گزارش داده می‌شود. ابتدا تغییرپذیری یا پراکندگی داده‌های سنسور LMC بررسی می‌شود. سپس، نتایج با اعتبارسنجی‌های

متقابل^۱ مختلف بدست می آید. اثر هریک از ویژگی های بدست آمده از سنسور LMC و تصویر دست جداگانه بررسی شده است. همچنین اثر همپوشانی بلوک ها در استخراج ویژگی هیستوگرام گرادیدان های جهت دار روی زمان پردازش و دقت بازشناسی بررسی می شود. روش پیشنهادی برای استخراج ویژگی از تصویر دست با دو روش دیگر که از تصویر دست استفاده می کنند، مقایسه می شود. روش پیشنهادی روی یک پایگاه داده ی دیگر نیز آزمایش می شود.

۴-۱- پراکندگی داده های LMC

در این بخش پراکندگی و تغییرپذیری داده های LMC بررسی می شود. بردار ویژگی بدست آمده از داده های LMC شامل ۲۳ مولفه است. ویژگی های ۱ تا ۱۴ زاویه های مفاصل هر انگشت را نشان می دهد. ویژگی های ۱۵ تا ۱۹ زاویه ی بین انگشتان است (شکل ۱۵). ویژگی های ۲۰ تا ۲۳ فاصله ی سرانگشت شست با سایر سرانگشتان است. شکل ۱۶ ماتریس واریانس ویژگی های LMC را نشان می دهد. برای محاسبه ی پراکندگی داده ی J-ام، واریانس ویژگی J-ام را برای تمام نمونه های یک کلاس بدست آورده شده است. ماتریس ویژگی های محاسبه شده برای سنسور LMC به صورت F_{LMC} در رابطه ی (۱۳) است که سطرهای آن نوع ویژگی (۲۳ ویژگی) و ستون های آن نمونه های پایگاه داده (۶۴۰۰۰ نمونه) را نشان می دهد. برای مثال ویژگی اول زاویه مفصل پایین انگشت کوچک است و ویژگی ۱۱ زاویه ی مفصل وسط انگشت اشاره است (شکل ۱۵). ماتریس واریانس ویژگی ها به صورت V_{LMC} در رابطه (۱۴) نمایش داده شده است که سطرهای آن نوع ویژگی (۲۳ ویژگی) و ستون های آن کلاس های پایگاه داده (۳۲ کلاس) را نشان می دهد. برای محاسبه ی مولفه v_{ik} از ماتریس V_{LMC} کفایت عناصر x_{ij} به ازای $j = k, k + 1, \dots, k + 2000 - 1$ (از هر کلاس ۲۰۰۰ نمونه وجود دارد) از ماتریس F_{LMC} را در نظر بگیریم و واریانس آنها را بدست آورده شده است.

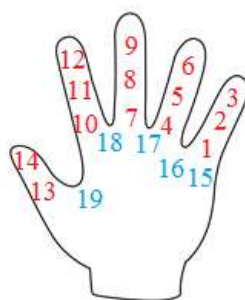
$$F_{LMC} = (x_{ij})_{23 \times 64000} \quad (13)$$

$$V_{LMC} = (v_{ik})_{23 \times 32} \quad (14)$$

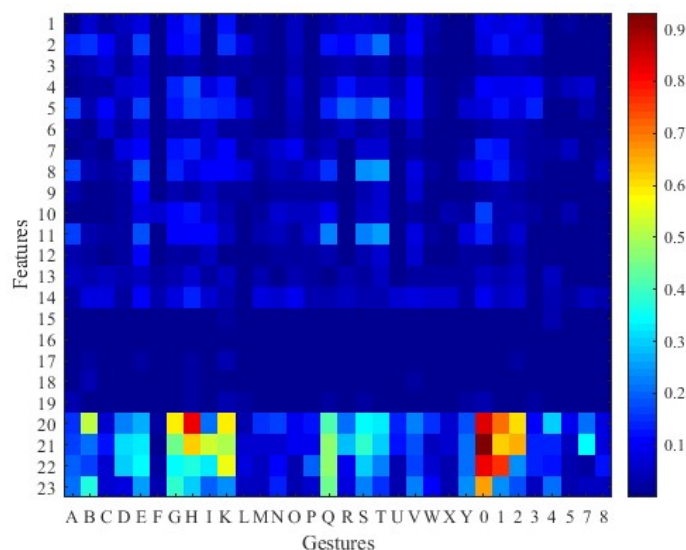
$$v_{ik} = variance(x_{ik}, x_{i(k+1)}, \dots, x_{i(k+2000-1)}) \quad (15)$$

همانطور که در این نمودار دیده می شود زاویه بین انگشتان (۱۵ تا ۱۹) واریانس بسیار کمی دارند. این امر می تواند به این خاطر باشد که زاویه بین انگشتان محدوده ی تغییر کمی دارد. فاصله سرانگشت شست با سرانگشتان دیگر بیشترین واریانس را دارند (ویژگی های ۲۰ تا ۲۳). واریانس بالا برای این ویژگی ها نشان می دهد که سنسور LMC نتوانسته است موقعیت دقیق انگشت شست را شناسایی کند. برای مثال انگشت شست در دو نمونه از یک کلاس می تواند نسبت به سایر انگشتان در دو مکان بسیار متفاوت قرار گیرد. به طوری که فاصله سرانگشت شست با سرانگشتان دیگر در دو نمونه مقادیر بسیار متفاوت داشته باشد.

