مجله کنترل
ISSN (print) 2008-8345
ISSN (online) 2538-3752
جلد ۱۳، شماره ۳، پاییز ۱۳۹۸، صفحه ۳۹-۲۹

تشخیص الفبای دستی فارسی ناشنوایان مبتنی بر اطلاعات نرمال سازی شده در تصاویر ژرفا

شهاب رجبی[۱]، امیر موسوی‌نیا[۲]

[۱] فارغ‌التحصیل کارشناسی ارشد مهندسی برق، گروه الکترونیک، دانشگاه صنعتی خواجه نصیرالدین طوسی، sh.rajabi@ee.kntu.ac.ir
[۲] دانشیار، دانشکدهٔ مهندسی کامپیوتر، گروه معماری کامپیوتر، دانشگاه صنعتی خواجه نصیرالدین طوسی، moosavie@kntu.ac.ir

دریافت: ۱۳۹۷/۰۶/۱۰ ویرایش : ۱۳۹۷/۰۹/۱۰ پذیرش: ۱۳۹۷/۱۰/۱۰

چکیده: پس از ارائه دستگاه کینکت، محصول شرکت مایکروسافت گزارشهای بسیاری از کاربرد این دستگاه در تشخیص حالت دست و انگشتان منتشر شده‌است. در بیشتر این کاربرد ها اطلاعات ژرفا تنها برای جداسازی تصویر دست از پس زمینه استفاده شده و پردازش اصلی بر روی تصاویر ویدیویی و در فضای دو بعدی انجام شده‌است. در این مقاله روشی ارائه می‌شود که اطلاعات ژرفا نقش پر رنگ تری در پردارش دارند. با کمک روش آستانه گذاری مبتنی بر ژرفا، ابتدا قالب دست شخص در فضای سه بعدی استخراج می‌شود. سپس در فضای سه بعدی، راستای عمود برکف دست پیدا شده و با استفاده از ماتریسهای دوران و انتقال، این راستا با راستای دوربین همسو می‌شود. به این ترتیب دورانهای دست حول محورهای پیچ و یاو از تصویر حذف شده و با استفاده از ماتریس انتقال، تصویر دست در فاصله مشخصی از دوربین قرار می‌گیرد. در مرحله بعد، از دو ابزار تبدیل موجک و یک توصیفگر جدید به نام توصیفگر دایروی که در این سیستم معرفی شده‌است برای استخراج ویژگی ها استفاده می‌شود. یک شبکه های عصبی، غربالگری اولیه را در ویژگی های استخراج شده توسط تبدیل موجک انجام داده و سپس توصیفگر دایروی با استفاده از ماشین بردار پشتیبان بازشناسی حرف موردنظر را به اتمام می‌رساند. در آزمایشهای عملی با کمک اطلاعات برخط سنسور کینکت دقت شناسایی حروف الفبای فارسی ٪۹۶/۷ و تاخیر ۲ ثانیه برای هر علامت بدست آمده‌است.

کلمات کلیدی: الفبای ایستای فارسی ناشنوایان، سنسور کینکت، تبدیل موجک، توصیفگر دایروی، شبکه عصبی

Persian sign language detection based on normalized depth image information

Shahab Rajabi, Amir Mousavinia

Abstract: There are many reports of using the Kinect to detect hand and finger gestures after release of device by Microsoft. The depth information is mostly used to separate the hand image in the two-dimension of RGB domain. This paper proposes a method in which the depth information plays a more dominant role. Using a threshold in depth space first the hand template is extracted. Then in 3D domain the perpendicular vector to the hand surface is found. Using the rotation matrix all the rotations along three axes are compensated in a way that the camera z- coordinate lies perpendicular to hand surface. Then the resulted 3d image is translated to a distance of 80 to 100 cm from the Kinect. Wavelet transform with a new descriptor, called Circular Descriptor are used to extract required features. A trained MLP neural network in conjunction with a SVM is used to classify the signs. Empirical results show an average accuracy of 96.7 % with a two seconds delay for online recognition of Persian Sign Language.

Keywords: Persian sign language, deaf people, Kinect sensor, wavelet, circular descriptor.

۱- مقدمه

امروزه بازشناسی زبان اشاره ناشنوایان به یک حوزه تحقیقاتی در حال رشد تبدیل شده و تحقیقات زیادی در زمینه تشخیص ژست دست و شناسایی الفبای ناشنوایان صورت گرفته‌است. استفاده ازتصاویر ویدئو و دستکشهای رنگی و یا نشان‌گذاری شده بسیار متداول است. در اغلب این روش‌ها، مجموعه‌ای از بردارهای ویژگی، یک فضای با ابعاد بالاتر را تشکیل می‌دهند که درآن احتمال طبقه‌بندی بهتر الگو افزایش می‌یابد[۱]. مدلسازی هندسی دست با در نظر گرفتن ساختار اسکلت آن نیز با موفقیت برای شناسایی حرکات و موقعیت دست نسبت به دوربین استفاده شده‌است [۲،۳].

اخیرا همراه با ارائه حسگرهای ژرفای ارزان قیمت مانند کینکت، محصول شرکت مایکروسافت، بازشناسی زبان اشاره بر اساس اطلاعات تصویر ژرفا به دلیل ارائه یک مدل سه بعدی نسبتا دقیق و عدم نیاز به پوشیدن دست‌کش و یا رنگ آمیزی خاص، مورد توجه بسیاری از محققین قرار گرفته‌است. بازشناسی بر اساس اطلاعات ژرفا معمولا دقیقتر بوده و گستره‌ی بیشتری از لغات را در مقایسه با تصاوتر رنگی دو بعدی در بر می‌گیرد [۴،۵]. کینکت علاوه بر تصویر ژرفا می‌تواند به کمک دوربین رنگی خود که محورش قدری با محور دوربین ژرفا متفاوت است، تصاویر ویدیوی دو بعدی را نیز ارائه دهد.

ورما و همکارانش [۶] روشی را ارایه کرده‌اند که با استفاده از حسگر کینکت، همزمان با حرکت دست کاربر به داخل صحنه، ضبط و پردازش تصاویر شروع شده و تا زمانی که هر دو دست کاربر پایین برود این رویه را ادامه می‌دهد. سپس تمام این فریمهای ضبط شده برای ترجمه توسط سیستم استفاده می‌شوند.

آقای یانگ نیز با استفاده از تصویر ژرفای کینکت مایکروسافت و ارائه یک میدان تصادفی شرطی سلسله مراتبی، ابتدا علامتهای حرکات دست را شناسایی و سپس با کمک یک نقشه‌ی تقویت شده، تشخیص نهایی را انجام می‌دهد [۷].

در روشی دیگر بر مبنای کینکت، از یک کدگذاری برای مدل سازی اشارات مختلف دست استفاده می‌شود. در این روش با توجه به تفاوت بین اشاره ها، با انتخاب تعدادی از نمونه ها در هر دسته از نشانه‌ها آموزش انجام می‌شود. سپس در هر فریم از ویدئوی جدید میزان شباهت با نمونه های آموزش دیده بررسی می‌شود. در نهایت بهترین نمونه‌ها برای یک چارچوب فرمول‌بندی شده و بطور همزمان یک طبقه‌بند ویدیویی برای تشخیص علائم تولید می‌شود[۸].

بنگیو نیز از اطلاعات به دست آمده از ژرفا توسط کینکت، با الگوریتم SAE[1] به فرآیند آموزش ویژگی‌ها پرداخته و با استفاده از الگوریتم آنالیز مولفه‌های اصلی[2] تحت طبقه‌بندی ماشین بردار پشتیبان[3] دقت نتایج خود را افزایش می‌دهد[۹]. در روشی دیگر از یک شبکه عصبی کانولوشن به جای ساختن ویژگی های پیچیده استفاده شده که به طور خودکار ویژگی ها را استخراج می کند و توانسته است با دقت ۹۱/۷٪ الفبای اشاره ایتالیایی را با استفاده از کینکت به دست آورد.

