

## سامانه تشخیص ارقام فارسی در نوشتار هوایی مبتنی بر تصویر عمق

\* رضا ملکی      \*\* شهرام محمدی

\* دکتری برق - الکترونیک، دانشگاه زنجان، سازمان تنظیم مقررات و ارتباطات رادیویی، تبریز، ایران

\*\* استادیار، دانشکده مهندسی برق، دانشگاه زنجان

تاریخ پذیرش: ۱۳۹۸/۹/۱۸      تاریخ دریافت: ۱۳۹۸/۱۱/۰۵

### چکیده

تشخیص دستنوشته از روی کاغذ، صفحه‌نمایش و یا در هوا از چالش‌هایی هستند که در بینایی ماشین وجود دارند. تشخیص نوشتار هوایی به خاطر سه‌بعدی بودن دارای چالش‌های زیادی است. در این کار تحقیقی تشخیص ارقام فارسی در نوشتار هوایی مدنظر است که در آن، کاربر ارقام صفر تا نه را در مقابل حس‌گر کینکت در هوا می‌نویسد و سامانه با استفاده از اطلاعات عمق حس‌گر قادر به تشخیص ارقام فوق است. در سامانه پیشنهادی، برای جداسازی دست و نوک انگشت از پس‌زمینه از روش خوشه‌بندی خودکار k-means، برای استخراج ویژگی از روش تغییر علامت شب خط پیشنهادی و جهت تشخیص ویژگی و تعیین رقم از دسته‌بند مدل مارکوف مخفی (HMM) استفاده شده است. دقت تشخیص سامانه پیشنهادی برای ارقام فارسی با دیتابیس محلی و باعتبار سنجی متقابل ۹۸ درصد است. سامانه پیشنهادی با نتایج چندین کار مشابه مقایسه گردید، این مقایسه‌ها نشان می‌دهند که سامانه پیشنهادی به صورت نسبی بهتر از سامانه‌های تحت مقایسه کار می‌کند.

واژه‌های کلیدی: حس‌گر کینکت، تصویر عمق، تغییرات علامت شب خط، مدل مارکوف مخفی

سه‌بعدی را تقلید نمایند. استفاده از ژست‌های دست، جایگزین جذاب و طبیعی برای تجهیزات واسط دو بعدی در محیط‌های مجازی هستند. یک ژست به عنوان حرکت فیزیکی دست، بازو، صورت و بدن باهدف انتقال اطلاعات یا معنی تعریف می‌شود [۱]. طبق تحقیقی که انجام شده است، دست به صورت گسترشده ای در مقایسه با سایر اجزای بدن جهت تعامل استفاده می‌شود [۲]. ژست‌های دست توسط تجهیزات تماسی و غیر تماسی قابل تشخیص هستند. تجهیزات غیر تماسی به خاطر طبیعی

### ۱- مقدمه

پیشرفت‌های عظیمی در حوزه‌های فناوری محاسباتی، ارتباطی و نمایشگرها انجام شده‌اند؛ اما پیشرفت چندانی در حوزه تعامل با این تجهیزات، حاصل نشده است؛ بنابراین، لازم است تنگناهای موجود مرتفع شوند. این الزام باعث شده است که در چند سال گذشته حوزه تعامل انسان - کامپیوتر (HCI) به یک حوزه تحقیقی فعالی تبدیل گردد. اگرچه در طراحی و ساخت صفحه کلیدها و موشاوه‌ها پیشرفت‌های زیادی انجام شده است؛ اما هنوز این تجهیزات، مخصوصاً در تعامل با فضاهای سه‌بعدی جهت HCI مناسب نیستند. موشاوه‌ها که دارای آزادی دو بعدی هستند نمی‌توانند فضای