¹ Cross validation

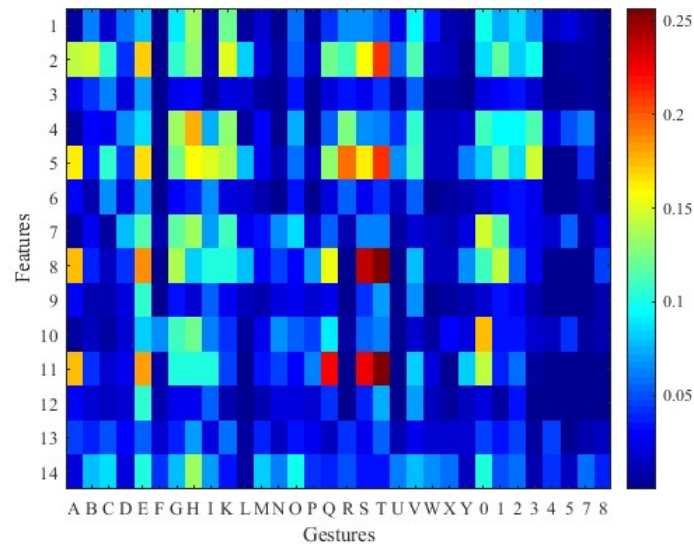


شکل ۱۵. زاویه‌های مفاصل هر انگشت (شماره‌های قرمز) و زاویه بین انگشت‌ها (شماره‌های آبی)



شکل ۱۶. واریانس ویژگی‌ها برای هر یک از حالت‌های دست

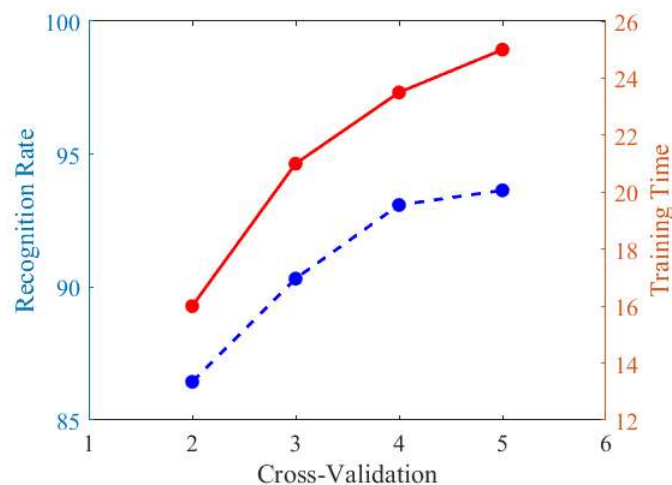
در شکل ۱۷ ویژگی‌های ۱۵ تا ۲۳ که کمترین و بیشترین واریانس را داشتند حذف شده است تا اختلاف بین سایر ویژگی‌ها با رنگ‌های متفاوت دیده شود. همانطور که در این شکل دیده می‌شود واریانس برای حروف S, E, A و T بالاست. دست در این چهار حروف حالت مشت دارد و همچنین واریانس حروف G, H, K, Q و 0 و 1 نسبت به سایر حروف بیشتر است. در این حروف بیش از یک انگشت دست در انسداده است. طبق نتایج بدست آمده سنسور LMC مدل سازی ضعیفی از انگشت‌هایی که در انسداده هستند ارائه داده است.



شکل ۱۲. واریانس ۱۴ ویژگی اول (زاویه مفصل هر انگشت) برای تمام حالت‌های دست

۴-۲- آزمایش با اعتبارسنجی‌های متقابل مختلف

اعتبارسنجی متقابل، رایج‌ترین روش برای ارزیابی کارایی یک سیستم است. در اعتبارسنجی متقابل k -لایه‌ای، کل پایگاه داده به k قسمت مساوی تقسیم می‌شود. از این k زیرمجموعه، هر بار یکی برای اعتبارسنجی و $k-1$ تای دیگر برای آموزش بکار می‌روند. این روال k بار تکرار می‌شود و همه داده‌ها دقیقاً یک بار برای آموزش و یک بار برای اعتبارسنجی بکار می‌روند. میانگین خطای حاصل از k آزمایش به عنوان خطای سیستم گزارش می‌شود. در این مقاله، آزمایش برای اعتبارسنجی‌های متقابل ۲، ۳، ۴ و ۵-لایه‌ای انجام می‌شود. نتایج بازشناسی و زمان صرف شده برای آموزش بر حسب ثانیه در شکل ۱۸ ارائه شده است. در تمام آزمایش‌ها ۳۰ درخت بکار برده شده است. هرچه لایه‌های اعتبارسنجی متقابل افزایش می‌یابد، نتایج بهتری برای بازشناسی حاصل می‌شود. ولی لایه‌های بیشتر برای اعتبارسنجی نیاز به زمان زیادی برای آموزش دارد. نتایج بازشناسی برای اعتبارسنجی متقابل ۴-لایه و ۵-لایه تفاوت زیادی باهم ندارند. ولی زمان صرف شده برای این دو اختلاف فاحشی دارند.



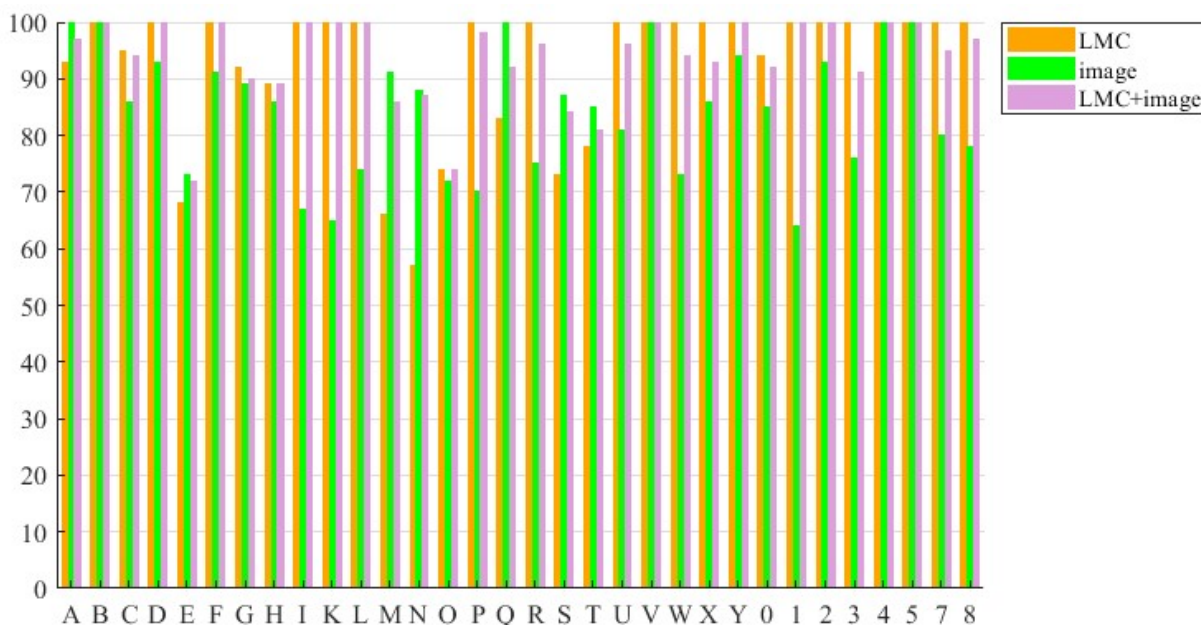
شکل ۱۸. نتایج بازشناسی و زمان آموزش برای اعتبارسنجی‌های متقابل مختلف

