بازشناسی حروف الفبای زبان اشاره فارسی ناشنوایان با استفاده از خوشه‌بندی kmeans و دنباله ای از تصاویر اشاره دست در پس زمینه پویا

زهرا حسن شاهی[۱]، دکتر احمدرضا نقش نیلچی[۲]

[۱]دانشگاه اصفهان ، parasto_1610@yahoo.com

[۲]دانشگاه اصفهان ، nilchi@eng.ui.ac.ir

چکیده

در این مقاله، یک سامانه بینایی ماشین برای بازشناسی حروف الفبای زبان اشاره فارسی ناشنوایان با استفاده از دنباله ای از تصاویر اشاره دست (به عنوان یک ابزار ورود اطلاعات) ارائه می‌شود.لازم است علامت‌های الفبای زبان اشاره فارسی در پس زمینه پویا را به زبان طبیعی ترجمه کند. در روش ارائه شده، در مرحله‌ی استخراج ویژگی، ابتدا با استفاده از روش خوشه‌بندی kmeans حروف الفبای ناشنوایان را به دو، سه یا چهار خوشه تبدیل می‌شود، به نحوی که در هر خوشه حروفی قرار بگیرند که بیشترین شباهت را با یکدیگر داشته باشند. سپس برای بازشناسی علامت‌ها با قابلیت اطمینان بالا، یک سیستم ترکیبی از چهار کلاس‌بند مختلف ماشین بردار پشتیبان و پرسپترون چند لایه و درخت‌های تصمیم و طبقه‌بندی بیزین برای هر نود این درخت پیشنهاد شده است، بطوری که در هر نود از یک دسته‌بند که بیشترین صحت بازشناسی برای آن نود را دارد، بکار رود. در این پژوهش، از یک مجموعه متشکل از ۱۸۰۰ تصاویر از ۳۶ حالت مختلف دست که از حروف الفبای زبان اشاره فارسی هستند، استفاده شده است. نتایج آزمون سیستم ارائه شده، نشان می‌دهد که در شرایط مختلف (دارای منبع یا بدون منبع روشنائی، نزدیک یا دور بودن رنگ دست با پس زمینه)، متوسط میزان تشخیص درست حالت‌های زبان اشاره ۹۹,۵۹٪ می باشد. همچنین، این سیستم می تواند با خطای کمتر از ۰,۴ ٪ ، حالات دست را به درستی تشخیص می‌دهد. نرخ بازشناسی و قابلیت اطمینان سیستم بر روی مجموعه آموزشی پایگاه داده PSL به ترتیب ۸۵,۸۳٪ و ۹۹,۲۱٪ است که عملکرد سیستم رضایت بخش و مطلوب می‌باشد.

کلید واژه

آشکارسازی دست، تشخیص حالت‌های دست، پردازش تصویر، بازشناسی علامت‌های زبان اشاره ، پس زمینه پویا

۱- مقدمه

زبان اشاره[۱]، یک زبان کامل و رسا است که مهم‌ترین روش ایجاد ارتباط بین افرادی است که از نقص شنوایی رنج می‌برند[۱۰]. به منظور راحت‌تر شدن ارتباط بین افراد ناشنوا و شنوا و یک‌پارچه شدن جامعه، می‌توان زبان اشاره را به عنوان زیر مجموعه ای از حالت‌ها بررسی و به طور خودکار به متن یا صحبت ترجمه کرد. علامت‌های این زبان را می‌توان به دو دسته حالت‌های ساکن و حالت‌های متحرک تقسیم کرد. حالت‌های ساکن در علامت‌ها، فقط با شکل دست معرفی می‌شوند. در صورتی که در تشخیص حالت‌های متحرک علاوه بر شکل دست، حرکت دست نیز تأثیرگذار است.

زبان اشاره فارسی، تقریبا شامل ۱۰۷۵ حالت ساکن و متحرک برای حروف الفبا و کلمات متداول زبان فارسی است. الفبای زبان اشاره فارسی، شامل ۳۷ علامت است که با یک دست ادا می‌شوند. از بین این علامت‌ها، ۲ حالت متحرک و ۳۵ حالت ساکن داریم.

پژوهش‌ها بر روی بازشناسی خودکار علامت‌های زبان اشاره از سال ۱۹۹۰ شروع شد. سیستم‌های موجود به دو دسته تقسیم می‌شوند. دسته‌ی اول، علامت دهنده‌ها لازم به استفاده از دستکش‌های خاصی

[1] Sign language

می‌شوند و با استفاده از اطلاعات بدست آمده از دستکش‌ها، سیستم به بازشناسی علامت‌ها می‌پردازد. روش‌های مبتنی بر دستکش، که مجهز به حسگرهایی برای اندازه‌گیری زوایای مفاصل دست هستند. حسگرهای تعبیه شده روی آن‌ها، زاویه‌های بین انگشتان را به عنوان معیاری برای مشخص کردن جهت دست و شکل قرارگیری آن محاسبه می‌کند. به عنوان مثال در [۱] ، برای تشخیص علامت‌های زبان اشاره از دستکش الکترونیکی استفاده شده است. این دستکش قادر است زوایای مفاصل را اندازه گرفته و داده‌های بدست آمده را به یک شبکه عصبی آموزش دهد. در [۲] Kadous یک سیستم تشخیص زبان اشاره استرالیایی را بر اساس دستکش‌های الکترونیکی پیشنهاد کرد. او این سیستم را بر روی یک مجموعه ۹۵تایی از زبان اشاره به کار گرفت. در [۳]، نیز از دستکش‌های مخصوصی که حالت‌های دست را مشخص می‌کنند، استفاده شده است.

دسته‌ی دوم شامل سیستم‌هایی هستند که مبتنی بر روش‌های بینایی ماشین و پردازش تصویر کار می‌کنند. در سیستم طراحی شده در [۴]، از تشخیص رنگ پوست ، با استفاده از تبدیل تصویر RGB به تصویر YCBCR است. در مرحله آموزش از روش ثابت لحظه و بردار ویژگی استفاده شده است که این بردار ویژگی به عنوان ورودی برای طبقه بندی چندکلاس شبکه‌های عصبی مصنوعی است. در مرحله تست، از وب کم برای گرفتن تصویر و چهار مقدار ویژگی برای طبقه بندی و شناخت حرکات دست استفاده می شود. همچنین در [۵]، روشی برای تشخیص وضعیت دست زبان اشاره تایلندی در یک پس زمینه پیچیده با استفاده از تلفیقی از عمق و تصویر رنگی است. این فناوری جدید از حسگرها، مانند مایکروسافت کینکت، برای پیدا کردن موقعیت دست در صحنه با کمک عمق ویدئویی فراهم می کند. برای استخراج ویژگی-های تصویر علامت دست نمودار هیستوگرام گرادیان مورد استفاده قرار می‌گیرد. این ویژگی ها به شبکه عصبی مصنوعی برای آموزش و شناخت منتقل می‌شود.