روش‌های اشاره شده غالبا اطلاعات زمانی را مدل می‌نمایند و قادر به انطباق گستره وسیعی از کلمات اشاره نیستند. برای حل این مشکل با توجه به اینکه حافظه طولانی کوتاه مدت می‌تواند اطلاعات توالی زمانی را به خوبی مدل کند، در [۱۰] با روش هدف-به-هدف، بازشناسی زبان اشاره بر اساس LSTM[4] پیشنهاد شده‌است. این روش با توجه به مسیرهای حرکتی، چهار مفصل اسلکتی به عنوان ورودی و بدون در نظر گرفتن دانش پیشین و عاری از ویژگی‌های صریح استفاده می‌شود.

آلمدیا روشی برای بازشناسی الفبای برزیلی ناشنوایان ارایه داده است که در ابتدا هفت ویژگی بینایی از تصاویر RGB به دست می‌آورد. هر ویژگی مربوط به یک، دو یا سه عنصر ساختاری در الفبای اشاره برزیلی[5] است. سپس رابطه بین ویژگی‌های استخراج شده و عناصر ساختاری بر اساس شکل حرکت و موقعیت آن محاسبه می‌شود. در پایان بوسیله طبقه‌بند بردار پشتیبان، بازشناسی انجام می‌شود. نتایج دقت ۸۰٪ را در این الگوریتم نشان می دهد[۱۱].

فیلترهای ذره‌ای نیز در بازشناسی ژست انسان نقش ویژه‌ای دارند. لیم در یک تحقیق، یک فیلتر سریالی مبتنی بر ماتریس کوواریانس را برای تشخیص زبان اشاره پیشنهاد داده است. پس از انجام مراحل پیش پردازش ورجداسازی دست، با کمک این فیلتر دست در دنباله‌ای از فریم‌ها علامت‌گذاری شده و به صورت هم‌زمان هر دو دست ردیابی می‌شود. سپس با تولید ماتریس کوواریانس ویژگی اطراف ناحیه ردیابی شده و کاهش ابعاد، بازشناسی زبان اشاره آمریکایی با نرخ ٪۸۷/۳۳ انجام می‌دهد[۱۲].

ژایو نیز مدل رفتاری را با یک گراف برچسب‌گذاری نموده و طبقه‌بندی را با تطابق گراف از یک پایگاه داده و تصویر ورودی انجام می‌دهد[۱۳]. برای جلوگیری از پیچیدگی‌های این تطابق، یک کرنل گراف برای افزایش سرعت و هم‌چنین دقت بالاتر معرفی می‌شود. برای جلوگیری از چالش تشخیص علامت‌ها در توالی‌های به هم پیوسته ویدیویی از یک رویکرد در مدل‌سازی علامت‌های زیرمجوعه‌ای استفاده شده است. این چارچوب با ترکیبی از تکنیک‌های خوشه‌بندی فضایی-زمانی و انحراف پویای زمانی کار می‌کند. از آنجایی که زبان اشاره شامل هر دو بردار ویژگی فضا و زمان است از پراکندگی زمان پویا برای اندازه‌گیری فاصله بین دو نشانه مجاور استفاده می‌شود. این فاصله به عنوان یک بردار ویژگی زمانی در هنگام خوشه‌بندی بردارهای ویژگی فضایی با استفاده از خوشه‌بندی مینیمم بی‌نظمی[6] مورد استفاده قرار می‌گیرد. این فرآیند به صورت بازگشتی انجام می‌شود تا تمام حرکات میانی را به‌صورت پویا بدون استفاده از مدل‌سازی صریح یا ضمنی خوشه‌بندی کند[۱۴].

در زمینه تشخیص الفبای ناشنوایان فارسی گزارش‌های زیادی وجود ندارد. اما یکی از بهترین سیستم‌های موجود در [۱۵] معرفی شده که با استفاده از تبدیل موجک و شبکه‌های عصبی، از تصاویر ویدیوی حاصل از یک دوربین دیجیتال اقدام به بازشناسی الفبای ناشنوایان می‌نماید. ابتدا قسمت‌های اضافه‌ی عکس‌های رنگی حذف شده، اندازه‌ی آنها تغییرکرده و به حالت سیاه و سفید تبدیل می‌شوند. سپس تبدیل

[1] Sparse-Auto Eencoder
[2] Principal Component Analysis
[3] Support Vector Machine
[4] Momory Term – short Long
[5] Brazilian Sign Language
[6] Minimum Entropy Clustering

موجک گسسته بر روی عکس‌های سیاه و سفید اعمال شده و تعدادی از ویژگی‌ها استخراج می‌شوند. در نهایت ویژگی‌های استخراج شده برای آموزش یک شبکه عصبی پرسپترون چند لایه استفاده شده‌است. در سیستم استفاده شده از هیچ نوع دستکش یا سیستم‌های نشانگر استفاده نشده و این سیستم تنها نیاز به تصاویر دست خالی برای بازشناسی نیاز دارد. لازم است که پس زمینه تصاویر مشکی باشد و تنها دست و آن هم در امتداد عمود بر دوربین در تصویر قرار گرفته باشد. نتایج تجربی نشان می‌دهد که این سیستم قابلیت شناسایی ۳۲ حرف الفبای فارسی انتخاب شده را به دقت طبقه‌بندی ۹۴/۰۶٪ را دارد.

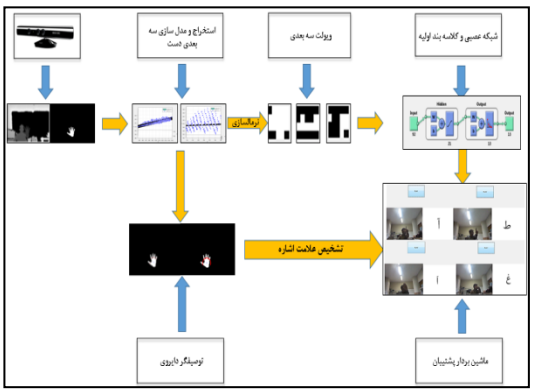
در اغلب روش‌های ارائه شده، محدودیت‌هایی برای کاربران مانند پوشیدن دستکش، ویا الزامات محیطی مانند وجود پس زمینه ساده وجود دارد. روشهایی با استفاده از فضای رنگی معرفی شدند که همانطور که مشخص هست در فضای نویزی دقت این الگوریتم ها به شدت کاهش می‌یابد[۱۶، ۱۷]. تعدادی از الگوریتم‌ها نیز از لحاظ زمانی و سنگینی محاسبات رنج می‌برند و برای برنامه های کاربردی و بلادرنگ نامناسب است[۱۸، ۱۹]. به علاوه در تعداد زیادی از الگوریتم‌های شناسایی حالت دست و هم‌چنین بیشتر الگوریتم‌های کارا در زمینه شناسایی الفبای فارسی ناشنویان، وابستگی زیادی به زاویه دست کاربر نسبت به دوربین دارند و از این رو اکثر آن‌ها از یک مجموعه داده خاص با پس زمینه ساده استفاده نموده‌اند[۲۰، ۲۱].

در این مقاله سیستمی معرفی می‌شود که بدون اعمال محدودیت پوششی برای دست افراد ناشنوا و با دقتی مناسب برای پیاده سازی آنی، شناسایی الفبای فارسی ناشنوایان در یک محیط معمولی و واقعی توسط سنسور کینکت انجام می‌شود. ابتدا با کمک اطلاعات ژرفا ناحیه اطراف دست استخراج شده و به این ترتیب پس زمینه حذف می شود. مرکز دست با اجرای الگوریتمهای متعارف مورفولوژی و محاسبه مرکز جرم بدست می‌آید. سپس با استفاده از تبدیل موجک و یک توصیف‌گر جدید که در این مقاله معرفی می‌شود، استخراج ویژگی انجام می‌پذیرد. برای کلاسه‌بندی از دو روش شبکه عصبی و بردارهای پشتیبان استفاده شده-است. در نهایت از یک فیلتر مدین برای حذف نویز در نتایج خروجی بازشناسی استفاده شده تا عملیات شناسایی حروف به درستی صورت گیرد.