۳-۴- کارایی هریک از ویژگی‌ها

سه آزمایش انجام شده است که در دو مورد فقط یک دسته از ویژگی‌های بدست آمده از سنسور LMC و ویژگی‌های تصویر دست بدست آمده از تصاویر رنگی بکار گرفته شده است و در آزمایش سوم همه‌ی ویژگی‌ها بکار برده شده است. در همه‌ی آزمایش‌ها از اعتبارسنجی متقابل ۴-لایه‌ای استفاده شده است. نتایج بازشناسی برای هریک از کلاس‌های پایگاه داده و برای هر دسته از ویژگی‌ها جداگانه در شکل ۱۹ نشان داده شده است. همانطور که در این شکل دیده می‌شود نتایج بازشناسی با استفاده از ویژگی‌های اطلاعات سنسور LMC برای حروف E، M، N، O، S و T پایین است. در تمام این حروف دست حالت مشت دارد و همانطور که قبلاً گفته شد سنسور LMC نمی‌تواند وقتی دست مشت می‌شود محل دقیق انگشت شست را شناسایی کند. همچنین با تغییر زاویه دست در حروف H، G و Q درصد بازشناسی نسبت به بقیه حروف کاهش یافته است.

نتایج بازشناسی با استفاده از ویژگی‌های تصویر برای حروف P، K و I و عدد ۱ پایین‌تر از ۷۰ درصد است. جهت مچ دست در حروف K و P نزدیک به یکدیگر است. درصد بازشناسی برای حروف L، E، O، R، W و اعداد ۳ و ۸ بین ۷۰ تا ۸۰ درصد است. تصویر حرف E شباهت زیادی به کانتور حروف C و O دارد. همچنین تصویر حرف R شباهت زیادی به D و U دارد. نتایج نرخ بازشناسی با استفاده از ویژگی‌های تصویر در مقایسه با ویژگی‌های سنسور LMC در حروف M، N، S، A، E، Q و T بیشتر و بهتر شده است. همانطور که در این شکل مشاهده می‌شود ترکیب دو ویژگی نتایج بسیار خوبی را ارائه داده است. در واقع این دو گروه ویژگی مکمل یکدیگر هستند.

جدول ۲ میانگین نرخ بازشناسی را برای هریک از ویژگی‌ها حالت‌های دست در ژست‌های مشت و غیر مشت جداگانه نشان داده است. ویژگی‌های تصویر ژست‌های مشت را بهتر بازشناسی کرده است.



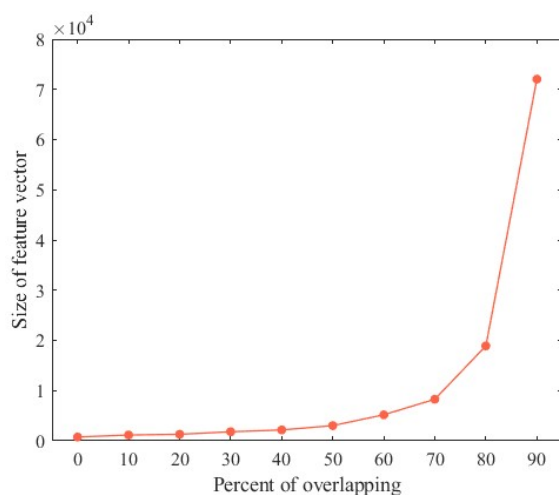
شکل ۱۹. نتایج بازشناسی برای هریک از حالت‌های دست با سه مجموعه ویژگی

جدول ۲. نتایج بازشناسی هریک از ویژگی‌ها روی حالت‌های مشت و غیر مشت دست

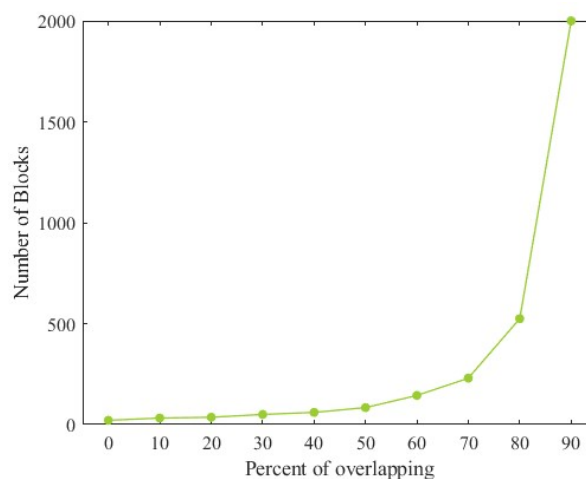
الگوریتم			درصد بازشناسی
کل پایگاه داده	حالت های مشت	حالت های غیر مشت	
۹۲/۵۶	۷۵/۳۸	۹۸/۲۹	ویژگی های داده های LMC
۸۴/۴۴	۸۵/۱۳	۸۴/۲۱	ویژگی های تصویر
۹۳/۶۹	۸۴/۱۳	۹۶/۸۸	کل ویژگی ها