از کدهای زنجیره ای، ماتریس شکل، تبدیل فوریه ، لحظات Hu7 نیز به عنوان ویژگی برای تشخیص حرکت دست استاتیک وپویا زبان اشاره هند و سه روش طبقه بندی نزدیک‌ترین متوسط طبقه بندی(NMC-K)، نزدیک‌ترین همسایگی (K-NN) طبقه بندی بیز استفاده شده است[۶]. همچنین برای تشخیص علامت‌های ساکن زبان اشاره فارسی [۷] استفاده شده است. در [۷] ابتدا تصاویر علامت‌ها به تصاویر سطح خاکستری تبدیل شده و سپس ناحیه دست از آن تصاویر استخراج می‌گرد. تبدیل ویولت روی نواحی دست اعمال و ویژگی‌های حاصل از آن به شبکه‌های عصبی آموزش داده می‌شود. علاوه بر شبکه‌های عصبی از مدل پنهان مارکوف[1] نیز به عنوان یکی از روش‌های بازشناسی الگو استفاده شده است. خیلی از محققین برای تشخیص ساختار عکس دست، از تقسیم بندی دست استفاده کرده‌اند [۱۲]. روش ارائه شده در فضای رنگ YCBCR ، مدل گاوسی تک و قاعده بیز است. ویژگی‌های ژست دست بر اساس فاصله شعاعی و تبدیل فوریه و در نهایت، فاصله اقلیدسی برای محاسبه شباهت بین علایم ورودی و بردارهای ویژگی استفاده می‌شود. Raboy و همکارانش با ایجاد یک برنامه تلفن همراه که حالت‌های دست را تشخیص و به متن تبدیل می‌کند. این نرم افزار با استفاده از فرایند تقسیم بندی دست و استخراج ویژگی برای دریافت اطلاعات مورد نیاز برای استخراج دست و مجموعه داده ها است [۱۳]. در[۱۴]، استخراج ویژگی علایم اشاره با استفاده از مدل رنگ HSV، ویژگی‌های به دست آمده با الگوریتم آنالیز اجزای اصلی (PCA) مقایسه شده است. پس از آن تست حداقل فاصله اقلیدسی برای به تشخیص علایم محاسبه می‌شود. در نهایت، ژست شناخته شده به متن و صدا تبدیل می‌شود. Kim از فیلتر کالمن برای پیدا کردن نوک انگشتان استفاده کرده است [۱۵]. وانگ[2]از گرادیان جهت‌دار حرکتی[3]برای استخراج خصوصیات حرکتی و اسپارس بیزین[4]برای کلاس‌بندی این خصوصیات به کلاس مربوطه استفاده کرده است [۱۶]. در زمینه زبان علامت فارسی کار کرمی در [۷] بیان شده است. در سیستمی که ایشان طراحی کردند از ۳۲ حروف ثابت زبان

[1] Hidden Marcov Model (HMM)

[2] Wong

[3] Motion Gradient Orientation (MGO)

[4] Sparse Bayesion

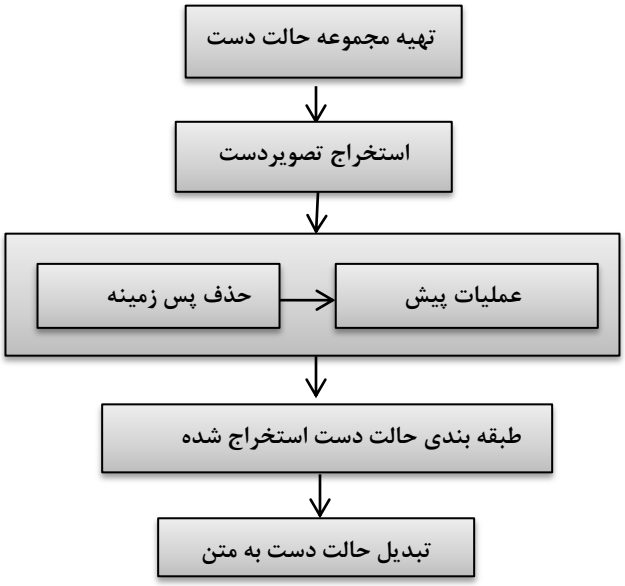
اشاره استفاده کرده اند. ایشان از تبدیل ویولت[1] برای استخراج ویژگی و از شبکه عصبی برای آموزش آن بهره برده‌اند. کار آقای مقدم نیز در [۸] آمده است ایشان نیز به مقایسه اثر استخراج ویژگی‌های KPCA[2] و KDA[3] بر نتیجه خروجی پرداخته‌اند و از متدهای K نزدیک‌ترین همسایه[4] و ماشین بردار پشتیبان[5] و شبکه‌های عصبی[6] بهره برده‌اند. Barkoky در [۹] به تشخیص اعداد ناشنوایان به صورت ثابت پرداخته است. او از متد نازک‌سازی برای مشخص کردن قسمت اصلی دست استفاده کرده است. کار کیانی در [۱۰] هم مشابه کار کرمی در [۷] می‌باشد با این تفاوت به جای ۳۲ حرف از ۸ کلمه که با علامت ثابت نشان داده شده است استفاده کرده است. ژیلا موسویان یک سیستم برای بازشناسی حروف الفبای زبان اشاره فارسی ارائه کردند. این سیستم بر مبنای هیستوگرام گرادیان‌های جهت‌دار و شبکه‌های عصبی طراحی شده است [۱۱].