در ادامه این مقاله و در بخش دوم کلیات سیستم پیشنهادی شامل روش بدست آوردن صفحه گذرنده از کف دست و چگونگی محاسبه مرکز و شعاع کف دست توضیح داده می شود. روش استخراج ویژگی در بخش سوم و توصیف گر دایروی در بخش چهارم بررسی می‌شوند. سرانجام نتایج شبیه سازی، مقایسه و جمع بندی نیز فصول پنجم و ششم را تشکیل می‌دهند.

۲- معرفی سیستم پیشنهادی

در این مقاله از ابزار کینکت که توسط شرکت مایکروسافت برای کنسول بازی ایکس باکس تهیه شده‌است، استفاده می‌نماییم. شکل ۱ بلوک دیاگرام سیستم ارائه شده را نشان می دهد. روش پیشنهادی از چهار بخش اصلی تشکیل می‌شود. در بخش نخست و با استفاده از سنسور کینکت دست کار بر از پس زمینه جدا می‌شود. در بخش دوم با استفاده از تکنیک رگرسیون سه بعدی، مختصات صفحه گذرنده از کف دست کاربر در فضای سه بعدی به دست آمده و سپس با دوران و انتقال دستگاهها در فضای سه بعدی راستای عمود بر کف دست با راستای دوربین کینکت موازی می‌شود. در بخش سوم و به شکل موازی دو سری ویژگی هندسی از دست استخراج می‌شود. ویژگی نخست از توصیف‌گر دایروی پیشنهادی این مقاله و ویژگی دوم از ضرایب تبدیل موجک اطلاعات دست در فضای سه‌بعدی به‌دست می‌آیند. در بخش چهارم نیز با ترکیب همزمان شبکه عصبی و ماشین بردار پشتیبان شناسایی انجام می‌شود. در ادامه هر یک از مراحل فوق با جزییات بیشتر شرح داده می‌شوند.



شکل ۱) بلوک دیاگرام روش پیشنهادی

۲-۱ استخراج دست انسان از تصاویر ژرفای

کاربر در مقابل دستگاه کینکت و در فاصله ۱۳۰ تا ۱۸۰ سانتی متری قرار گرفته و دست خود را در مقابل بدن و در برابر دستگاه کینگت قرار می‌دهد. این فاصله به شکل تجربی و با توجه به خروجیهای دستگاه کینکت بدست آمده است. اینک در تصویر عمق با انتخاب یک آستانه مناسب مبنتی بر ژرفا، β ، می توان دست را از پس زمینه جدا کرد. این مقدار آستانه با توجه به تفاوت‌های ناچیز در اندازه دست کاربران مختلف، مقدار تقریبا ثابتی بر اساس آزمایش‌ها به دست آمده است و بر اساس قطر متوسط دست انسان نرمال سازی شده‌است. اگر $D(n)$ تصویر عمق بدست آمده از دستگاه کینکت باشد و th ضخامت متوسط دست انسان در اینصورت رابطه ۱ روش جداسازی دست از پس زمینه را در تصویر ژرفا نشان میدهد.

$$I(n) = \{D(n): D_{min} \le D(n) \le D_{min} + th\} \qquad (۱)$$

در این رابطه n شماره پیکسل در تصویر ژرفا ، D_{min} مقدار ژرفای نزدیکترین جسم به دستگاه کینکت (فرض شده که نزدیکترین جسم به دستگاه دست کاربر باشد) و $I(n)$ تصویر عمق دست است. th در سیستم پیشنهادی و برای یک دست متعارف مقدار ۷۵ میلیمتر انتخاب شده‌است. در شکل ۲ یک نمونه تصویر ویدیو و ژرفای بدست آمده از دوربین کینکت قبل از استخراج دست مشاهده می‌شود. با استفاده از روش پیشنهاد شده تصویر ژرفای دست مطابق شکل ۳ استخراج شده-است. برخی مواقع بعلت وجود نویز بخشهای کوچک دیگری نیز در خروجی آستانه گیر مشاهده می شود. برای رفع این مشکل بزرگترین جسم را بعنوان دست انتخاب می کنیم.

شکل ۲) تصویر اتخاذ شده از دوربین کینکت: بالا تصویر ویدیو. پایین تصویر ژرفا

شکل ۳) دست استخراج شده از شکل ۲ با استفاده از روش آستانه‌گذاری در فضای ژرفا

۲-۲- استخراج صفحه گذرنده از کف دست

دست کاربر می‌تواند در فضای آزاد در راستاهای مختلف قرار گیرد که دوران راستای نسبی انگشتان دست نسبت به دوربین کینکت را تغییر خواهد داد. برای حل مشکل ابتدا با کمک رگرسیون خطی صفحه فرضی عبور کننده ازکف دست در فضا شناسایی شده و سپس به کمک ماتریس‌های دوران و انتقال، راستای عمود بر این صفحه با محور کینکت در یک راستا قرار می‌گیرد. به این ترتیب اثرات مربوط به چرخش در سمت (یاو) و ارتفاع (پیچ) حذف خواهد شد. برای حذف اثر چرخش حول محور طولی (رول) نیز از توصیف گر دایروی که در بخش بعدی معرفی می شود استفاده خواهد شد. رابطه ۲ معادله یک صفحه در فضای سه بعدی را نشان می‌دهد[۲۲، ۲۳].

$$ax + by + z + d = 0 \qquad (۲)$$

فاصله یک نقطه با مختصات $(x_p,y_p,z_p)^T$ تا صفحه فوق نیز به کمک رابطه ۳ محاسبه میشود:

$$D = \frac{ax_p + by_p + z_p + d}{\sqrt{a^2 + b^2 + 1}} \qquad (۳)$$

برای بدست آوردن مشخصات بهترین صفحه‌ای که از نقاط مرتبط با کف دست در تصویر ژرفا می گذرد یک مساله بهینه سازی برای مجموعه نقاط در فضا طرح می‌شود. اگر تعداد نقاط استخراج شده در تصویر ژرفای دست را R در نظر بگیریم، نرم مجموع فواصل این نقاط نسبت به صفحه، D_T، از رابطه ۴ بدست می‌آید:

$$D_T = \sum_{i=1}^{R} |D_i|^2 = \sum_{i=1}^{R} \left| \frac{ax_i + bx_i + z_i + d}{\sqrt{a^2 + b^2 + 1}} \right|^2 \qquad (۴)$$

با مشتق گیری نسبت به مولفه های a و b و d سه رابطه حاصل می شود که در روابط ۵-۷ نشان داده شده‌است.

$$\frac{dD_T}{db} = 0 \longrightarrow \sum_{i=1}^{R} \begin{matrix} (y_i{}^2)b^3 + [y_i(ax_i + z_i + d - 1)]b^2 + \\ (a^2y_i{}^2 + y_i{}^2 - ax_i - z_i - d)b + \\ (a^3x_iy_i + a^2y_iz_i + a^2y_id + ax_iy_i + z_iy_i + dy_i) \end{matrix} = 0 \qquad (۵)$$

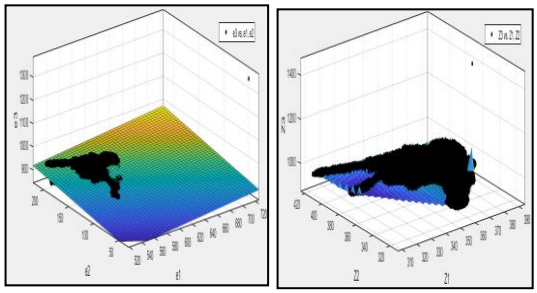
$$\frac{dD_T}{da} = 0 \longrightarrow \sum_{i=1}^{R} \begin{matrix} (x_i)^2a^2 + [x_i(by_i + z_i + d - 1)]a^3 + \\ (b^2x_i{}^2 + x_i{}^2 - by_i - z_i - d)a + \\ (b^3y_ix_i + b^2x_iz_i + b^2x_id + bx_iy_i + z_ix_i + dx_i) = 0 \end{matrix} \qquad (۶)$$

$$\frac{dD_T}{dd} = 0 \longrightarrow \sum_{i=1}^{R} ax_i + by_i + z_i + d = 0 \qquad (۷)$$

با قرار دادن سه معادله ۵، ۶ و ۷ در یکدیگر، مولفه های a و b و هم‌چنین d مطابق با معادله ۸ محاسبه می شود:

$$d = \frac{\sum_{i=1}^{R} ax_i + by_i + z_i}{R} \qquad (۸)$$

نرم افزار MATLAB توسط محیط گرافیکی Cftool می‌تواند این مساله را حل نماید. شکل ۴ صفحات استخراج شده برای دو مجموعه نمونه از نقاط مرتبط با تصاویر ژرفای دست را نشان می‌دهد. صفحه پیشنهاد شده به رنگ تیره و پیوسته در تصویر قابل تشخیص هستند.