۴-۴- بررسی اثر پارامترهای هیستوگرام گرادیان‌های جهت‌دار

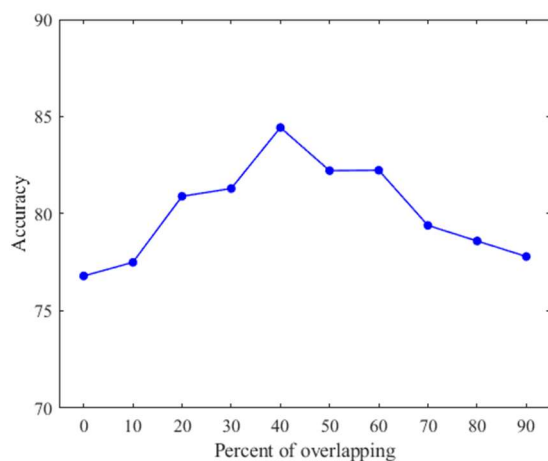
در این بخش اثر میزان همپوشانی بلوک‌های هیستوگرام گرادیان‌های جهت‌دار روی زمان پردازش، دقت بازشناسی، تعداد بلوک‌های تشکیل شده و اندازه بردار ویژگی بررسی شده است. این بررسی برای ۶۴۰۰۰ تصویر با اندازه‌ی ۱۵۲×۳۱۲ انجام شده است. اندازه سلول‌ها ۴×۴ ، اندازه بلوک‌ها ۱۰×۱۰ و تعداد جهت‌ها (ستون‌های هیستوگرام) ۹ در نظر گرفته شده است. همانطور که در شکل ۲۰ دیده می‌شود با افزایش مقدار همپوشانی، تعداد بلوک‌ها به طرز چشمگیری افزایش می‌یابد (شکل ۲۰ الف). رابطه بین اندازه بردار ویژگی، زمان محاسبه و تعداد بلوک‌ها رابطه‌ی خطی با یکدیگر دارد (شکل ۲۰ الف، ب، ج). اندازه بردار ویژگی از حاصلضرب تعداد ستون‌ها و تعداد سلول‌های هر بلوک و تعداد بلوک‌ها بدست می‌آید. بنابراین وجود رابطه‌ی خطی بین تعداد بلوک‌ها و اندازه بردار ویژگی قابل پیش‌بینی بود. زمان محاسبه هم با جمع زمان محاسبه برای بلوک بدست می‌آید. بنابراین رابطه بین تعداد بلوک‌ها و زمان محاسبه هم خطی خواهد بود. رابطه بین افزایش همپوشانی بلوک‌ها و درصد بازشناسی در شکل ۲۰ قسمت د نشان می‌دهد که با افزایش همپوشانی درصد بازشناسی افزایش می‌یابد. اما از یک نقطه به بعد این روند نزولی خواهد بود زیرا اطلاعات مضاعف زیادی ایجاد شده است و از آنجایی که جنگل تصادفی در نودها ویژگی‌ها را به صورت تصادفی انتخاب می‌کند ممکن است از تمام ناحیه‌ی تصویر ویژگی انتخاب نشود.



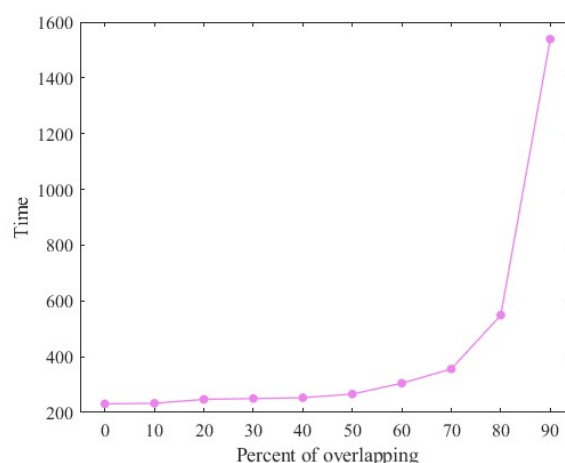
ب



الف



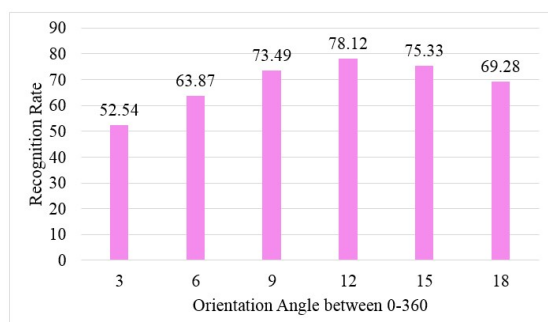
د



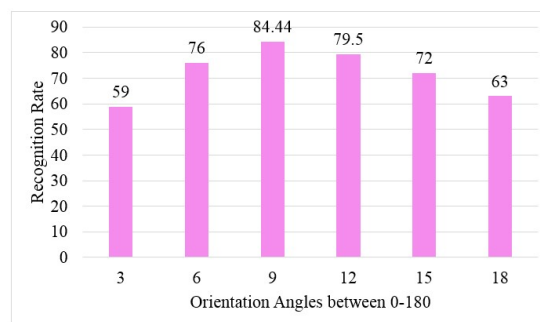
ج

شکل ۲۰. اثر افزایش همپوشانی بلوک‌های ویژگی هیستوگرام گرادینان‌های جهت‌دار بر (الف) تعداد بلوک‌ها، (ب) اندازه بردار ویژگی، (ج) زمان محاسبه و (د) دقت بازشناسی.

یکی دیگر از پارامترهای هیستوگرام گرادینان‌های جهت‌دار تعداد ستون‌هایی است که برای آن در نظر گرفته می‌شود. دو نوع زاویه می‌توان برای جهت گرادینان در نظر گرفت: زاویه بدون علامت که بین ۰ تا ۱۸۰ است و زاویه علامت‌دار که بین ۰ تا ۳۶۰ در نظر گرفته می‌شود. با توجه به شکل ۲۱ بالاترین درصد بازشناسی برای ۹ زاویه بدون علامت بدست آمده است.



ب



الف

شکل ۲۱. اثر تعداد ستون‌ها بر دقت بازشناسی.