هدف از این مقاله، طراحی یک سیستم برای بازشناسی علامت‌های حروف الفبای زبان اشاره فارسی می‌باشد. در این سیستم، ما با در نظر گرفتن ۳۶ علامت الفبای زبان اشاره‌ی فارسی و استفاده از روش‌های پردازش تصویری به بازشناسی علامت‌ها می‌پردازیم. در این مقاله، ابتدا سامانه باید بتواند نواحی مربوط به دست را تشخیص داده و حالت‌های دست را شناسایی کند. سپس برای طبقه بندی علامت‌ها، با استفاده از روش خوشه‌بندی kmeans یک سیستم ترکیبی از چهار کلاس‌بند مختلف ماشین بردار پشتیبان و پرسپترون چند لایه و درخت‌های تصمیم و طبقه‌بندی بیزین برای هر نود این درخت پیشنهاد می‌کنیم. بطوری که در هر نود می‌توان از یک دسته‌بندی که دقت بازشناسی بالاتری برای آن نود دارد استفاده کنیم. ادامه این مقاله به ترتیب زیر نوشته شده است: در بخش ۲ به معرفی سیستم پیشنهادی می پردازیم .نتایج بازشناسی سیستم در بخش ۳ ارائه شده است .در نهایت جمع بندی و نتیجه گیری مطالب در بخش ۴ آورده شده است.

[1] Wavelet Transform
[2] Kernel Principle Component Analysis
[3] Kernel Discriminant Analysis
[4] K-nearest neighbors
[5] Support Vector Machine
[6] Artificial Neural Network

۲- سیستم پیشنهادی

سیستم پیشنهادی برای بازشناسی علامت‌های الفبای زبان اشاره‌ی فارسی طراحی شده است. در این مقاله سعی ما بر آن است یک سیستم بازشناسی جدید طراحی نمائیم. این سیستم دارای سه مرحله-ی پیش پردازش، استخراج ویژگی و بازشناسی علامت‌ها مبتنی بر بردارهای ویژگی بدست آمده می‌باشد. در ادامه هر کدام از مراحل سیستم پیشنهادی به ترتیب بررسی خواهند شد. در شکل زیر روند اجرای سیستم تشخیص حالت دست و تبدیل به متن مشخص شده است.



شکل۱ :روند اجرای سیستم تشخیص حالت دست و تبدیل به متن

۲-۱- تهیه مجموعه حالات دست

مجموعه داده متشکل از افراد مختلف با پشت زمینه پویا فراهم شده است. برای انجام آزمایش‌ها ما از پایگاه داده[7] ژست‌های دست [۴] و [۵] تصاویری که تهیه شده استفاده نموده‌ایم. تصاویر این پایگاه داده شامل ۳۶ علامت ساکن زبان اشاره فارسی می‌باشد که با یک دست بدون هیچ‌گونه پوششی در مقابل یک پس زمینه پیچیده گرفته شده-اند. برای هر علامت، ۵۰ تصویر وجود دارد که با فاصله‌های متفاوت از

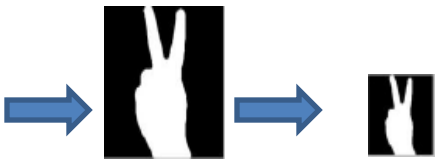
[7] Data Base

دوربین و با علامت دهنده‌های متفاوت تهیه شده‌اند. اندازه همه تصاویر ۲۰۰*۲۰۰ پیکسل است.

۲-۲- استخراج تصویر دست از پس زمینه پویا

قدم بعدی برای تشخیص حالت دست این است که بتوان خود دست را با استفاده از روش تشخیص رنگ پوست بدون نویز از پس زمینه آن جدا کرد تا بتوان علامت نشان داده آن‌را جدا از پس زمینه‌ای که وجود دارد طبقه‌بندی نمود. این مرحله را به سه مرحله اصلی حذف پس‌زمینه در فضای YC_BC_R و عملیات پیش‌پردازش و استاندارد سازی تصاویر تقسیم می‌شود. در مرحله حذف پس‌زمینه تصویر عکس از پس‌زمینه آن جدا شده و در مرحله عملیات پیش‌پردازش عکس جدا شده از پس‌زمینه، چنان آماده می‌شود تا بتوان خصوصیات آن‌را استخراج کرد. برای محاسبه‌ی مؤلفه‌های اصلی شکل دست لازم است ناحیه دست از پس زمینه جدا گردد. برای این منظور، ابتدا تصاویر اصلی را به تصاویر دودویی تبدیل کرده و با اعمال فیلترهای مورفولوژی، شکل دست را در تصاویر پر می‌کنیم. برای رسیدن به یک استاندارد یگانه برای همه تصاویر می‌بایست به گونه‌ای پیکسل‌های پس زمینه را حذف کرد. پس از اینکه عملیات مورفولوژیک روی تصویر انجام شد، تصویری باینری از حالت دست که دارای پس زمینه سیاه رنگ است تهیه می‌شود. با یافتن جعبه مرزی[1] تصویر، طول و عرض مجموعه پیکسل‌های تشکیل دهنده حالت دست مشخص می‌شود. با در نظر گرفتن ماکزیمم طول و عرض، تصویر دست در مرکز یک قاب مربعی شکل که طول و عرض آن برابر با ماکزیمم طول و عرض جعبه مرزی است، قرار داده می شود. بعد از اینکه تمام کانتورها در محدوده رنگی پوست دست مشخص شد، باید کانتور مربوط به دست را از بقیه کانتورها جدا کنیم. در اینجا ما کانتوری که نسبت به بقیه کانتورها مساحت بزرگتری دارد را پیدا کرده و آن را به عنوان دست انتخاب می‌کنیم تا سایر عملیات تشخیص ژست را روی آن انجام دهیم. برای این کار، هنگامی که در برنامه به دنبال پیدا کردن کانتورها هستیم، برای هر کانتور، مساحت آن را با دیگر کانتورهای آشکار شده مقایسه می‌کنیم. سپس شماره مربوط به کانتور بزرگ‌تر را ذخیره کرده تا سایر عملیات پردازش تصویر را بر روی آن اعمال کنیم. در شکل۲ نحوه انجام این مراحل نمایش داده شده است. سپس در ادامه، منحنی پیرامونی دست را با استفاده از لبه یاب الگوریتم فیلترگابور بدست می‌آوریم.