شرایط نوری محیط شدیداً وابسته است. در کار دیگری، تصاویر رنگ و عمق دوربین کینکت برای آشکارسازی ارقام انگلیسی صفتاً نه نوشتۀ شده در هوا مقایسه می‌شوند [۶]. نتایج نشان می‌دهند که استفاده از تصویر عمق برای جداسازی دست و تشخیص ارقام، خطای کمتری را ایجاد می‌کند. نتایج ارزیابی سامانه‌های گزارش شده در حوزه تشخیص نوشتار هوایی نشان دادند که استفاده از تصویر عمق بجای تصویر رنگ دقیق‌تر تشخیص بالاتری را ارائه می‌دهد؛ بنابراین از روی نتایج سامانه‌های گزارش شده فوق، در این کار تحقیقی از تصویر عمق ژست دست استفاده می‌شود که با استفاده از خوشه بند خودکار k-means، دست و نوک انگشت از پس‌زمینه در هر فریم جداسازی می‌شود و درنهایت، خط سیر از اتصال نقاط نوک انگشت هر فریم حاصل می‌گردد. روش استخراج ویژگی از خط سیر ژست دست نیز مهم‌ترین مسئله‌ای است که بهصورت مستقیم روی دقیق‌ترین تشخیص ژست‌ها مؤثر است. مهم‌ترین ویژگی‌هایی که می‌توانند بردار ویژگی را از خط سیر استخراج کنند عبارت‌اند از: شبیه منحنی، طول، نقاط ماکزیمم و می‌نیمم منحنی و تعداد نقاط برخورد با منحنی. بردار تشخیص مؤثر، بردار تشخیصی است که نسبت بهاندازه، انتقال و چرخش ژست یا خط سیر مقاوم است، ابعاد بسیار کمی را ایجاد می‌کند و همچنین برای هر ژست منحصر به‌فرد است. در کار حاضر روشی پیشنهاد می‌شود که برای هر ژست دست، بردار تشخیص مؤثری را ایجاد می‌کند که این بردار به عنوان ورودی برای دسته‌بند HMM جهت تشخیص ژست دست استفاده می‌گردد. مشابه کارهای [۵] HMM [۱۲-۱۳]، در سامانه پیشنهادی از دسته‌بند HMM استفاده شده است. HMM دارای تئوری ریاضی بسیار قوی است و در صورتی که خوب آموزش ببیند دارای دقیق‌تر در تشخیص کلاس‌ها است. زمان آموزش HMM نسبت به دسته‌بندهای دیگر مانند SVM و KNN بسیار بالا است؛ اما بعدازاین که آموزش دید زمان تشخیص نسبت به دسته‌بندهای فوق بسیار پایین‌تر خواهد بود.

- 
1. Air-writing
  2. Dynamic Space Time Warping
  3. Most Discriminating Segments

بودن تعامل، ارجحیت بالاتری برای کاربر و محققین دارند. از مهم‌ترین کاربردهای تعامل با تجهیزات غیرتماسی نوشتار هوایی<sup>۱</sup> است که مانند صفحه کلید سه‌بعدی عمل می‌کند. سامانه تشخیص نوشتار هوایی در محیط‌هایی که امکان استفاده از موسواره و صفحه کلید وجود ندارد مانند محیط‌های روغی، گازی قابل استفاده است. سامانه تشخیص نوشتار هوایی می‌تواند کاراکترهای خیلی بیشتری نسبت به صفحه کلید عادی منتقل کند. این سامانه می‌تواند در حوزه بازی و گیم، صدور دستورات کنترلی، تعامل با ربات و ماشین‌ها بهصورت گسترده استفاده گردد. در حوزه تشخیص نوشتار هوایی کاراکترهای فارسی کار معابری تابه‌حال گزارش نشده است؛ بنابراین، در کار حاضر تلاش و تمرکز ما روی تشخیص نوشتار هوایی ارقام فارسی، اعداد و کلمات و کارهای آینده روی تشخیص حروف فارسی، اعداد و کلمات و عبارت‌های فارسی نوشته شده در هوا متمرکز می‌شویم. این مقاله بهصورت زیر تنظیم شده است. در بخش دوم کارهای مرتبط با تشخیص نوشتار هوایی کاراکترها بیان می‌شود. در بخش سوم الگوریتم‌های استفاده شده در بلوک‌های ساختاری سامانه تشریح می‌گردد. در بخش چهارم نتایج عملی سامانه پیشنهادی بیان می‌گردد و در قسمت پنجم نتیجه‌گیری و کارهای آینده و در انتهای مراجع ارائه می‌شوند.