۴-۵- مقایسه با روش های کانتوری دیگر

در این بخش، ویژگی پیشنهاد شده برای تصویر با دو روش تصویری دیگر مقایسه شده است. آزمایش ها روی پایگاه داده ساخته شده در این مقاله انجام شده و الگوریتم آنها روی این پایگاه داده پیاده سازی شده است. در روش ریورا-آکوستا و همکاران [۲۷]، ابتدا شیب نقاط کانتور دست بدست می آید. نقاطی شاخص در این روش نقاط هستند که در آنها علامت شیب تغییر می کند. هر کانتور دست تعداد متفاوتی نقاط شاخص دارد. برای همسان سازی تعداد نقاط شاخص، در مناطقی که تعداد نقاط شاخص بیشتر است تعداد را کاهش می دهد و به مناطق دیگر اضافه می کند. در آخر، مختصات نقاط شاخص به عنوان بردار ویژگی به طبقه بند داده می شود. این روش نسبت به نویز و چرخش های کوچک بسیار حساس است. فرایند حذف نقاط شاخص از مکان های پرتراکم برای یکسان سازی اندازه بردار ویژگی اطلاعات مفید را از بین می برد. این روش نه تنها روی حالت های مشت بلکه روی حالت های دیگر دست هم عملکرد چندانی ندارد.

در روش رن و همکاران [۲۶]، ابتدا جهت دست پیدا شده است. برای یافتن جهت دست روش چند-مقیاسی بکار برده شده است. در این روش ابتدا مستطیلی محیطی به کانتور دست رسم می شود. سپس، مستطیل به چند مستطیل با اندازه های مختلف تقسیم می شود. بنابراین مستطیل محیطی به سلول های زیادی تقسیم شده است. در هر سلول جهت کانتور محاسبه شده است. هر سلول با توجه به موقعیت آن وزن های متفاوتی می گیرند. برای مثال سلول هایی که در ناحیه دست قرار می گیرند و هیچ نقطه ی کانتوری را شامل نمی شوند، وزن صفر دریافت می کنند. سلول هایی که شامل نقاط محدب و مقعر کانتور دست می شوند، وزن های کمتری دریافت می کنند. در آخر، هیستوگرام جهت سلول ها محاسبه می شود. ستونی از هیستوگرام که بلندترین ستون را دارد به عنوان جهت دست انتخاب می شود. پس از تعیین جهت دست، با شروع از یک نقطه مرجع فاصله نقاط کانتور از مرکز دست به عنوان بردار ویژگی محاسبه می شود. با توجه به اینکه نقاط مقعر و محدب وزن کمتری دریافت می کنند در حالت هایی که دست مشت می شود، این روش به خوبی جهت دست را شناسایی نمی کند. زیرا تفاوت حالت های دست در کلاس هایی که به صورت مشت هستند دقیقاً در محل تحدب و تفرع نقاط کانتوری است.

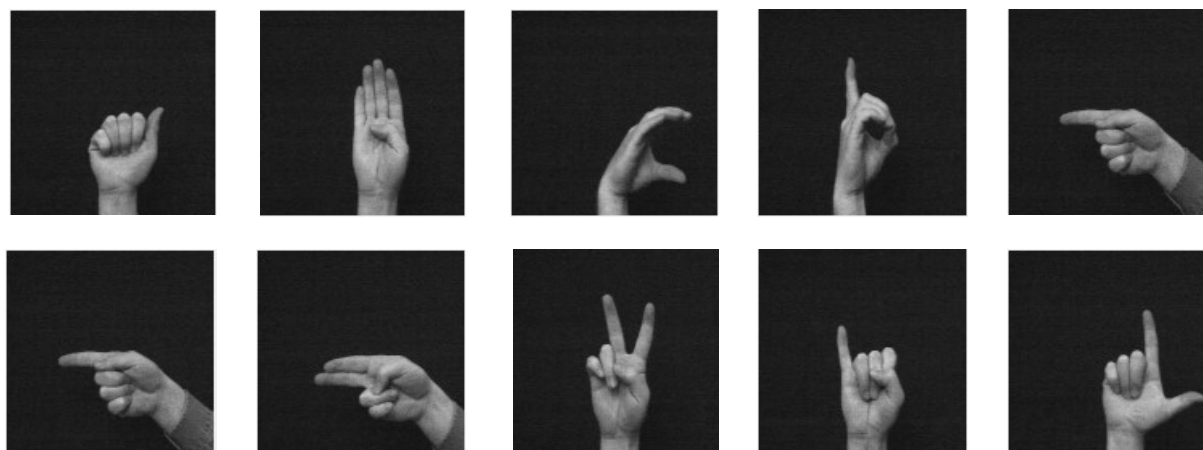
در روش پیشنهادی نه تنها از اطلاعات نقاط کانتوری بلکه از اطاعات بافت ناحیه دست هم ویژگی استخراج می شود. ترکیب فاصله نقاط کانتور از مرکز دست و هیستوگرام گرادینان های جهت دار ناحیه دست اطلاعات بیشتری برای بازشناسی فراهم می کند. در جدول ۳ روش پیشنهادی با دو روش ذکر شده مقایسه شده است. همانطور که در این جدول مشاهده می شود برای حالت های مشت، روش پیشنهادی عملکرد خوبی داشته است.

جدول ۳. مقایسه ی ویژگی پیشنهادی برای تصویر با دو روش تصویری دیگر روی پایگاه داده

الگوریتم	درصد بازشناسی		
	کل پایگاه داده	حالت های مشت	حالت های غیر مشت
نقاط شاخص [۲۷]	۵۷/۲۱	۴۰/۵۸	۶۵/۳۹
هیستوگرام سلول ها [۲۶]	۷۰/۲۸	۴۶/۱	۷۸/۳۴
روش پیشنهادی	۸۴/۴۴	۸۵/۱۳	۸۴/۲۱

۴-۶- آزمایش روی پایگاه داده‌ی دیگر

روش پیشنهادی روی پایگاه داده‌ی دیگری نیز آزمایش می‌شود. پایگاه داده‌ای که روش‌های زیادی روی آن آزمایش شده پایگاه داده‌ی Jochen Triesch [۳۰] است که شامل ۱۰ حالت دست است که توسط ۲۴ نفر با سه پس زمینه تاریک، روشن و پیچیده انجام شده است. کل پایگاه داده شامل ۷۱۸ نمونه است که نمونه‌هایی از آن در شکل ۲۲ نشان داده است. جدول ۴ نتایج آزمایش روش‌های مختلف را روی تصاویر این پایگاه داده با پس زمینه ساده را نشان می‌دهد. درصدهای ذکر شده در این جدول طبق اعداد گزارش شده این پژوهش‌ها است.