تصویر اصلی تصویر آستانه گیری شده و استفاده از عملگرهای مورفولوژی تصویر در جعبه مرزی



تصویر نرمال شده تصویر نهایی با اندازه استاندارد

شکل۲ :مراحل استانداردسازی تصاویر حالت دست

پارامترهای تشخیص کارایی استخراج دست از پس‌زمینه

در اینجا پنج پارامتر اصلی که به ما در دقت نتیجه به‌دست آمده کمک می‌کند عبارتند از:

۱. TP معادل تعداد پیکسل‌هایی از دست استخراج شده که با دست واقعی در عکس اصلی مطابقت دارد.

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \qquad (۱)$$

۲. FP معادل تعداد پیکسل‌هایی از دست استخراج شده که معادل آن در دست واقعی، واقع در عکس اصلی مطابقت ندارد.

[1] Bounding Box

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - SPC \quad (۲)$$

۳. TN تعداد پیکسل‌هایی از پس‌زمینه که به‌درستی تشخیص داده شده‌اند.

$$SPC = \frac{TN}{N} = \frac{TN}{FP + TN} \quad (۳)$$

۴. FN تعداد پیکسل‌هایی که به اشتباه به عنوان پس‌زمینه شناخته شده‌اند.

$$FNR = \frac{FN}{P} = \frac{FN}{FN + TP} \quad (۴)$$

۵. قابلیت اطمینان[1] ، میزان اعتبار تصمیم نهایی کلاس‌بند را برای هر کلاس نشان می‌دهد و طبق رابطه زیر تعریف می‌شود.

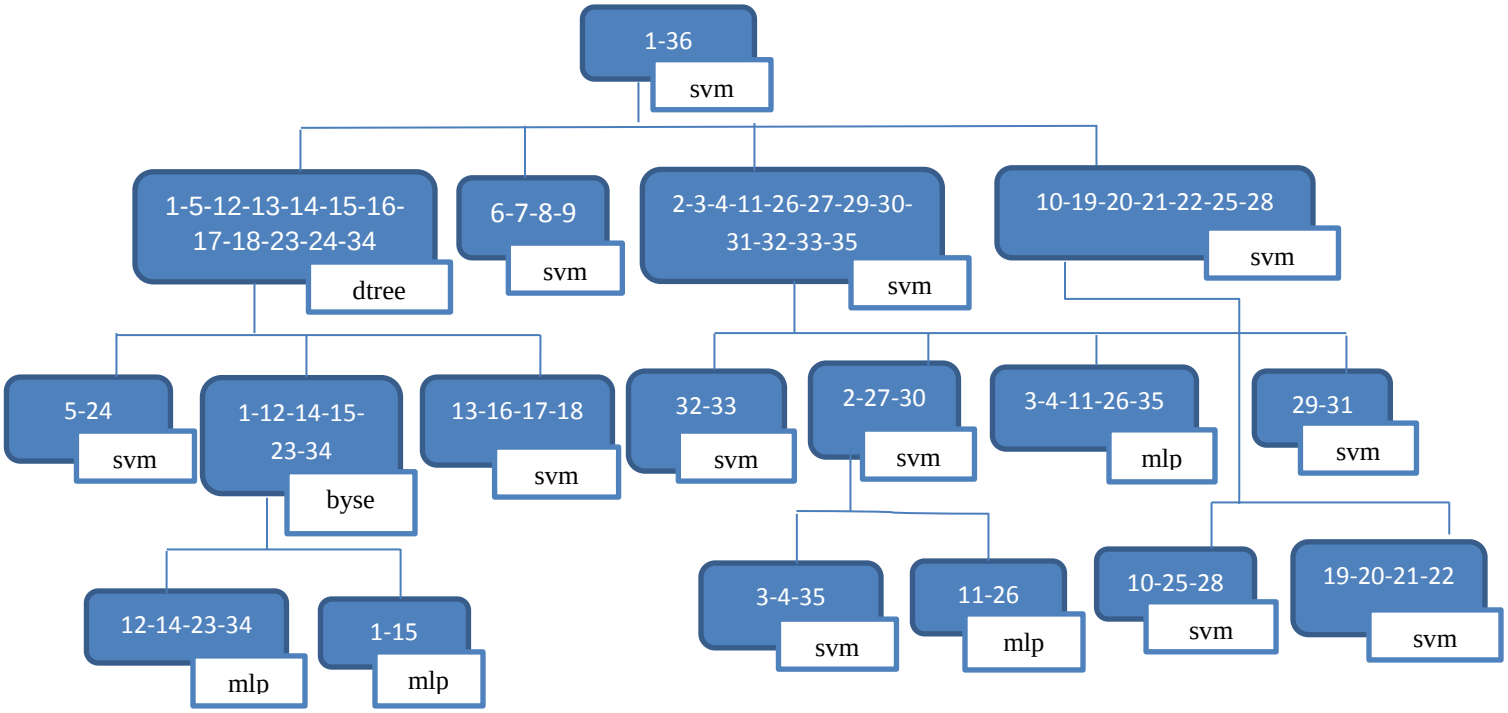
$$ACC = \frac{TP + TN}{P + N} \quad (۵)$$

۲-۳- بازشناسی علامت‌ها

پس از مرحله استخراج ویژگی از تصاویر آماده شده در مرحله پیش پردازش، در مرحله آخر برای کلاس‌بندی علامت‌ها از ویژگی‌های استخراج شده، ازچهار کلاس‌بند ماشین بردار پشتیبان، پرسپترون چند لایه، بیزین، درخت تصمیم استفاده می‌کنیم. در مرحله بازشناسی علامت‌ها، یک سیستم ترکیبی نیز با استفاده از نتایج بازشناسی چهار کلاس‌بند فوق طراحی می‌کنیم. قانون بازشناسی در سیستم جدید به این صورت است: در بازشناسی زبان اشاره، حروف شبیه به هم گاهاً اشتباه بازشناسی می‌شود. یعنی مثلاً حروف (د و ذ) که زبان اشاره آن شبیه به هم است به جای یکدیگر به اشتباه شناسایی می‌شود. برای افزایش دقت بازشناسی می‌توان در مرحله اول طی یک روند مشخص تعیین کرد کدام حروف ممکن است اشتباهاً به جای یکدیگر شناسایی شوند. به عنوان مثال حروف(ح، خ، د، ذ، ر، ز، ژ) خیلی از مواقع به جای یکدیگر شناسایی می‌شوند. با استفاده از الگوریتم خوشه‌بندی kmeans این حروف را به دو، سه یا چهار خوشه تبدیل کنیم به نحوی که در هر خوشه حروفی قرار بگیرند که بیشترین شباهت را با یکدیگر دارند. اگر خوشه‌ای تعداد کلاس‌های آن بیشتر از چهار بود این عملیات را برای آن تکرار می‌کنیم، به این ترتیب یک درخت به وجود می‌آید که نودهای انتهایی آن چهار یا کمتر از چهار برگ دارد سپس برای هر نود از این درخت یک دسته‌بند دارای دقت و صحت بالاتر استفاده می‌کنیم و تعداد دسته‌بندها با تعداد نودهای درخت برابر است. با اعمال بردارهای ویژگی بدست آمده به هر یک از کلاس‌بندها و سیستم ترکیبی، نتایج بازشناسی را در همه‌ی حالت‌ها بدست آورده و عملکرد الگوریتم پیشنهادی را در هر یک از حالت‌ها ارزیابی می‌شود. در شکل۳ در صورتی که تمام نودهای درخت از کلاس‌بندی svm استفاده کنند، نتیجه آزمایش در سیستم جدید چندان تغییری نمی‌کند. ولی کلاس‌بند svm بدون استفاده از الگوریتم خوشه‌بندی kmeans به نحوی که در بازشناسی زبان اشاره، حروف شبیه به هم اشتباه بازشناسی می‌شود، با توجه به شکل۴ نتیجه خوبی ندارد.