**۲- کارهای مرتبط با تشخیص نوشتار هوایی کاراکتر**  
برای حوزه تشخیص نوشتار هوایی مقالات و سامانه‌های زیادی گزارش شده‌اند. استفان و همکاران از یک دوربین Rnگ با وضوح ۴۰×۳۲۰ و دسته‌بند DSTW<sup>۲</sup> برای شناسایی نوشتار هوایی ژست‌های دست ارقام انگلیسی استفاده کردند [۳]. استرن با استفاده از مفهوم MDS<sup>۳</sup> (بخشی از ژست کاراکتر است که از بخش‌های ژست‌های دیگر متفاوت است) و دسته‌بند MDSLCS (توسعه یافته LCS است) ارقام انگلیسی نوشته شده در هوا را با حس گر PrimeSense 3D تشخیص دادند [۴]. المازین و همکاران برای تشخیص نوشتار هوایی اعداد انگلیسی از تصاویر رنگ و عمق دوربین استریو استفاده کردند [۵]. آنها برای انجام این کار از دسته‌بند HMM استفاده کرده‌اند. چون از تصویر رنگ برای جداسازی دست در کار آنها استفاده شده است؛ بنابراین، تشخیص ارقام در کار آنها به

مهم‌ترین این روش‌ها استفاده از آشکارسازی پوست در تصاویر رنگی، استفاده از اطلاعات عمق دست در تصاویر عمق، استفاده از نقاط مفصل در تصاویر اسکلتی و یا استفاده از اطلاعات رنگ و عمق همزمان دست در تصاویر رنگ و عمق هستند. تصاویر رنگی شدیداً به شرایط نوری محیط وابسته می‌باشند؛ بنابراین، آشکارسازی و جداسازی دست در تصاویر رنگی با تغییرات روشنایی و سایه‌ها دچار خطای زیادی است. استفاده تنها از تصاویر اسکلتی برای آشکارسازی دست نیز در شرایط خاص امکان‌پذیر است و معمولاً استفاده از تصویر اسکلتی برای آشکارسازی دست، همراه با تصاویر رنگی انجام می‌شود<sup>[14]</sup>. تصاویر عمق در مقایسه با تصاویر رنگ جزئیات کاملی از ژست دست را انتقال نمی‌دهند؛ اما چون جزئیات انتقالی شامل نوک انگشت و مرکز دست را به صورت کامل شامل می‌باشند و این نقاط مستقل از تغییرات روشنایی محیط هستند گزینه بسیار عالی برای جداسازی دست در تصاویر هستند؛ به همین دلیل، در این کار تحقیقی از تصویر عمق ژست دست استفاده می‌شود. در اینجا، برای استخراج ناحیه دست از k-means تصویر عمق بکار گرفته شده است. خوش‌بندی k-means روش تقریب سازی برداری است که نقاط تصویر را به چندین ناحیه اختصاص می‌دهد. هر ناحیه یک خوش نامیده می‌شود. نقاط داخل هر خوش برچسب‌های برابری دارند. چگونگی قرار دادن یک نقطه در داخل یک خوش وابسته به نزدیک بودن آن نقطه به مرکز خوش است. رابطه (۱) شرط قرار دادن نقاطی مانند  $X_p$  در درون خوش  $S_i$  نشان می‌دهد<sup>[7]</sup>. در این رابطه  $k$  و  $m$  به ترتیب تعداد خوش‌ها و مرکز خوش‌ها هستند.