شکل ۴) استخراج صفحه عبور کننده از کف دست در فضای سه بعدی

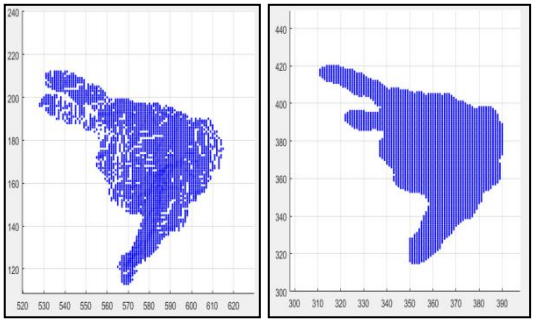
پس از این مرحله، با استفاده از زوایای صفحه تقریب زده شده نسبت به بردارهای یکه دستگاه مختصات کینکت و نیز اعمال ماتریس‌های دوران ارائه شده برای دو محور x و y مطابق رابطه (۹) بر تصویر ژرفای

دست، می‌توان چرخش و زاویه دست ناخواسته کاربر نسب به محورهای مختصات x و y را حذف نمود[۲۴].

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \quad (۹)$$

$$R_y = \begin{bmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{bmatrix}$$

شکل ۵ نتیجه انجام دوران بر روی مجموعه نقاط سه بعدی مرتبط با شکل ۴ را با بزرگنمایی نشان می‌دهد.



شکل ۵) هم‌راستا سازی اطلاعات ژرفای دست با اعمال ماتریس دوران بر نقاط سه بعدی برای شکل ۴

پس از حذف دوران از اطلاعات ژرفای دست لازم است که اندازه دست نیز با توجه به معیارهای هندسی موجود در دست انسان نرمال‌سازی شود. به این ترتیب اثر تغییرات فاصله تا دوربین و یا بزرگی و کوچکی دست نیز حذف خواهند شد. معادلات ۱۰ و ۱۱ روش نرمال‌سازی شعاع کف دست و فاصله آن تا دوربین کینکت را نشان می‌دهند:

$$N_r = \frac{R_{New}}{R_{Refrence}} \quad (۱۰)$$

$$N_D = \frac{D_{New}}{D_{Refrence}} \quad (۱۱)$$

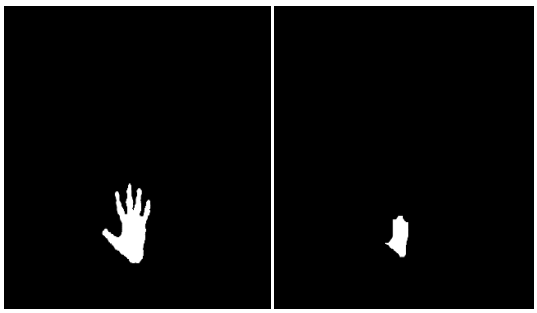
در روابط فوق R_{New} شعاع کف دست کاربر و $R_{Refrence}$ شعاع کف دست مرجع و N_r ضریب نرمال سازی شعاعی است. به همین ترتیب N_D ضریب نرمال سازی ژرفا، D_{New} فاصله کف دست کاربر با کینکت و $D_{Refrence}$ فاصله کف دست مرجع نسبت به دوربین کینکت است. برای انجام این عمل ابتدا لازم است که مرکز و شعاع دست بدست آیند.

۲-۳- استخراج مرکز و شعاع کف دست

امروزه از الگوریتم‌های مورفولوژی در علم پردازش تصویر برای کاربردهای گوناگونی از جمله شناسایی محل ترک خوردگی، نقاط آسیب دیده و هم‌چنین ترمیم تصویر استفاده می‌شود. در این سیستم استخراج مرکز و شعاع کف دست با استفاده از تکنیک سایش انجام شده‌است. سایش تصویر A بر اساس روابط ریاضی در تصاویر باینری توسط ساختار عنصر B از رابطه (۱۲) پیروی می‌کند.

$$A \ominus B = \{ z \in E | B_z \subseteq A \} \quad (۱۱)$$

در این رابطه E یک فضای اقلیدسی و A یک تصویر باینری است. B_z در واقع برگردان B بوده و با حرکت روی تصویر باینری آن را سایش میدهد. شکل ۶ تصویر کف دست را قبل و بعد از عملیات سایش نشان میدهد. همان‌طور که مشاهده می‌شود، توده‌ی کف دست انسان به جا مانده و انگشت‌های در شکل ۶ ب حذف شده‌است.



(ب) (الف)

شکل ۶) عملیات سایش :الف)تصویر دو سطحی دست قبل از سایش ب)استخراج توده کف دست بوسیله عملیات سایش

برای به دست آوردن مرکز کف دست ابتدا با استفاده از اصل به هم پیوستگی، توده‌های V را به ترتیب بر اساس بالاترین بهم پیوستگی تا پایین‌ترین آن برچسب گذاری می‌کنیم. این برچسب گذاری در حقیقت نماینده وزن آن توده به هم پیوسته می باشد و در بازه $[1, n]$ قرار می‌گیرد.

$$W_V = (n - V + 1)X \quad (۱۲)$$

در این رابطه n تعداد توده‌های شامل یک یا چند پیکسل در تصویر شامل دست بوده و X پایین‌ترین وزن یک توده در تصویر می‌باشد. در رابطه (۱۳) مجموع وزن های موجود در تصویر را مشاهده می‌نمایید.

$$\sum_{V=1}^{n} W_V = 1 \quad (۱۳)$$

این رابطه می‌تواند بصورت رابطه زیر ساده شود:

$$\sum_{L=0}^{n-1} (n - L)X = 1 \quad then \; X = \frac{1}{\sum_{L=0}^{n-1}(n - L)} \quad (۱۴)$$

سپس مرکز توده O از رابطه ۱۵ محاسبه می شود :

$$O = \frac{\sum_{V=1}^{n} \sum_{K=1}^{CV} W_V q_{k_v}}{\sum_{V=1}^{n} W_V C_V} \quad (۱۵)$$

در این رابطه C_V ، تعداد پیکسلهای توده v ام و q_{k_v} مختصات پیکسل k ام از خوشه v ام را نشان می‌دهد. پس از به دست آوردن مرکز دست، به محاسبه شعاع کف دست، r می‌پردازیم. مقدار r را به صورت بهینه از

کمترین مقدار ممکن غیر صفر آغاز نموده و تابع هزینه را مطابق رابطه (۱۶) در نظر می گیریم.