همانطور که درشکل۴ مشخص است بدترین حالت ممکن در این مساله سیستم پرسپترون چند لایه با متوسط دقت سیستم ۹۵,۸۱٪، صحت سیستم ۶۰,۵۴٪، نرخ بازشناسی صحیح ۹۷,۸۴٪، نرخ بازشناسی اشتباه ۲,۱۵ ٪ می‌باشد. بهترین حالت استفاده از سیستم پیشنهادی با متوسط دقت سیستم، صحت سیستم، نرخ بازشناسی صحیح، نرخ بازشناسی اشتباه حالات زبان اشاره به ترتیب ۹۹,۲۱٪ و ۸۷,۵۸٪ و ۹۹,۵۹٪ و ۰,۴٪ می‌باشد.

[1] Reliability



شکل ۳:بازشناسی زبان اشاره فارسی با kmeans

۲-۳-۱ روش خوشه‌بندی K-means

در این روش ابتدا به تعداد خوشه‌های مورد نیاز، نقاطی به صورت تصادفی انتخاب می‌شود. سپس در داده‌ها با توجه به میزان نزدیکی(شباهت) به یکی از این خوشه‌ها نسبت داده می‌شوند و بدین ترتیب خوشه‌های جدیدی حاصل می‌شود. با تکرار همین روال می‌توان در هر تکرار با میانگین‌گیری از داده‌ها مراکز جدیدی برای آنها محاسبه کرد و مجدداً داده‌ها را به خوشه‌های جدید نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر تغییری در داده‌ها حاصل نشود [۱۷] . تابع زیر به عنوان تابع هدف مطرح است.

$$j = \sum_{j=1}^{k} \sum_{i=1}^{x} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (۶)$$

که $\| \ \|$ معیار فاصله بین نقاط و c_j مرکز خوشه j ام است.

۲-۳-۲ روش ماشین بردار پشتیبان[1]

در این روش مجموعه عکس‌هایی از حروف ناشنوایان داریم که عکس دست از پس زمینه جدا شده و ویژگی‌های آن استخراج گردیده است. حال صورت مسئله را با جدا کردن نمونه‌های کلاس‌های مختلف در مجموعه آموزشی D تعریف می‌کنیم. در این وضعیت D مجموعه آموزشی که شامل n نمونه Xi می‌باشد. مثال را با دو کلاس پیش برده و در قسمت بعدی SVM دو کلاسی به چند کلاس تعمیم داده می‌شود. زیرا اساس SVM بر پایه دو کلاس می‌باشد. در این صورت Xi را برچسب‌های $ci \in \{-1,1\}_{i=1}^{n}$ از دو کلاس مثلاً حرف "د" از حروف ناشنوایان و حرف "ذ" از حروف ناشنوایان تقسیم می‌کنیم و ابعاد ویژگی را P در نظر می‌گیریم به عبارتی بردار ویژگی P بعدی هستند.

$$d = \{(x_i, c_i)\} \Big| X_{i \in R,}{}^{P} c_i \in \{-1,1\}_{i=1}^{n} \qquad (۷)$$

[1] Support Vector Machine

۲-۳-۳ شبکه عصبی پرسپترون چند لایه[1]

یک شبکه عصبی مصنوعی از یک گروه به هم پیوسته‌ی نورون‌های مصنوعی تشکیل شده است. معمولا یک شبکه عصبی پرسپترون چند لایه از مجموعه‌ای از واحدهای حسی تشکیل شده است. که یک لایه ورودی، یک یا چندین لایه‌ی پنهان و یک لایه خروجی از گره‌های محاسباتی را تشکیل می‌دهد. سیگنال ورودی به واسطه شبکه در جهت رو به جلو، بصورت لایه به لایه انتشار می‌یابد. این شبکه‌های عصبی تعمیمی از پرسپترون تک لایه را ارائه می‌دهد [۱۸] .که قادر هستند هر تابع پیوسته‌ای را با هر دقتی تخمین بزنند. اگرچه آموزش یک شبکه عصبی چند لایه توسط الگوریتم پس انتشار[2] کند است و توانایی تعمیم پذیری به وزن‌های اولیه بستگی دارد. پرسپترون‌های چندلایه از شبکه عصبی مصنوعی پیشخور با لایه‌های متعدد می‌باشد و به گونه‌ای است که هر خروجی نورون در یک لایه، ورودی نورون‌های دیگر را فراهم می‌سازد.

۲-۳-۴ درخت‌های تصمیم[3]

یادگیری درخت تصمیم روشی برای تقریب توابع هدف با مقادیر گسسته است. این روش نسبت به نویز داده‌ها مقاوم بوده و قادر است ترکیب فصلی گزاره‌های عطفی را یاد بگیرد. این روش جزو مشهورترین الگوریتم‌های یادگیری استقرائی است. اغلب الگوریتم-های یادگیری درخت تصمیم بر پایه یک عمل جستجوی حریصانه[4]

بالا به پائین[5] در فضای درخت‌های موجود عمل می‌کنند و در نهایت به گره‌های برگ می‌رسد. علت نام‌گذاری آن با درخت تصمیم این است که این درخت فرایند تصمیم گیری برای تعیین دسته ورودی را نشان می‌دهد.