مرکز خوش توسط متوسط گیری نقاط درون خوش‌ها به دست می‌آید؛ که این موضوع در رابطه (۲) نشان داده شده است. در شروع، مرکز خوش‌ها به صورت تصادفی روی تصویر انتخاب می‌شوند. سپس بروز رسانی خوش‌ها با استفاده از

### ۳- الگوریتم‌های بلوک‌های ساختاری سامانه

در حالت کلی سامانه‌های گزارش شده در حوزه تشخیص نوشتار هوایی از چهار بلوک اصلی تشکیل شده‌اند که عبارت‌اند از: ۱- بلوک اخذ داده: وظیفه تبدیل کردن ژست‌های دست به تصاویر را انجام می‌دهد. ۲- پیش‌پردازش و جداسازی دست: وظیفه استخراج دست، نوک انگشت و خط سیر ژست دست را دارد و همچنین نویز و پس‌زمینه را حذف می‌کند. ۳- استخراج ویژگی: وظیفه استخراج بردار ویژگی از خط سیر را دارد. ۴- دسته‌بند: شناسایی کاراکتر از روی بردار ویژگی را انجام می‌دهد. در پیاده‌سازی سامانه پیشنهادی دقیقاً چهار بلوک فوق در نظر گرفته شده است. شکل ۱ بلوک دیاگرام سامانه پیشنهادی را نشان می‌دهد؛ که در ادامه، ساختار سامانه پیشنهادی تشریح می‌گردد.

### ۱-۱- بلوک اخذ داده

در کار حاضر، حس گر استفاده شده برای تبدیل ژست دست به تصویر، دوربین کینکت XBOX360 مایکروسافت است. این حس گر در ابتدا توسط شرکت مایکروسافت برای کنسول بازی ساخته شده بود؛ اما محققان موفق شدند بعد از هک کردن درایورهای این حس گر، آن را برای کارهای تحقیقاتی و بخصوص در حوزه بینایی ماشین بکار بگیرند. خروجی دوربین کینکت تصاویر رنگ، عمق، اسکلتی و مادون‌قرمز است. در این کار تحقیقاتی از تصویر عمق کینکت که دارای ۳۰ فریم بر ثانیه با وضوح  $640 \times 480$  است استفاده می‌شود. برای استفاده از داده‌های عمق حس گر کینکت، از کلاس‌ها و متدهای خاصی که در چارچوب Kinect SDK وجود دارند استفاده شد.

### ۱-۲- پیش‌پردازش و جداسازی دست

نوشتار هوایی توسط نوک انگشت یا به‌وسیله مرکز دست انجام می‌شود. برای استخراج نقاطی که نوشتار هوایی را انجام می‌دهند، لازم است که دست از پس‌زمینه تصویر در هر فریم استخراج شود. برای استخراج دست از پس‌زمینه روش‌های مختلفی در مقالات پیشنهاد شده است که

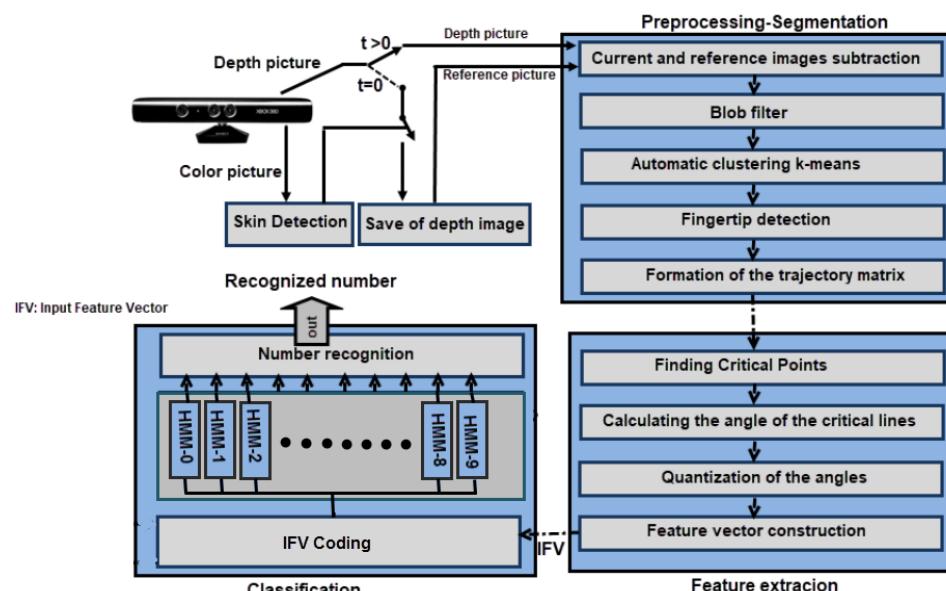
کمی از خوشها استخراج شود؛ اما زمانی که فاصله افزایش پیدا می‌کند با خوشها کم نمی‌توان ناحیه هدف را استخراج کرد. شکل ۲ این موضوع را نشان می‌دهد. چنانچه از این شکل دیده می‌شود، جهت آشکارسازی صحیح ناحیه هدف باید تعداد خوشها متناسب با فاصله ناحیه هدف از دوربین باشد.