(۱۶)

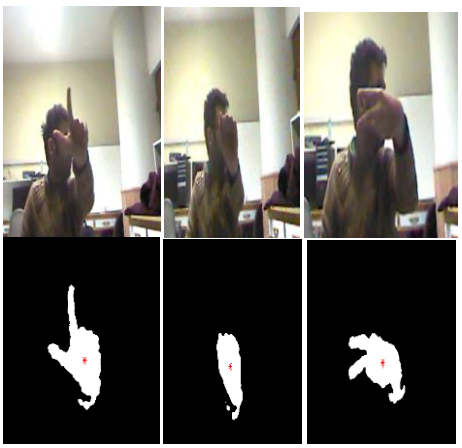
$$J(r)=\sum_{\theta=0}^{2\pi-Res(\theta)}|(r-rp_\theta)(Cos(\theta)+Sin(\theta)|^2$$

مقدار $Res(\theta)$ در واقع رزولوشن اندازه‌گیری بر محیط دایره رسم شده را نشان می دهد و هر چه مقدار کوچکتر انتخاب شود، r با دقت بهتری محاسبه شده و زاویه θ با مقدار آن در رابطه $J(r)$ افزایش می‌یابد. rp_θ در واقع شعاع در راستای زاویه θ می‌باشد که مقدار آن با افزایش طول در راستای زاویه θ، تا جایی که به اولین مقدار غیر صفر در آن توده برسد، محاسبه می‌شود و مقدار ثابتی برای هر یک از زوایای مورد نیاز به دست می‌آید. در رابطه (۱۸) با مشتق گیری نسبت به r تابع هزینه را برابر با صفر قرار داده و از آن طبق روابط (۱۹) مقدار r بهینه محاسبه می‌شود. .شکل ۵ مرکز دست استخراج شده را برای سه نمونه نشان می دهد.

(۱۷)

$$J'(r)=0 \quad \sum_{\theta=0}^{2\pi-Res(\theta)}(-rp_\theta)(Cos(\theta)+Sin(\theta))^2(r-rp_\theta)=0$$

$$\rightarrow \quad r=-\frac{\sum_{\theta=0}^{2\pi-Res(\theta)}(rp_\theta(Cos(\theta)+Sin(\theta))^2}{\sum_{\theta=0}^{2\pi-Res(\theta)}rp_\theta(Cos(\theta)+Sin(\theta))^2}$$



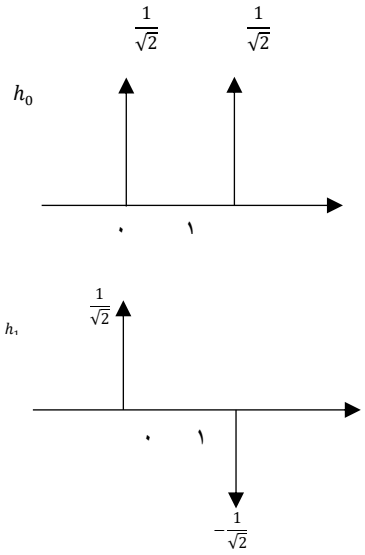
شکل ۷) استخراج مرکز دست برای سه نمونه از حروف الفبای ناشنوایان

۳- استخراج ویژگی

پس از طی مراحل ساده‌سازی تصویر و پیدا نمودن مدل هندسه دست، حال نوبت به استخراج ویژگی می‌رسد. در این روش، ابتدا با استفاده از تبدیل موجک هار یک فرآیند استخراج ویژگی صورت گرفته سپس از یک توصیف‌گر دایروی برای قدرتمندتر نمودن ابزار شناسایی استفاده می شود که است انگشتان ایستاده در تصویر را بیابد. خروجی هر یک از این دو ابزار استخراج ویژگی به ترتیب توسط شبکه‌های عصبی و بردارهای پشتیبان دسته‌بندی شده که در ادامه به تشریح آن‌ها خواهیم پرداخت.

۳-۱ تبدیل موجک هار برای استخراج ویژگی

در این مقاله از تبدیل موجک هار برای کدکردن تصاویر ژرفای حروف الفبای فارسی برای خانواده‌های مختلف تعریف شده استفاده شده‌است. شکل ۸ بردارهای پایه h_0 و h_1 در بانک فیلترهای ویولت هار را نشان می‌دهد.



شکل۸) ضرایب پایه های برداری ویولت هار در بانک فیلتر

برای اجرای یک مرحله از تبدیل ویولت، تقریب تصویر و جزییات افقی و عمودی و قطری تصویر برای یک تصویر F با ابعاد m*n روابط (۱۸) تا (۲۱) برقرار است.

(۱۸)

$$A_1=\frac{1}{2}(F(2m,2n)+F(2m-1,2n)+F(2m,2n-1)+F(2m-1,2n-1))$$

(۱۹)

$$H_1=\frac{1}{2}(F(2m,2n)+F(2m-1,2n)-F(2m,2n-1)-F(2m-1,2n-1))$$

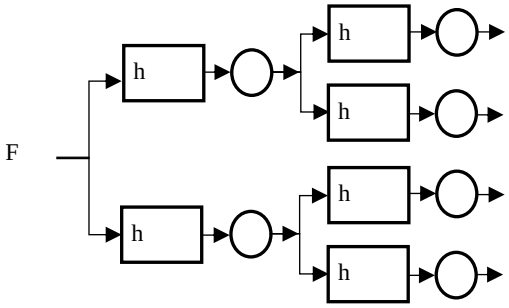
(۲۰)

$$V_1=\frac{1}{2}(F(2m,2n)-F(2m-1,2n)+F(2m,2n-1)-F(2m-1,2n-1))$$

(۲۱)

$$D_1=\frac{1}{2}(F(2m,2n)-F(2m-1,2n)-F(2m,2n-1)+F(2m-1,2n-1))$$

نمای کلی از تبدیل ویولت بر اساس بانک فیلتر در شکل ۹ نشان داده شده‌است.



شکل۹) شمای کلی تبدیل ویولت استفاده شده[۱۵]

ابتدا سایز هر فریم ورودی حاصل از پردازش تصاویر عمق را به مقدار ۳۰۰×۲۰۰ تغییر داده تا برای اعمال تبدیل موجک آماده شود. فرآیند گفته شده را عملیات پیش‌پردازش جهت آماده‌سازی تصاویر برای اعمال تبدیل موجک می‌نامیم که در شکل ۱۰ نمونه‌ای از آن قابل مشاهده است.

شکل۱۰) نصویر کف دست بر مبنای اطلاعات ژرفا

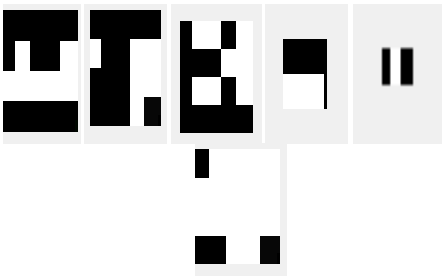
سپس تبدیل موجک اعمال می‌شود. بر اساس آزمایش‌های متعدد و به دلیل سادگی و کارا بودن در این تحقیق، از تبدیل موجک هار استفاده شده‌است و بعد از ۷ مرحله تبدیل موجک که هر مرتبه روی تقریب موجک مرحله قبل اجرا می‌شود، بردار ویژگی رابطه (۲۲) به دست می‌آید. از این بردار ویژگی اهداف زیر دنبال می‌شود:

۱- اختصاص یک کد به هر تصویر و کاهش ابعاد فضای ویژگی

۲- بردار ویژگی مناسب و کافی و بهینه برای تمییز دادن حروف مختلف

$$F = [A(6)\ H(6)\ V(6)\ D(6)\ H(7)\ D(7)] \qquad (۲۲)$$

مولفه‌های A، H، Vو D به ترتیب ضرایب تقریب تصویر، جزییات افقی، جزییات عمودی و جزییات قطری را در بر داشته و اعداد داخل پرانتز سطح تبدیل موجک گرفته شده را نشان می‌دهد. ضرایب را به صورت ستونی تعریف نموده و از F یک بردار ویژگی ۹۲ بعدی ساخته خواهد شد که در واقع کد مورد نظر برای حالت دست می‌باشد. شکل ۱۱ تصاویر حاصل از کدسازی شکل ۱۰ را نشان می‌دهد.