[1] Multi Layer Perceptron

[2] Back Propagation

[3] Decision Tree

[4] Greedy

[5] Top-Down

۲-۳-۵ طبقه‌بندی بیزین[6]

طبقه‌بندی بیزی برای مسائلی که هر نمونه x درآن توسط مجموعه‌ای از مقادیر صفات و تابع هدف f(x) از مجموعه‌ای مانند V انتخاب می‌گردد کاربرد دارد. روش بیزی برای طبقه‌بندی نمونه جدید این است که محتمل‌ترین طبقه یا مقدار هدف v_{MAP} را با داشتن مقادیر صفات $< a1,\cdots a_n >$ که توصیف کننده نمونه جدید است شناسایی کند،

$$v_{MAP} = \arg\max_{v_j \in V} P(v_j \mid a_1, a_2, ..., a_n) \qquad (۸)$$

با استفاده از قضیه بیز می‌توان عبارت بالا را به صورت زیر بازنویسی کرد،

$$\begin{aligned} v_{MAP} &= \arg\max_{v_j \in V} \frac{P(a_1, a_2, \ldots, a_n, \mid v_j) P(v_j)}{P(a_1, a_2, \ldots, a_n)} \\ &= \arg\max_{v_j \in V} P(a_1, a_2, \ldots, a_n, \mid v_j) P(v_j) \end{aligned} \qquad (۹)$$

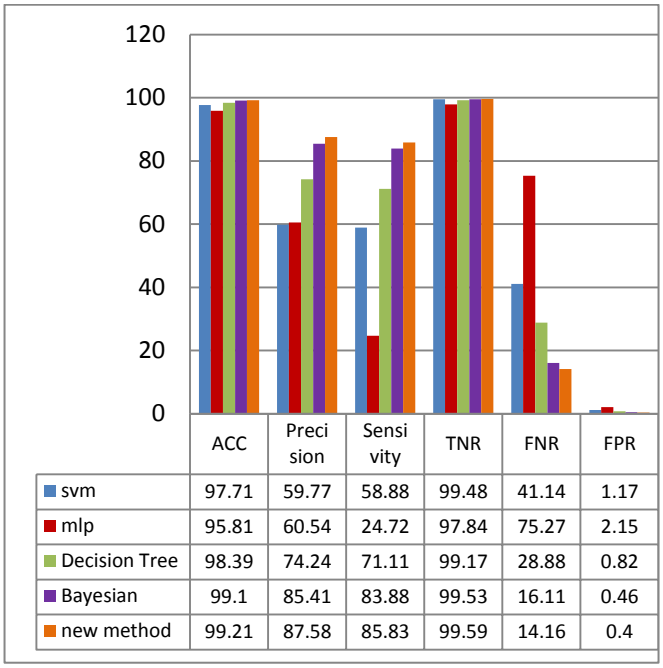
با استفاده از داده‌های آموزشی دو جمله معادله بالا را تخمین می-زنیم. محاسبه از روی داده‌های آموزشی به این صورت که میزان تکرار v_j در داده‌ها چقدر است. اما محاسبه جملات مختلف $P(a_1, a_2, ..., a_n \mid v_j)$ به این صورت قابل قبول نخواهد بود مگر اینکه حجم بسیار زیادی از داده‌های آموزشی در اختیار داشته باشیم. مشکل اینجاست که تعداد این جملات برابر تعداد نمونه‌های ممکن ضرب در تعداد مقادیر تابع هدف می‌باشد. بنابراین باید هر نمونه را چندین بار مشاهده کنیم تا تخمین مناسبی از آن بدست آید .

روش طبقه‌بندی بیزی بیانگر این است که به شرط مشاهده خروجی تابع هدف احتمال مشاهده صفات $a_1, a_2, ..., a_n$ برابر ضرب احتمالات هر صفت به طور جداگانه می‌باشد. اگر این را جایگزین معادله بالا کنیم روش طبقه‌بندی بیزین را نتیجه می‌دهد،

[6] Bayesian Belief Networks

$$v_{NB} = \arg\max_{v_j \in V} P(v_j) \prod_i P(a_i \mid v_j) \qquad (۱۰)$$

که v_{NB} خروجی طبقه‌بندی ساده بیزی برای تابع هدف می‌باشد. نتیجه اینکه یادگیری بیزین سعی در تخمین مقادیر مختلف $P(v_j)$ و $P(a_i \mid v_j)$ با استفاده از میزان تکرار آنها در داده‌های آموزشی دارد. هر گاه فرض مستقل شرطی بودن روش طبقه بندی ساده بیزی بر آورده شود طبقه‌بند بیزی معادل طبقه MAP خواهد بود.

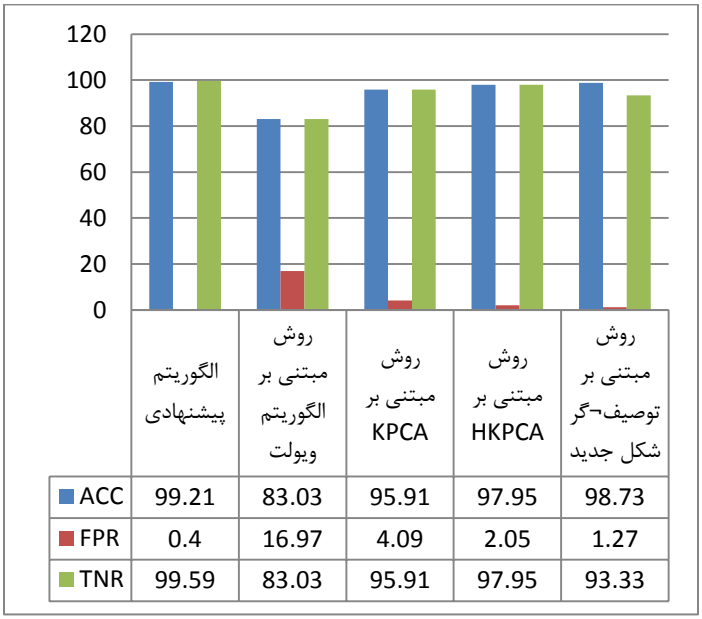


	ACC	Precision	Sensivity	TNR	FNR	FPR
svm	97.71	59.77	58.88	99.48	41.14	1.17
mlp	95.81	60.54	24.72	97.84	75.27	2.15
Decision Tree	98.39	74.24	71.11	99.17	28.88	0.82
Bayesian	99.1	85.41	83.88	99.53	16.11	0.46
new method	99.21	87.58	85.83	99.59	14.16	0.4