روابط (۱) و (۲) انجام می‌شود. در بروز رسانی، مراکز خوشها حرکت می‌کنند. زمانی که تعداد نقاط در داخل خوشها تقریباً ثابت ماند، حرکت خوشها متوقف می‌شود و خوشبندی همگرا می‌شود.

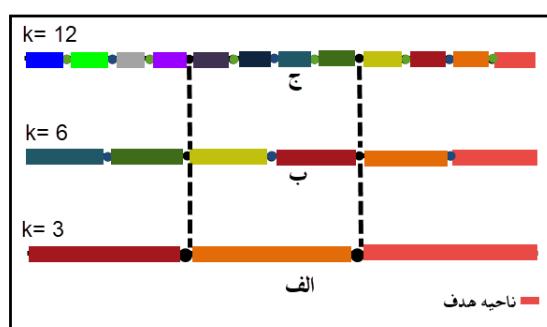
استخراج ناحیه هدف از تصویر کامل با استفاده از خوشبندی k-means به صورت شدید وابسته به تعداد خوشها است. در فاصله نزدیک که ناحیه هدف بیشترین سطح تصویر را پوشش می‌دهد، ناحیه هدف می‌تواند با تعداد

$$S_i^{(t)} = \{X_p : \|X_p - m_i^{(t)}\|^2 \leq \|X_p - m_j^{(t)}\|^2 \forall j, 1 \leq j \leq k\} \quad (1)$$

$$m_i^{t+1} = \frac{1}{S_i^{(t)}} \sum_{X_j \in S_i^{(t)}} X_j \quad (2)$$



شکل ۱- بلوك دیاگرام سامانه پیشنهادی برای شناسایی ارقام فارسی در نوشتار هوایی.



شکل ۲- ارتباط تعداد خوشها با فاصله ناحیه هدف از دوربین، الف) ناحیه هدف در فاصله نزدیک، ب) ناحیه هدف در فاصله دور، ج) ناحیه هدف در فاصله دورتر.

رابطه (۳) برابر  $k=[5.66-0.1]=5$  خواهد بود. در اینجا، خوشبندی توسط روابط (۱) و (۲) انجام شده و کوچکترین مقدار عمق در هر خوش بندی به عنوان برچسب آن خوش انتخاب می‌شود. چون دست نزدیکترین ناحیه به دوربین است بنابراین بعد از خوشبندی خودکار، خوش با برچسب کمینه انتخاب می‌شود. این خوش، نزدیکترین خوش به دوربین است و ناحیه نوک انگشت نیز در این ناحیه خواهد بود. بالاترین نقطه این تصویر نسبت به کف که دارای کمترین پهنا است به عنوان نوک انگشت خواهد بود (شکل ۳.۵). اتصال نقاط نوک انگشت پیدا شده در فریم‌های متوالی، خط سیر نوک انگشت را تشکیل می‌دهد. این خط سیر در واقع همان ژست دست کاربر است که نوشتار هوایی ارقام را انجام داده است. در کار حاضر، برای حذف اشیایی که در مقایسه با دست به دوربین نزدیک‌تر است از تغیریق هر فریم از تصویر مرجع استفاده شده است که در آن، تصویر مرجع تصویری است که قبل از نوشتار هوایی از صحنه گرفته می‌شود. خروجی تغیریق ممکن است شامل یک سری حباب‌های کوچک ناشی از حرکت‌های جزئی اشیای فوق و یا انعکاست نوری باشد که برای حذف آن از فیلتر حباب استفاده شده است که این موضوع در شکل ۶ نشان داده شده است.