(الف) (ب) (پ) (ت) (ث) (ج)

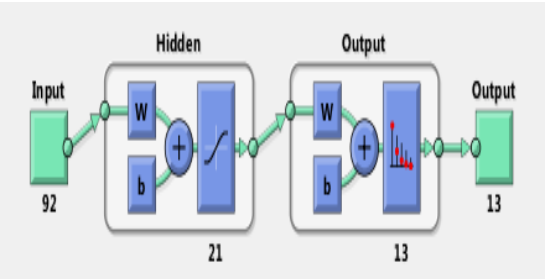
شکل ۱۱) کدسازی شکل ۸ با اعمال تبدیل موجک: الف) تقریب سطح ششم، ب)جزییات افقی سطح ششم، پ)جزییات عمودی سطح ششم، ت)جزییات قطری سطح ششم، ث)جزییات افقی سطح هفتم، ج) جزییات عمودی سطح هفتم

۳-۲- شبکه عصبی برای استخراج ویژگی

در سالهای اخیر، استفاده از دسته‌بندهای مبتنی بر شبکه‌های عصبی برای حل مسائل پیچیده غیرخطی رشد روزافزونی داشته است.پس از کد نمودن هر یک از حالت‌های دست، در این سیستم همانند مرجع [۱۵] از شبکه‌های عصبی برای دسته‌بندی ویژگی‌های به دست آمده از تبدیل موجک استفاده می کنیم. در شبکه عصبی طراحی شده از تابع غیر خطی سیگموئید مطابق رابطه (۲۳) در لایه میانی استفاده شده-است[۲۶].

$$Tansig(n) = \frac{2}{1+e^{-2n}} - 1 \qquad (۲۳)$$

شکل ۱۲ ساختار کلی شبکه عصبی مورد استفاده در این مقاله را نشان میدهد که شامل لایه ورودی با ۹۲ نرون، لایه خروجی با ۱۳ نرون و و لایه مخفی با ۲۱ نرون است.



شکل ۱۲) بلوک دیاگرام شبکه عصبی طراحی شده

به شکل تجربی مشخص شد که شبکه عصبی قادر به تشخیص تمامی حروف الفبا نیست به همین دلیل تنها ۱۳ نرون در لایه خروجی برای تشخیص ۱۳ خانواده متفاوت که در جدول ۱ مشخص شده اند، منظور شده‌است. در مرحله آموزش برای هر خانواده ۲۰ فریم متفاوت و در مجموع ۲۶۰ فریم در نظر گرفته شد. الگوریتم پس انتشار خطا برای آموزش وزنها در شبکه عصبی بکار گرفته شد.

پس از اتمام آموزش از مجوعه داده جدیدی برای تست شبکه عصی استفاده شد.

جدول۱: گروه‌بندی خانواده حرف‌های پیشنهادی توسط شبکه عصبی

خانواده حرف پیشنهادی	حرف پیشنهادی
ا ، آ ، ع ، غ	اَ
ا ، آ ، ی ، ه	آ
ع ، غ ، ن	ع
ع ، ه ، گ ، ن ، ل	گ
اَ، ع ، ه ، غ ، ن	غ
ه ، ن	ه
آ ، ل ، گ ، ی	ل
م ، ع	م
ه ، م ، ن ، ی ، گ	ن
گ ، ل ، اُ	اُ
ل ، ط ، ی ، م	ط
ل ، ی ، گ	ی

۴- توصیف‌گر دایروی

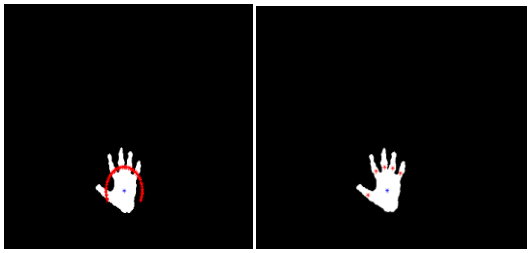
در این قسمت به معرفی توصیف‌گری جدید خواهیم پرداخت که می‌تواند تمایزی قابل توجهی در حالات مختلف دست پدید آورده و اثر غلت را نیز در تصاویر ورودی خنثی سازد. پس از به دست آوردن مرکز و شعاع دست و هم‌چنین از بین بردن چرخش ناخواسته کاربر در سمت و ارتفاع، روشی ارائه می‌شود که می‌تواند تعداد انگشتان باز و بسته را به دست آورد. به این منظور بعد از نرمال‌سازی تصویر ژرفای دست، دایره‌ای از مرکز کف دست ترسیم می‌شود که بتواند انگشتان دست را قطع نماید. شعاع این دایره با بررسی مشخصات هندسی دست بیش از ۵۰ نفر و با استفاده از رابطه ۲۴ بدست می‌آید:

$$R = \begin{cases} C \times N_D \times N_r & , D_{New} < D_{Refrence} \\ \dfrac{C \times N_r}{N_D} & , D_{New} \geq D_{Refrence} \end{cases} \qquad (۲۴)$$

C مقدار ثابتی است که با توجه به آزمایشهای متنوع در دو شرایط خاص بنا به فاصله دست تا دوربین و مقدار N_r به دست می‌آید. دایره توصیف‌گر ویژگی‌های جالب توجهی همچون زوایای انگشتان ایستاده و ضخامت انگشتان برای تشخیص به هم چسبیدگی دو یا چند انگشت را در اختیار قرار می‌دهد. به این ترتیب می‌توان معیاری ارائه کرد که وضعیت انگشتان دست را در هر حالت نشان دهد. شکل ۱۳ دایره رسم شده و محل برخورد آن با انگشتان را نشان داده است.

با در نظر گرفتن مرکز کف دست به عنوان نقطه مرجع می توان برای هر انگشت زاویه‌ای مطابق رابطه ۲۵ ارائه کرد:

$$\theta = tan^{-1}\left(\frac{y_{PALM} - y_{FINGER}}{x_{PALM} - x_{FINGER}}\right) \qquad (۲۵)$$

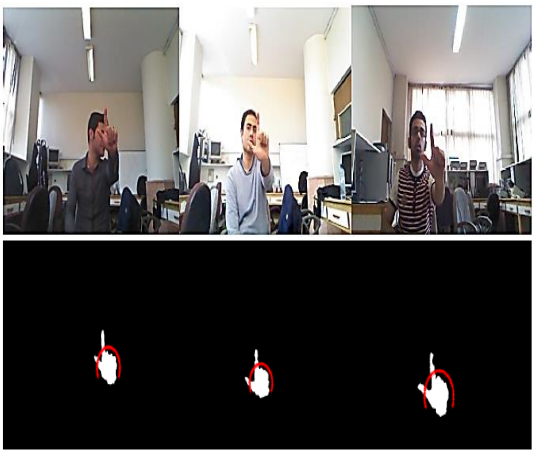


شکل ۱۳) توصیفگر دایروی: نمایش مرکز و نماینده هر انگشت در تصویر جداسازی شده دست

این زوایا هر کدام به عنوان یک ویژگی در ورودی بردار ویژگی حاصل از این قسمت در نظر گرفته خواهد شد. ضخامت هر انگشت نیز با توجه به تعداد نقاط برجسته قرارگرفته شده روی هر قسمت به هم پیوسته از دایره توصیفگر مشخص می‌شود. بدین ترتیب بردار ویژگی

$$X= [T_1\ T_2\ T_3\ T_4\ T_5\ \theta_1\ \theta_2\ \theta_3\ \theta_4\ \theta_5]$$

حاصل از این توصیفگر برای هر دست شامل ۱۰ ویژگی است که به ترتیب پهنا و زاویه هر انگشت را نشان می‌دهد. در صورت عدم وجود هر کدام از انگشت‌ها مقادیر صفر برای زاویه و ضخامت آن در نظر گرفته می‌شود. در شکل ۱۴ سه کاربر مختلف یک حرف را به نمایش گذاشته‌اند. توصیف‌گر دایروی بر روی تصویر ژرفای استخراج شده اعمال شده‌است. همان‌گونه که قابل مشاهده است وضعیت دو انگشت اول به خوبی تمییز داده می‌شود.