شکل۴:مقایسه عملکرد روش پیشنهادی برای مجموعه‌های آموزشی

۳- نتایج آزمایش

نتایج و مقایسه با کارهای ذکر شده در شکل۴ نشان داده شده است. به منظور امکان ارزیابی عملکرد روش پیشنهادی در مقایسه با روش-هایی که بر مبنای این پایگاه‌داده انجام گرفته‌اند، ابتدا تعداد داده-های آموزشی و آزمایشی را مشابه با آنها انتخاب می‌کنیم. بنابراین برای انجام آزمایش‌ها با استفاده از هر کدام از کلاس‌بندها، ابتدا ۱۸۰۰ تصویر (۲۵ تصویر برای هر علامت را برای آموزش و ۲۵ تصویر برای هر علامت را برای آزمایش) استفاده شده است. جهت مقایسه عملکرد سیستم پیشنهادی نسبت به روش‌ها [۶] و [۷] که بر مبنای پایگاه داده‌ی استفاده شده در این تحقیق به بازشناسی علامت‌های حروف الفبای زبان اشاره فارسی پرداخته‌اند، استفاده می‌کنیم. نتایج مقایسه حاکی از آن است که در شرایط یکسان آزمایشی، الگوریتم پیشنهادی، به لحاظ نرخ بازشناسی نسبت به روش مبتنی بر تبدیل ویولت[۶] عملکرد بهتر و نسبت به روش های KPCAو HKPCA [۷] و روش بازشناسی علامت‌های ساکن زبان اشاره‌ی فارسی مبتنی بر توصیف‌گر شکل جدید[۷] عملکرد نسبتاً قوی داشته است. علاوه بر آن، همان طور که در شکل۵ نشان داده شده است، قابلیت اطمینان سیستم پیشنهادی نسبت به همه روش-های موجود بالاتر می‌باشد. و همچنین سیستم ترکیبی پیشنهاد شده عملکرد بهتری نسبت به چهار دسته‌بند ماشین بردار پشتیبان، پرسپترون چند لایه، روش بیزین، درخت تصمیم دارد.



	الگوریتم پیشنهادی	روش مبتنی بر الگوریتم ویولت	روش مبتنی بر KPCA	روش مبتنی بر HKPCA	روش مبتنی بر توصیف‌گر شکل جدید
ACC	99.21	83.03	95.91	97.95	98.73
FPR	0.4	16.97	4.09	2.05	1.27
TNR	99.59	83.03	95.91	97.95	93.33

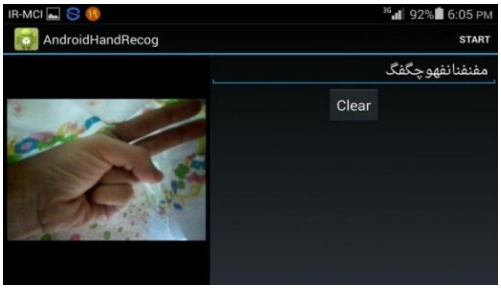
شکل۵:مقایسه الگوریتم پیشنهادی و روش‌های موجود

۳-۱- تشخیص دست در محیط‌های متفاوت

تشخیص دست در محیط ساده و با نوردهی مناسب در شکل۶ مشخص است، محیط و دست دارای تضاد مناسبی هستند. به همین دلیل است که دست به درستی تشخیص داده می‌شود و تبدیل به متن مورد نظر می‌شود. نتیجه دقت این الگوریتم برای چنین محیطی بسیار بالا است. پس اگر محیط دارای روشنایی مناسب و تضاد کامل با دست باشد، می‌توان گفت که الگوریتم با تقریب خوبی به درستی عمل می‌کند.

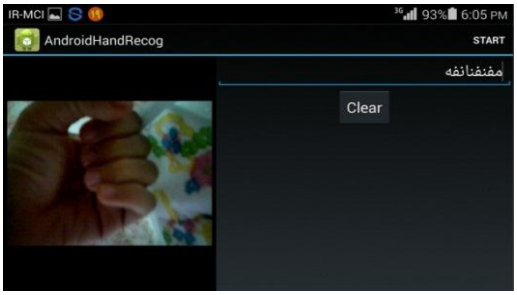
شکل۶ :مثالی از تشخیص دست در محیط با شرایط مساعد و تبدیل به متن

در شکل۷ تشخیص دست، به طوری که رنگ پس زمینه و رنگ پوست دست در محدوده رنگی تقریبا مشابهی قرار داشته باشند. در این حالت نیز برنامه کاربردی به راحتی نمی‌تواند دست را تشخیص دهد و در برخی مواقع دچار اشکال می‌شود.



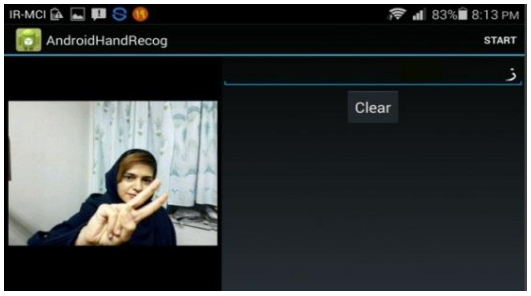
شکل ۷ :تشخیص دست در محیط با پس زمینه نزدیک با پوست دست

برنامه را در محیطی با میزان روشنایی ضعیف قرار می‌دهیم. تصاویر به دست آمده در شکل۸ نتایج را نشان می‌دهد. در تصویر اندکی اشکال در تشخیص دست مشاهده می‌شود. روشنایی صفحه نمایش به دلیل فلاش دوربینی است که از تصاویر عکس برداری شده است.



شکل ۸ :تشخیص دست در محیط با نور کم

تشخیص دست با الگوریتم بیان شده که بزرگترین کانتور را در نظر می‌گیرد، در شکل۹ نشان داده شده است. چون دست جلوتر از صورت قرار گرفته، کانتور بزرگتری را ساخته است و بنابراین دست به درستی تشخیص داده می‌شود.



شکل ۹ :تشخیص درست دست از صورت و تبدیل به متن

در شکل ۱۰ به جای دست، صورت شخص تشخیص داده شده است. در این تصویر چون محدوده انتخاب شده در قسمت دست بوده که در این تصویر همرنگ با صورت شخص است، به جای دست، صورت به عنوان کانتور بزرگتر پیدا شده است. بنابراین برای جلوگیری از بروز اینگونه مشکلات باید از کاربر خواسته شود که در شرایط مناسبی قرار گیرد.