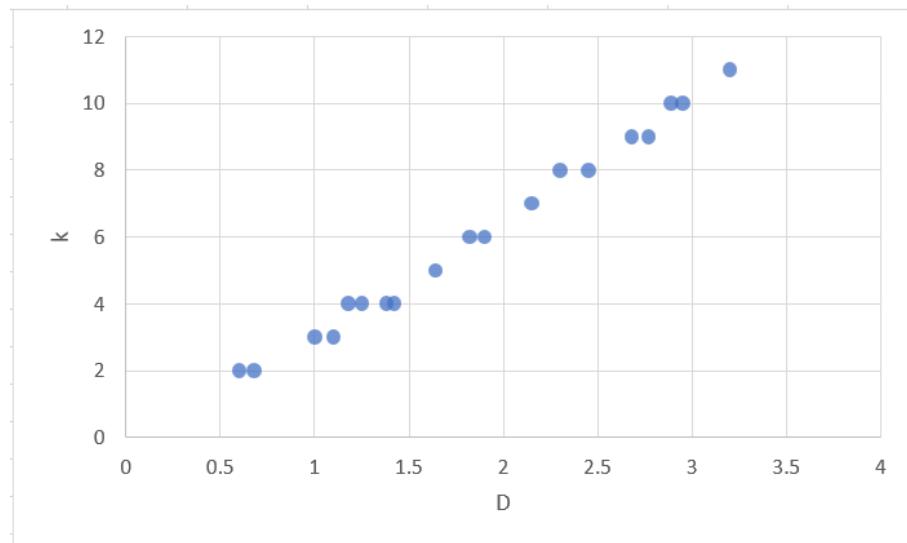
اگر تعداد خوش‌ها متناسب با فاصله ناحیه هدف از دوربین نباشد امکان استخراج ناحیه هدف وجود نخواهد داشت. شکل ۳ این مطلب را نشان می‌دهد. در این شکل ناحیه هدف که دست کاربر است در صورتی که  $k=7$  باشد قابل استخراج است (شکل ۳.الف) اما اگر  $k=3$  باشد علاوه بر دست، صورت کاربر نیز استخراج می‌شود (شکل ۳.ب) همچنین اگر  $k=9$  باشد فقط قسمتی از دست استخراج می‌شود (شکل ۳.ج). در این مقاله، برای پیاده‌سازی ارتباط بین تعداد خوش‌ها و فاصله ناحیه هدف از دوربین، خوشبندی خودکار k-means پیشنهاد می‌شود. در این روش، تعداد خوش‌ها متناسب با فاصله کاربر از دوربین است و تعداد خوش‌ها با دور شدن کاربر از دوربین به صورت خودکار افزایش پیدا می‌کند. نتیجه ۲۰ آزمایش عملی که توزیع آن در شکل ۴ آمده است، نشان داد که رابطه تعداد خوش‌ها نسبت به نزدیکترین نقطه به دوربین از رابطه (۳) به دست می‌آید. در این رابطه  $k = [D \times 3.56 - 0.1]$  به ترتیب تعداد خوش‌ها، نزدیکترین فاصله دوربین از کاربر بر حسب متر و جزء صحیح عدد هستند. رابطه (۳) از طریق درون‌بایی خطی با روش حداقل مربعات به دست آمده است [۱۵].