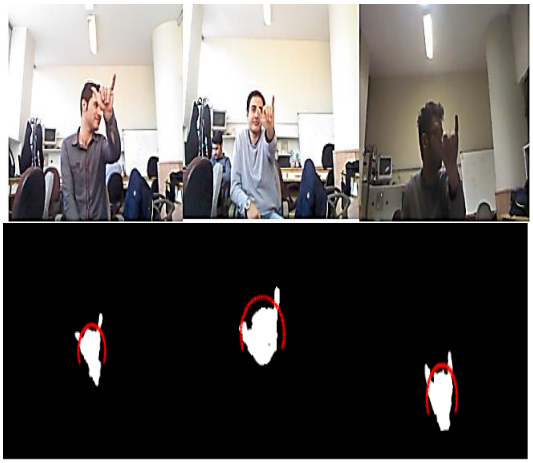


شکل ۱۴) تصویر حرف الف توسط ۳ شخص متفاوت و اعمال توصیفگر دایروی

با انجام آزمایش‌های متعدد، طول شعاع مرجع مقدار ثابت ۳۶ پیکسل و مقدار ژرفای مرجع ۸۶ سانتیمتر به دست آمده است. نکته حائز اهمیت در شعاع توصیفگر مقدار C می‌باشد. به دلیل وابستگی خطای اندازه‌گیری عمق در کینکت به فاصله کاربر تا دوربین، مقدار ثابتی برای این ضریب نمی توان یافت. اگر کاربر دست خود را در فاصله ای بین ۸۰ تا ۱۳۰ سانتیمتر از دستگاه کینکت قرار دهد رابطه ۲۶ مقدار C مناسب را نشان میدهد.

(۲۶)
$$C = \begin{cases} 65, & (N_D < 1\ OR\ N_D > 1.11) AND (N_r \le 0.85) \\ 50 & Otherwise \end{cases}$$

حال پس از تعین مقدار C و مقادیر مرجع، با داشتن شعاع و مینیمم فاصله در هر فریم، ضرایب نرمال‌سازی به دست می آید. سپس دایره ای به شعاع توصیف‌گر و رزولوشن $Res(\theta) = \frac{\pi}{50}$ رسم می نماییم تا ضخامت انگشتان و محل آنها یافت می‌شود. در شکل ۱۵ نتیجه توصیف‌گر پیشنهادی برای هنگامی‌که دست در فواصل به ترتیب از چپ، ۱۰۱ سانتیمتری، ۱۲۷ سانتیمتری و ۸۴ سانتیمتری قرار گرفته است، نشان می‌دهد.



شکل ۱۵) نمایش حرف "ی" توسط ۳ کاربر متفاوت و اعمال توصیفگر دایروی

همانگونه که شرح داده شد شبکه عصبی استفاده شده حروف را به سیزده خانواده مختلف دسته بندی می کند. این خانواده ها در جدول ۱ ارائه شده‌اند. اینک و برای جداسازی کلاس‌های موجود در هر خانواده از یک ماشین بردار پشتیبان تحت یک کرنل خطی بر مبنای حد حاشیه نرم[1] استفاده می‌شود[۲۸، ۲۹]. ایده استفاده از کرنل خطی به جای استفاده از انواع غیرخطی آن، حجم محاسبات پایین آن و مناسب‌تر بودن برای پیاده‌سازی برخط می‌باشد. در مرحله آموزش برای هر کلاس ۳۰ نمونه از داده‌های آموزشی با استفاده از ویژگی‌های آناتومیکی ساخته شده‌است. با اعمال مد آماری، کلاسی که در رقابت با کلاس‌های دیگر بیشترین موفقیت را داشته است به عنوان کلاس برنده انتخاب می‌شود. برای تصمیم‌گیری نهایی خروجی‌های شبکه عصبی و ماشین بردار پشتیبان برای فریم‌های متوالی تصویر ژرفا بدست آمده و یک فیلتر مدین بر بروی آخرین ده نمونه اجرا می‌شود. نتیجه این فیلتر خروجی اصلی خواهد بود. به این ترتیب اگر در بیش از شش فریم خروجی بردارهای پشتیبان به یک حرف اشاره نمایند حرف موردنظر نمایش داده می‌شود. هدف از این عملیات، حذف نویز از خروجی حاصل از نمایش یک فیلم و قابل اعتماد نمودن الگوریتم می‌باشد.

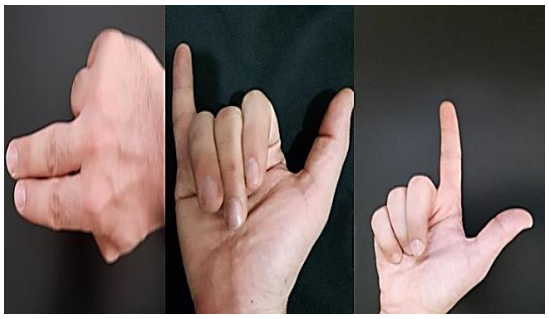
۵- نتایج شبیه‌سازی

در این بخش نتایج حاصل از پیاده سازی الگوریتم پیشنهادی نشان داده می‌شود. در شکل ۱۷ نمایش سه حرف آ، طا و اَ در رابط گرافیکی کاربر نوشته شده در نرم‌افزار متلب را مشاهده می‌کنید.

شکل ۱۷) نمایش چهار حرف از الفبای ناشنوایان تحت سیستم پیشنهادی

برای مقایسه روش پیشنهادی، عملکرد الگوریتم مرجع [۱۵] که در زمان نگارش این مقاله معتبرترین روش موجود در زمینه بازشناسی الفبای فارسی ناشنوایان می‌باشد، استفاده شده‌است. این مرجع از اطلاعات تصویر رنگی استفاده کرده و این تصاویر در شرایط ایده آل و با پس‌زمینه کاملا ساده و سیاه رنگ تهیه شده‌اند. معیار مقایسه انتخابی نیز میزان موفقیت‌آمیز بودن تشخیص حروف است که با رابطه (۲۷) تعریف می‌شود.

(۲۷)
$$\text{نرخ بازشناسی} = \frac{\text{تعدادشناسایی درست}}{\text{تعداد کل آزمونها}} \times ۱۰۰$$



شکل ۱۸) بخی از مجموعه داده ساخته شده برای مقایسه با مرجع [۱۵]

برای ایجاد پایگاه داده مورد نیاز جهت پیاده سازی و مقایسه با مرجع [۱۵] در مجموع ۴۲۳ فریم تهیه شد. شکل ۱۸ نمونه‌ای از این مجموعه داده را نشان می‌دهد. یک دوربین گالکسی A8 با رزولوشن ۱۶

[1] Soft Margin

مگاپیکسل برای تصویر برداری استفاده شده‌است. دو کاربر متفاوت الگوهای مورد نیاز را با قرار دادن دست خود در فاصله ۵۰ سانتی متری از دوربین و در مقابل یک صفحه سیاه رنگ تامین کرده‌اند. از این میان ۲۷۶ الگو برای آموزش شبکه عصبی استفاده شده‌است. جدول ۲ نتایج پیاده سازی روش پیشنهادی و روش ارائه شده در مرجع [۱۵] را برای تشخیص تک تک حروف الفبای فارسی ناشنوایان نشان میدهد.

جدول ۲: مقایسه میزان موفقیت تشخیص حروف در روش پیشنهادی و روش [۱۵]

حرف	دقت در [۱۵]	دقت در سیستم پیشنهادی
أ	٪۸۱/۸۱	٪۹۶
آ	٪۸۱/۸۱	٪۹۸/۷۱
ع	٪۹۰/۹	٪۸۹/۲۸
گ	٪۹۵/۴۵	٪۹۶
غ	٪۱۰۰	٪۱۰۰
ه	٪۹۰/۹	٪۹۵/۸۵
ل	٪۱۰۰	٪۹۴
م	٪۹۰/۹	٪۱۰۰
ن	٪۹۰/۹	٪۹۱/۵
اُ	٪۸۱/۸۱	٪۱۰۰
ط	٪۱۰۰	٪۱۰۰
ی	٪۹۰/۹	٪۱۰۰

همانطور که مشاهده می‌کنید نرخ بازشناسی در سیستم پیشنهادی رشد و بهبود یافته است و نرخ بازشناسی کل سیستم پیشنهادی برابر ٪ ۹۶/۷ می باشد.