شکل۱۰ :تشخیص نادرست دست از صورت و تبدیل نادرست به متن

۴ - نتیجه گیری

در این مقاله یک برنامه کاربردی تحت سیستم عامل اندروید و با استفاده از کتابخانه OpenCV پیاده سازی شد. برای تشخیص ژست دست و تبدیل به متن، از سامانه بینایی ماشین استفاده شد که به صورت بلادرنگ تصاویر را از ورودی می‌گیرد و بر روی آن‌ها پردازش انجام می‌دهد. روش استفاده شده، ترکیبی از فیلتر رنگ پوست، شناسایی لبه ها، پیدا کردن کانتور دست است. در این مقاله دست در محیط‌های مختلف با روشنایی‌های متفاوت شناسایی می-شود. این سیستم دارای سه مرحله تهیه مجموعه حالات دست و استخراج تصویر از پس زمینه پویا و طبقه‌بندی حالت دست استخراج شده است. در مرحله‌ی استخراج ویژگی، ابتدا با استفاده از روش خوشه‌بندی kmeans حروف الفبای ناشنوایان را به دو، سه یا چهار خوشه تبدیل کردیم به نحوی که در هر خوشه حروفی قرار گرفتند که بیشترین شباهت را با یکدیگر داشته باشند. سپس برای بازشناسی علامت‌ها با قابلیت اطمینان بالا، یک سیستم ترکیبی از چهار کلاس‌بند مختلف ماشین بردار پشتیبان و پرسپترون چند لایه و درخت‌های تصمیم و طبقه‌بندی بیزین برای هر نود این درخت پیشنهاد کردیم. بطوری که در هر نود م از یک دسته‌بندی که دقت بازشناسی بیشتری برای آن نود دارد استفاده کردیم. نتایج آزمون سیستم ارائه شده، نشان می‌دهد که در شرایط مختلف (دارای منبع یا بدون منبع روشنائی، نزدیک یا دور بودن رنگ دست با پس زمینه)، متوسط میزان تشخیص درست حالت‌های زبان اشاره ۹۹,۵۹٪ می‌باشد. همچنین، این سیستم با خطای کمتر از ۴٪ ، حالات دست را به درستی تشخیص می‌دهد. نرخ بازشناسی و قابلیت اطمینان سیستم بر روی مجموعه آموزشی پایگاه داده PSL به ترتیب ۹۹,۴۸٪ و ۹۹,۲۱٪ است که عملکرد سیستم رضایت بخش و مطلوب می‌باشد.

مراجع

[1] S.-F. Wong, and R. Cipolla, "Real-time adaptive hand motion recognition using a sparse Bayesian classifier." *Computer Vision in Human-Computer Intraction, Springer:* 170-179. 2005.

[2] L. Anton-Canalis, E. Sanchez-Nielsen, and M. Castrillon-Santana, "Fast and accurate hand pose detection for human-Computer Intraction. " *Pattern Recognition and Image Analysis, Springer:* 553-560. 2005.

[3] A. B. Jmaa, W.Mahdi, Y. B. Jemaa, and A. B. Hamadou, "Hand localization and fingers features extraction: application to digit recognition in sign language." *Intelligent Data Engineering and Automated Learning*(IDEAL), *Spring:* 151-159. 2009.

[4] A. Kumar, Prof.N.M. Wagdarikar "A Hand Gesture Recognition for Human Computer Interaction" *International Journal of Innovative Research in Computer and Communication Engineering* (IJIRCCE), *Spring:* 10957-10961. 2016.

[5] Ch. Chansri, J. Srinonchat "Hand Gesture Recognition for Thai Sign Language in Complex Background Using Fusion of Depth and Color Video" *International Electrical Engineering Congress, iEECON, Spring:* 257 – 260. 2016.

[6] G. K. Kharate and A. S. Ghotkar "Vision based multi-feature hand gesture recognition for indian sign language manual signs" *International Journal on Smart Sensing and Intelligent Systems* (ISSN), *Spring:* 12 4-147.2016.

[7] A. Karami, B. Zanj, and A. K. Sarkaleh, "Persian sign language (PSL) recognition using wavelet transform and neural network." *Expert System with Applications* 38(3): 2661-2667. 2011.

[8] M. Moghaddam, M. Nahvi, and R. H. Pak,"Static Persian Sign Language Recognition Using Kernel-Based Feature Extraction." *Machine Vision and Image Processing*(MVIP), 2011 *7th Iranian,IEEE*. 2011.

[9] A. Barkoky, and N. M. Charkari, "Static hand gesture recognition of Persian sign numbers using thinning method." *Multimedia Technology*(ICMT), 2011 *International Conference on, IEEE*. 2011.

[10] A. Kiani Sarkaleh, F. Poorahangaryan, B. Zanj, and A. Karami, "A Neural Network based system for Persian sign language recognition." *Signal and Image Processing Applications*(ICSIPA), 2009 *IEEE International Conference on, IEEE*. 2009.

[۱۱] حسین ابراهیم نژاد، ژیلا موسویان، دانشگاه صنعتی سهند، " بازشناسی علامت های زبان اشاره فارسی بر اساس تبدیل هیستوگرام گرادیان های جهت دار و شبکه عصبی احتمالاتی "، تبریز، ایران، مهرماه ۱۳۹۲

[12] B. Jalilian, A. Chalechale " Persian Sign Language Recognition Using Radial Distance and Fourier Transform" *I.J. Image, Graphics and Signal Processing*, 40-46.2014.

[13] M. Raboy, D. Canlas, R. Renejane, J. Mojica, S. Bandiala "American Sign Language Alphabet Translator Android Application: Hand Shapes Into Text " *International Conference in Education, Psychology, and Social Science* (ICEPSS), 1-7.2015

[14] Sh. Narayan Sawant " Sign Language Recognition System to aid Deaf-dumb People Using PCA " *International Journal of Computer Science & Engineering Technology* (IJCSET), *Spring:* 570-574.2014

[15] A. Chaudhary, J. Raheja, K. Das, and S. Raheja, "A survey on hand gesture recognition in context of soft computing." *Advanced Computing, Springer*: 46-55. 2011.

[16] L. Shi, Y. Wang, and J. Li, "A real time vision-based hand gestures recognition system." *Advances in Computation and Intelligence, Springer:* 349-358. 2010.

[17]شیری، مبانی خوشه بندی, http://ceit.aut.ac.ir/~shiry/lecture/ machine-learning/tutorial/ clustering/K_Means.htm دانشگاه صنعتی امیرکبیر

[18] H. Simon “Neural Networks A comprehensive Foundation,” *second Edition, Prentice Hall*, 1999.