$$(3) \quad K = [D \times 3.56 - 0.1]$$

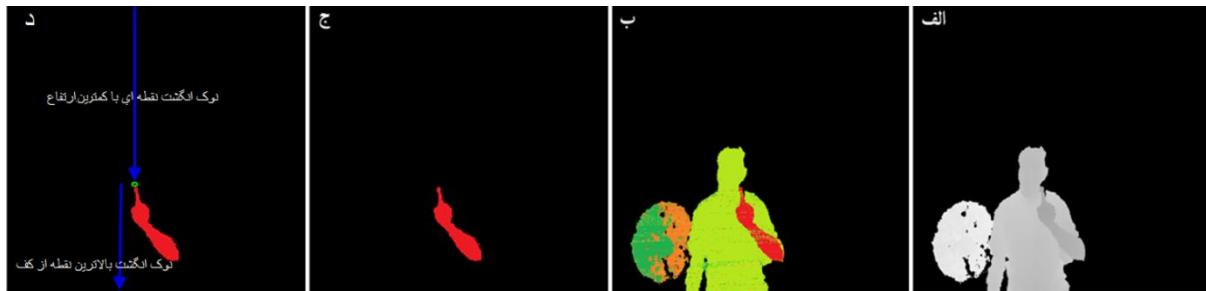
یک مثال از خوشبندی خودکار در شکل ۵ نشان داده شده است. چنانچه از این شکل دیده می‌شود، برای یک شخص در فاصله ۱۵۹ سانتی‌متری از دوربین تعداد خوش‌ها طبق



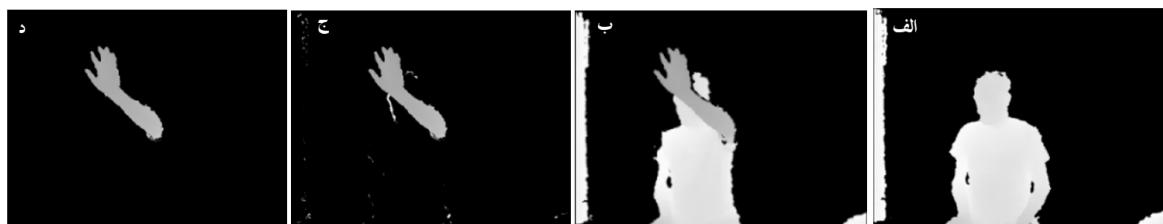
شکل ۳- نقش انتخاب  $k$  صحیح در خوشبندی k-means برای استخراج ناحیه هدف، الف) خوشبندی با تعداد خوش صحیح (۹)، ب) خوشبندی با تعداد خوش کمتر از خوش صحیح (۳)، ج) خوشبندی با تعداد بیشتر از خوش صحیح (۷).



شکل ۴- توزیع تعداد خوشه‌ها نسبت به کمترین فاصله نقطه تصویر از دوربین



شکل ۵- نحوه استخراج نقطه نوک انگشت (الف) تصویر عمق در فاصله ۱۵۹ سانتی‌متری از دوربین، (ب) نتیجه خوشه‌بندی خودکار، (ج) انتخاب نزدیک‌ترین خوشه به دوربین (د) تعیین نقطه نوک انگشت.



شکل ۶- استفاده از فیلتر Blob برای حذف حباب‌های ریز در تصویر (الف) تصویر مرجع (ب) تصویر جاری، (ج) تفریق تصویر جاری از تصویر مرجع، (د) نتیجه اعمال فیلتر Blob

یکی از متداول‌ترین روش‌های استخراج بردار ویژگی که در بیشتر مقالات استفاده شده است روش کدهای زنجیره‌ای است [۵] [۸]. در این روش، زاویه خط اتصال نقاط مجاور نسبت به افق پیدا می‌شود و با سطوح کوانتیزه مقایسه می‌گردد و برچسب مربوط به آن تعیین می‌گردد. شکل ۷،

### ۳-۳- استخراج بردار ویژگی

خط سیر استخراج شده دارای یکسری ویژگی است که این ویژگی‌ها، خط سیر مذکور را از خط سیرهای دیگر جدا می‌کند. طول، زاویه، شیب، نقاط شکست، سرعت تغییرات نقاط خط سیر از مهم‌ترین ویژگی‌های یک خط سیر هستند.