الگوریتم پیشنهادی توسط نرم افزار Matlab و بر روی یک کامپیوتر با پردازنده ۲/۳ گیگاهرتز و ۴ گیگابایت حافظه اجرا شده‌است. به شکل متوسط برای هر فریم از تصویر ژرفا ۲۵۰ میلی‌ثانیه زمان لازم است. برای حذف نویز در خروجی از یک فیلتر مدین بر روی آخرین هشت نمونه استفاده می شود. به این ترتیب خروجی نهایی با حدود دو ثانیه تاخیر قابل مشاهده است.

۶- نتیجه‌گیری

روش ارائه شده در این تحقیق با کمک دستگاه کینکت این امکان را فراهم می‌سازد که به شکل عملی و بدون در نظر گرفتن شرایط خاص برای محیط تصویر برداری و یا پوشیدن دستکش‌های مخصوص، بتوان الفبای ناشنوایان فارسی را تشخیص داد. تشخیص برخی از این علایم با توجه به اینکه تنها با جابجایی مختصر انگشتان در محور طولی متمایز شده‌اند، با استفاده از تصاویر عادی بسیار سخت است. در روش پیشنهادی استفاده از تصاویر ژرفا امکان این تمایز فراهم شده‌است. اجرای الگوریتم در این فضا امکان بی‌اثر کردن دوران دست کاربر در راستای محورهای مختصات را فراهم ساخته و به این ترتیب دقت تشخیص را تا ٪۹۶/۷ افزایش می‌دهد. مطمئنا در صورت پیاده‌سازی الگوریتم در محیطی غیر از متلب می‌توان سرعت پردازش را از چهار فریم در ثانیه به مقادیر بیشتر نیز افزایش داد.

۷- مراجع

[1] Moghadam M., Nahvi, M. Hassanzadeh, 2011, "Static Persian Sign Language Recognition Using Kernel-Based Feature Extraction", 7th Iranian IEEE Machine Vision and Image Processing (MVIP).

[2] Wang, L. C., Wang, R., Kong, D., Yin, B., 2014, "Similarity Assessment Model for Chinese Sign Language Videos", IEEE Transaction on Multimedia, 16, pp.751 – 761.

[3] Huang, J., Zhou, W., Li, H., Li, W., 2015, "Sign language recognition using 3d 1430 convolutional neural networks", Multimedia and Expo (ICME), IEEE International Conference on IEEE, pp.1–6.

[4] Agris, U, V., Zieren, J., Canzler, U., Bauer, B., Kraiss, K. 2008, "Recent developments in visual sign language recognition", Universal Access in the Information Society, 6(4), pp.323-362.

[5] Shah, N, K., Rathod, R, K., Agravat, J, S. 2014, "A survey on Human Computer Interaction Mechanism Using Finger Tracking", International Journal of Computer Trends and Technology (IJCTT). 7(3), pp. 174-177.

[6] Verma, H. V., Aggarwal, E., Chandra, S., 2013, "Gesture recognition using kinect for sign language translation", IEEE Second International Conference on IEEE Image Information Processing (ICIIP).

[7] Yang, H.D., 2014, "Sign language recognition with the kinect sensor based on conditional random fields", Sensors 15(1), pp.135-147.

[8] Sun, C., Zhang, T., Bao, B., Xu, C., Mei, T., 2013, "Discriminative Exemplar Coding for Sign Language Recognition with Kinect", IEEE Transaction on Cybernetics, 43, pp.1418 – 1428.

[9] Li, S.Z., Yu, B., Wu, W., Su, S.Z., Ji, R.R., 2015, "Feature learning based on SAE–PCA network for human gesture recognition in rgbd images", Neurocomputing, 151, pp.565–573.

[10] Liu, T., Zhou, W., Li, H., 2016, "Sign language recognition with long short-term memory", IEEE International Conference on Image Processing (ICIP), pp.2871–2875.

[11] Almeida, S.G.M., Guimar˜aes, F.G., Ram´ırez, J.A., 2014, "Feature extraction in Brazilian sign language recognition based on phonological structure and using rgb-d sensors", Expert Systems with Applications, 41, pp.7259–7271.

[12] Lim, M.K., Tan, W.C. A., Tan, C.S., 2016, "A feature covariance matrix with serial particle filter for isolated sign language recognition", Expert Systems with Applications, 54, pp.208-218.

the backpropagation method", Biological Cybernetics, 59, pp.257–263.

[27] M. T., Demuth, H. B., Beale, M. H., De Jesús O., Neural network design, 2nd edition, Hagan, (2014),

[28] Duda, R. O., Hart, P. E., Stork, D. G., 1973, "Pattern classification", Journal of Classification, 24, pp.305–307.

[29] Fukunaga, Keinosuke, Introduction to statistical pattern recognition, Elsevier, 2013.

[13] Zhao, R., Martinez, M. A., 2016, "Labeled Graph Kernel for Behavior Analysis", IEEE Transaction on Pattern Analysis and Machine Intelligence, 38, pp.1640 – 1650.

[14] Elakkiya, R., Selvamani, K. 2017, "Enhanced dynamic programming approach for subunit modelling to handle segmentation and recognition ambiguities in sign language", Journal of Parallel and Distributed Computing.

[15] Karami, A., Zanj B., Kiani A. 2011, "Persian sign language (PSL) recognition using wavelet transform and neural networks", Expert Systems with Applications, 38(3), pp.2661-2667.

[16] Wang, C., Liu, Z., Chan S., 2015, "Superpixel-Based Hand Gesture Recognition with Kinect Depth Camera", IEEE Transaction on Multimedia, 17, pp. 29 – 39.

[17] Neiva D. H., Zanchettin C. 2018, "Gesture Recognition: a Review Focusing on Sign Language in a Mobile Context", Expert Systems with Applications.

[18] Albrecht, I., Haber, J., Seidel, H. 2003, "Construction and animation of anatomically based human hand models", Proceedings of ACM SIGGRAPH / Eurographics Symposium on Computer Animation SCA '03, pp.98-109.

[19] Caillette, F., Galata, A., Howard, T., 2008, "Real-time 3-d human body tracking using learnt models of behavior", CVIU, 109(2), pp 112-125.

[20] Francke, H., Solar J, R., Verschae, R., 2007, "Real time hand gesture detection and recognition using boosted classifiers and active learning", Advances in Image and Video Technology, 4872, pp.533-547.

[21] Krotosky, S., Trivedi, M., 2006, "Registration of Multimodal Stereo Images Using Disparity Voting from Correspondence Windows", IEEE International Conference on Video and Signal Based Surveillance. pp. 91-91.

[22] Markelj, P., Tomaževič, D., Likar, B., Pernuša, F. 2012, "A review of 3D/2D registration methods for image-guided interventions", Medical image analysis, 16(3), pp.642-661.

[23] Pizzoli, M., Forster, C., Scaramuzza, D. 2014, "REMODE: Probabilistic, monocular dense reconstruction in real time", IEEE International Conference on Robotics and Automation (ICRA).

[24] Craig, J, J. 2009, Introduction to Robotics: Mechanics and Control (3rd Edition).

[25] Van den Bergh, M., Van Gool L., 2011, "Combining RGB and ToF cameras for real-time 3D hand gesture interaction", IEEE Workshop on Applications of Computer Vision (WACV).

[26] Vogl, T. P., Mangis, J.K., Rigler, A.K., Zink, W.T., Alkon, D.L. 1988, "Accelerating the convergence of