شکل ۲۲. نمونه‌هایی از تصاویر پایگاه داده‌ی Jochen Triesch

در جدول ۴ همانطور که مشاهده می‌شود الگوریتم‌هایی که از ترکیب چند ویژگی استفاده کرده‌اند نتایج بهتری دارند. کائو و همکاران از ترکیب ویژگی‌های هیستوگرام گرادیان‌های جهت‌دار، مفهوم شکل^۱ [۳۱] و از ماشین‌های بردار پشتیبان برای بازشناسی حالت‌های دست استفاده کرده‌اند. یولانگ و همکاران [۳۲] ناحیه دست را به نواحی کوچک‌تر تقسیم کردند. سپس از هر ناحیه یک مجموعه ویژگی به صورت سلسه مراتبی استخراج کردند. ژانگ و همکاران [۳۳] از ممان‌های زرنیک و هیستوگرام گرادیان‌های جهت‌دار و از سنجش (حسگری) فشرده^۲ برای طبقه‌بندی استفاده کردند. هانگ و همکاران [۳۴] از فیلترهای گابور برای استخراج ویژگی استفاده کرده‌اند و پس از کاهش ابعاد ماشین‌های بردار پشتیبان را در مرحله‌ی طبقه‌بندی اعمال کرده‌اند. تریش و همکاران [۳۰] پس از اعمال فیلترهای گابور و استخراج یک مجموعه نقطه خاص به تصویر یک گراف را اختصاص دادند. جاست و همکاران [۳۵] برای دسته‌بندی حالت‌های دست از تبدیل Census استفاده کردند.

جدول ۴. مقایسه‌ی روش‌های مختلف روی پایگاه داده‌ی Jochen Triesch

الگوریتم	درصد بازشناسی
[۳۵] MCT feature + AdaBoost	۸۹/۹۷
[۳۰] Elastic graph matching	۹۰
[۳۴] PCA + Garbor, SVM	۹۱/۵
[۳۳] Compressive Sensing + Zernike	۹۲/۱
[۳۱] MKL	۹۹/۱۷
[۳۲] BoF + Spectral-HIK	۹۹/۹۷
روش پیشنهادی	۱۰۰

^۱ Shape context

^۲ Compressive sensing

روش پیشنهادی روی سیستم ویندوز ۷، ۶۴ بیت، با مقدار حافظه ۸ گیگ و با پردازنده Core 7i با فرکانس 2.00GHz و با نرم افزار MATLAB 2018a پیاده سازی شده است.

۵- نتایج و بحث

آزمایش‌های مختلفی برای بررسی جنبه‌های مختلف روش پیشنهادی انجام گرفت. اولین آزمایش مطالعه روی پراکندگی داده‌های سنسور LMC بود. این آزمایش به این منظور انجام گرفت که مشخص شود یک ویژگی خاص بین نمونه‌های یک کلاس به چه میزان پایدار باقی مانده است. برای مثال شکل ۱۶ نشان می‌دهد که ویژگی ۲۰ در کلاس حرف H واریانس بالایی دارد زیرا در محدوده قمرز طیف قرار گرفته است. این امر نشان می‌دهد که بین ۲۰۰۰ نمونه‌ی کلاس H ویژگی شماره ۲۰ (فاصله انگشت شست با انگشت کوچک) مقادیر متفاوتی دارد. به طور مشابه ویژگی ۴ در کلاس حرف A در طیف رنگ آبی تیره است و واریانس بسیار پایینی دارد. به عبارت دیگر ویژگی شماره ۴ بین تمام نمونه‌های کلاس A مقادیر نزدیک به هم دارند. پس ویژگی شماره ۴ نسبت به ویژگی شماره ۲۰ قابل اعتمادتر و مقاوم‌تر است. با دقت در شکل ۱۶ مشاهده می‌شود که واریانس در قسمت پایین ماتریس پراکندگی یعنی ویژگی ۲۰ تا ۲۳ نسبت به سایر ویژگی‌ها مقادیر بالایی دارند. ویژگی ۲۰ تا ۲۳ از بردار ویژگی مربوط به فاصله سرانگشت شست با سایر سرانگشتان است. بنابراین سنسور LMC در مدل سازی موقعیت دقیق انگشت شست ضعیف عمل کرده است. انگشت شست نسبت به سایر انگشتان آناتومی پیچیده‌تری و آزادی عمل بیشتری دارد. واریانس ویژگی‌های ۱۵ تا ۱۹ که زاویه بین انگشتان را نشان می‌دهد (حرکت آبدوکیون و آدوکیون) بسیار پایین است. این امر به این دلیل است که به طور طبیعی زاویه بین انگشتان محدوده تغییر کمتری نسبت به زاویه‌ی بین مفاصل هر انگشت (حرکت فلکسیون و اکستنسیون) دارد. در طی این پژوهش برای ارائه بهتر واریانس بین ویژگی‌های ۱ تا ۱۴ که زاویه بین مفاصل هر انگشت را نشان می‌دهند، ویژگی‌های ۱۵ تا ۲۳، از طیف حذف شده است (شکل ۱۷). همانطور که در شکل ۱۷ دیده می‌شود واریانس برای حروف A، E، S و T و عدد 0 بالاست. دست در این چهار حروف حالت مشت دارد. بنابراین یکی دیگر از ضعف‌های سنسور LMC مواجه شدن با حالت‌های مشت دست است. همچنین واریانس حروف G، H، K، Q و عدد 1 نسبت به سایر حروف بیشتر است. در این حروف بیش از یک انگشت دست در انسداد است. سنسور LMC قادر به مدل سازی انگشتانی را که در انسداد هستند نیست.

اثر هریک از ویژگی‌های بدست آمده از سنسور LMC و تصویر دست جداگانه بررسی شد. با توجه به شکل ۱۹ ویژگی‌های بدست آمده از داده‌های سنسور LMC برای بازشناسی حالت‌های دست که به صورت مشت است (N، E، M، O، T و S) دقت کمتر از ۸۰ درصد دارد و برای حالت‌های دست که انگشتان در انسداد هستند و پشت دست روبروی دوربین قرار گرفته است (H و Q) عملکرد بین ۸۰ تا ۹۰ درصد دارد. نتایج بازشناسی با استفاده از ویژگی‌های تصویر برای حروف P، K و I و عدد 1 پایین‌تر از ۷۰ درصد و برای حروف L، E، O، R و W و اعداد 3 و 8 بین ۷۰ تا ۸۰ درصد است. کانتور I و 1، حروف K و P، حروف E و C و O شباهت زیادی به یکدیگر دارند برای همین درصد باز شناسی با استفاده از ویژگی‌های تصویر برای این دو حالت پایین است. تفاوت کلاس‌های K و P فقط در اختلاف اندکی در جهت مچ دست است. همچنین تصویر حرف R شباهت زیادی به D و U دارد. نرخ بازشناسی با استفاده از ویژگی‌های تصویر در مقایسه با ویژگی‌های سنسور LMC در حروف M، N، S، A، E، Q

T بهتر شده است. همانطور که در این شکل مشاهده می شود ترکیب دو ویژگی نتایج بسیار خوبی را ارائه داده است. در واقع این دو گروه ویژگی مکمل یکدیگر هستند.

اثر پارامترهای هیستوگرام گرادیان های جهت دار مانند میزان همپوشانی بلوک ها و تعداد ستون هایی که برای هیستوگرام در نظر گرفته می شود بررسی و مشاهده شد که همپوشانی بلوک ها بین ۴۰ تا ۶۰ درصد دقت بازشناسی را افزایش می دهد. اما میزان بالای همپوشانی علاوه بر افزایش اندازه ی بردار ویژگی و زمان محاسبه ی ویژگی، دقت بازشناسی را کاهش می دهد. همچنین بهترین درصد بازشناسی برای ۹ ستون (تعداد جهت های گرادیان) که بین ۰ تا ۱۸۰ درجه بود بدست آمد.

الگوریتم پیشنهادی و دو الگوریتم دیگر که بر مبنای کانکتور می باشند، بر روی پایگاه داده پیشنهادی این پژوهش و یک پایگاه داده دیگر پیاده سازی شده اند که به ازای هر دو پایگاه داده، الگوریتم پیشنهادی درصد بازشناسی بالاتری را کسب کرده است.

۶- نتیجه گیری

در این مقاله، سنسور کنترل کننده حرکت و تصویر دست برای بازشناسی زبان اشاره آمریکایی سه بُعدی بکار گرفته شده اند. سنسور کنترل کننده حرکت اطلاعات مربوط به محل سه بُعدی مفاصل دست را فراهم می کند. زاویه های مفاصل دست ویژگی متمایز کننده ای برای بازشناسی حالت دست محسوب می شوند. وقتی که دست مشت می شود یا در زاویه ی درستی نسبت به سنسور قرار نمی گیرد، اطلاعات سنسور کنترل کننده حرکت راجع به موقعیت مفاصل انگشتان قابل اعتماد نیست. ویژگی های استخراج شده از تصویر دست کمک می کند تا ژست دست دقیق تر بازشناسی شود. در این مقاله پایگاه داده ای شامل ۶۴ هزار نمونه ساخته شده است. پایگاه داده شامل ۳۲ کلاس (۲۴ حرف و ۸ عدد انگلیسی) است. هر نمونه شامل محل سه بُعدی مفاصل دست، جهت مچ دست، جهت کف دست و تصویر رنگی دست است. برای طبقه بندی از جنگل تصادفی استفاده شده است که نوعی یادگیری ماشین گروهی است و در آن تعداد زیادی درخت تصمیم گیری وجود دارد. هر درخت تصمیم گیری روی زیرمجموعه کوچتری از پایگاه داده آموزش داده می شود. همه درخت ها مستقل از یکدیگر هستند و می توانند همزمان آموزش داده شوند که این امر سرعت پردازش را بالا می برد. آزمایش های زیادی برای بررسی کارآمدی سیستم پیشنهادی انجام شده است. روش های دیگر روی پایگاه داده ی ما پیاده سازی شده اند و همچنین روش ما روی پایگاه داده های دیگران بررسی شده است. نتایج آزمایش نشان داد که ویژگی های استخراج شده از داده های سنسور LMC و ویژگی استخراج شده از تصویر دست مکمل یکدیگر هستند. نتایج آزمایش ها نشان می دهند که روش پیشنهادی نسبت به روش هایی که از تصاویر عمق استفاده می کنند بهتر عمل می کند. همچنین نتایج آزمایش ثابت کردند که استفاده از چند ورودی برای استخراج ویژگی های مکمل نتایج بازشناسی را بسیار بهبود می بخشد.

مراجع

- [1] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Computer Vision Image Understanding*, 2007, vol. 108, no. 1-2, pp. 52-73.
- [2] E. Togootogtokh et al., "3D finger tracking and recognition image processing for real-time music playing with depth sensors," *Multimedia Tools and Applications*, 2018, vol. 77, no. 8, pp. 9233-9248.
- [3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 886-893.
- [4] R. E. J. M. I. Schapire, "The strength of weak learnability," *Machine learning*, 1990, vol. 5, no. 2, pp. 197-227.
- [5] D. Tang, 3D hand pose regression with variants of decision forests, Imperial College London, 2015.
- [6] M. Ibrahim, "Scalability and performance of random forest based learning-to-rank for information retrieval," in *ACM SIGIR Forum*, 2017, vol. 51, no. 1, pp. 73-74.
- [7] L. Breiman, *Classification and regression trees*. Routledge, 2017.
- [8] M. Ristin-Kaufmann, *Large-Scale Image Recognition with Random Forests*, Doctoral dissertation, ETH Zurich, 2015.
- [9] M. Liu, M. Wang, J. Wang, and D. Li, "Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar," *Sensors Actuators B: Chemical*, 2013, vol. 177, pp. 970-980.
- [10] L. Zhang, Q. Ni, G. Zhang, M. Zhai, J. Moreno, and C. Briso, "Random forests-enabled context detections for long-term evolution network for railway," *IET Microwaves, Antennas Propagation*, 2019, vol. 13, no. 8, pp. 1080-1086.
- [11] C. Lei et al., "A comparison of random forest and support vector machine approaches to predict coal spontaneous combustion in gob," *Fuel*, 2019, vol. 239, pp. 297-311.
- [12] W. Nai, Y. Liu, D. Rempel, and Y. J. P. R. Wang, "Fast hand posture classification using depth features extracted from random line segments," *Pattern Recognition*, 2017, vol. 65, pp. 1-10.
- [13] C. Wang, Z. Liu, and S.-C. J. I. t. o. m. Chan, "Superpixel-based hand gesture recognition with kinect depth camera," *IEEE transactions on multimedia*, 2015, vol. 17, no. 1, pp. 29-39.
- [14] F. Pedersoli, S. Benini, N. Adami, and R. J. T. V. C. Leonardi, "XKin: an open source framework for hand pose and gesture recognition using kinect," *The Visual Computer*, 2014, vol. 30, no. 10, pp. 1107-1122.
- [15] N. Pugeault and R. Bowden, "Spelling it out: Real-time ASL fingerspelling recognition," in *IEEE International conference on computer vision workshops (ICCV workshops)*, 2011, pp. 1114-1119.
- [16] C. Zhang, Y. J. C. V. Tian, and I. Understanding, "Histogram of 3D facets: A depth descriptor for human action and hand gesture recognition," *Computer Vision and Image Understanding*, 2015, vol. 139, pp. 29-39.
- [17] C. Keskin, F. Kırac, Y. E. Kara, and L. Akarun, "Hand pose estimation and hand shape classification using multi-layered randomized decision forests," in *European Conference on Computer Vision*, 2012, pp. 852-863.
- [18] B. Feng et al., "Depth-projection-map-based bag of contour fragments for robust hand gesture recognition," *Transactions on Human-Machine Systems*, 2017, vol. 47, no. 4, pp. 511-523.
- [19] W. Aly, S. Aly, and S. Almotairi, "User-Independent American Sign Language Alphabet Recognition Based on Depth Image and PCANet Features," *IEEE Access*, 2019, vol. 7, pp. 123138-123150.
- [20] G. Marin, F. Dominio, P. J. M. T. Zanuttigh, and Applications, "Hand gesture recognition with jointly calibrated leap motion and depth sensor," *Multimedia Tools and Applications*, 2016, vol. 75, no. 22, pp. 14991-15015.
- [21] B. Hisham and A. J. J. Hamouda, "Arabic Static and Dynamic Gestures Recognition Using Leap Motion," *Journal of Computer Science*, 2017, vol. 13, no. 8, pp. 337-354.
- [22] P. Kumar, P. P. Roy, and D. P. J. I. S. Dogra, "Independent bayesian classifier combination based sign language recognition using facial expression," *Information Sciences*, 2018, vol. 428, pp. 30-

- 48.
- [23] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti, and C. J. I. T. o. M. Massaroni, "Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures," *IEEE Transactions on Multimedia*, 2019, vol. 21, no. 1, pp. 234-245.
- [24] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri, "A Modified-LSTM Model for Continuous Sign Language Recognition using Leap motion," *IEEE Sensors Journal*, 2019, vol. 19, no. 16, pp. 7056-7063.
- [25] H. Li, I. Wu, H. Wang, C. Han, W. Quan, and J. P. Zhao, "Hand gesture recognition enhancement based on spatial fuzzy matching in Leap Motion," *IEEE Transactions on Industrial Informatics*, 2019, vol. 16, no. 3, pp. 1885-1894.
- [26] Y. Ren, X. Xie, G. Li, Z. J. I. T. o. C. Wang, and S. f. V. Technology, "Hand gesture recognition with multiscale weighted histogram of contour direction normalization for wearable applications," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, vol. 28, no. 2, pp. 364-377.
- [27] M. Rivera-Acosta, S. Ortega-Cisneros, J. Rivera, and F. J. S. Sandoval-Ibarra, "American Sign Language Alphabet Recognition Using a Neuromorphic Sensor and an Artificial Neural Network," *Sensors*, 2017, vol. 17, no. 10, pp. 2176.
- [28] T. Pavlidis, *Algorithms for graphics and image processing*. Springer Science & Business Media, 2012.
- [29] A. Criminisi and J. Shotton, *Decision forests for computer vision and medical image analysis*. Springer Science & Business Media, 2013.
- [30] J. Triesch, C. J. I. von der Malsburg, and V. Computing, "Classification of hand postures against complex backgrounds using elastic graph matching," *Image and Vision Computing*, 2002, vol. 20, no. 13-14, pp. 937-943.
- [31] J. Cao, S. Yu, H. Liu, P. J. M. T. Li, and Applications, "Hand posture recognition based on heterogeneous features fusion of multiple kernels learning," *Multimedia Tools and Applications*, 2016, vol. 75, no. 19, pp. 11909-11928.
- [32] C. Yue-long, C. Ling, and C. J. J. o. Z. U. Gen-cai, "Hierarchical Bag-of-Features with Spectral-HIK filter based hand posture recognition," *Journal of ZheJiang University (Engineering Science)*, 2013, vol. 47, no. 9, pp. 1531-1536.
- [33] H. Zhang, H. Li, and M. J. J. o. H. U. Zhou, "Hand posture recognition based on multi-feature and compressive sensing," *Journal of Hunan University (Natural Sciences)*, 2013, vol. 40, no. 3, pp. 87-92.
- [34] D.-Y. Huang, W.-C. Hu, and S.-H. Chang, "Vision-based hand gesture recognition using PCA+ Gabor filters and SVM," in *fifth international conference on intelligent information hiding and multimedia signal processing*, 2009, pp. 1-4.
- [35] A. Just, Y. Rodriguez, and S. Marcel, "Hand posture classification and recognition using the modified census transform," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 2006, pp. 351-356.