نقشه انجام می‌شود. شکل ۹، سه نقطه بحرانی E، F و M را که با استفاده از این روش پیدا شده‌اند نشان می‌دهد. تائزانت‌های معکوس خطوط مابین نقاط A با E، E با F و F با M به ترتیب اولین، دومین و سومین مؤلفه از بردار ویژگی ورودی (IFV) هستند؛ بنابراین بردار ویژگی برای این خط سیر به صورت زیر خواهد بود:

$$\text{IFV} = \{\theta_1, \theta_2, \theta_3\}$$

برای محدود کردن تعداد زاویه‌ها، می‌توان از کوانتایزر آی تقریب کننده استفاده کرد [۸]. ساده‌ترین روش تقریب، تقسیم کردن کل زاویه  $0^{\circ}$  تا  $360^{\circ}$  درجه به قسمت‌های مساوی و اختصاص یک برچسب برای هریک از نواحی است. یک مثال نمونه از تقریب زاویه در شکل ۱۰ نشان داده شده است. در این شکل ناحیه  $360^{\circ}$  درجه به قسمت‌های  $30^{\circ}$  درجه تقسیم شده است؛ و برای هریک از نواحی و همچنین محورهای افقی و عمودی برچسب‌های ۱ تا ۱۶ استفاده گردیده است. استفاده از بلوک تقریب کننده باعث می‌شود که بردار با مؤلفه‌های زاویه‌ای به بردار با مؤلفه‌های عددی محدود تبدیل شوند. شکل ۱۱ بردار ویژگی که با روش SVD برای رقم ۵ به دست می‌آید را نشان می‌دهد. زاویه نرمال از تقسیم کردن مؤلفه‌های زاویه‌ای بردار ویژگی به SVD  $360^{\circ}$  درجه به دست می‌آید. بردار ویژگی که با روش SVD به دست می‌آید، به صورت کامل به اندازه و انتقال خط سیر مقاوم است. این در حالی است که به چرخش خط سیر فقط تا  $30^{\circ}$  درجه مقاوم است. ارزیابی ریاضی نشان می‌دهد که برای مقاوم بودن به  $360^{\circ}$  درجه، کافی است که مؤلفه‌های مجاور بردار ویژگی مطابق با رابطه (۵) از هم دیگر تفرق IFV گردد. در اینجا  $L < j \leq 0$  است و طول بردار است. مقاوم بودن روش SVD با رابطه (۵) در قالب یک مثال در شکل ۱۲ نشان داده شده است. در اینجا، بردار ویژگی IFVD6 یک نسخه

## 1.Slope Variations Detection

## 2.Quanizer

این روش را به صورت ساده نشان می‌دهد. چنانچه از شکل ۷.ج مشاهده می‌گردد، این روش ابعاد زیادی را برای بردار ویژگی ایجاد می‌کند و همچنین این روش شدیداً وابسته به اندازه خط سیر و نرخ فریم است. ما از این روش استفاده کرده و روشی را توسعه دادیم که ابعاد خیلی کمی را برای بردار ویژگی ایجاد می‌کند و همچنین بردار ویژگی که از این روش تولید می‌شود مستقل از اندازه، انتقال، چرخش خط سیر و نرخ فریم است. شکل ۷.د روش پیشنهادی را روی یک خط سیر نمونه نشان می‌دهد. در کار حاضر، برای استخراج ویژگی خط سیر از روش پیشنهادی آشکارساز تغییرات شبیه<sup>۱</sup> (SVD) استفاده می‌شود. در این روش، نقاط خارجی خطوطی که از نقاط قبلی و بعدی یک نقطه در روی خط سیر عبور می‌کنند ارزیابی می‌شوند. شکل ۸ این خطوط را برای نقطه (X<sub>1</sub>, Y<sub>1</sub>) نشان می‌دهد. در اینجا، نقاط (X<sub>i</sub>, Y<sub>i</sub>) به ازای  $i=0, 1, 2, \dots, n$  مختصات نوک اंگشت در فریم از  $n$  فریم هستند که روی خط سیر قرار دارند. (X<sub>Cj</sub>, Y<sub>Cj</sub>) مختصات نقطه بحرانی  $j$  روی خط سیر را نشان می‌دهند-  $\alpha$  و  $\beta$  شبیه یا تائزانت خطوط هستند. تائزانت‌ها از رابطه (۴) محاسبه می‌شوند.

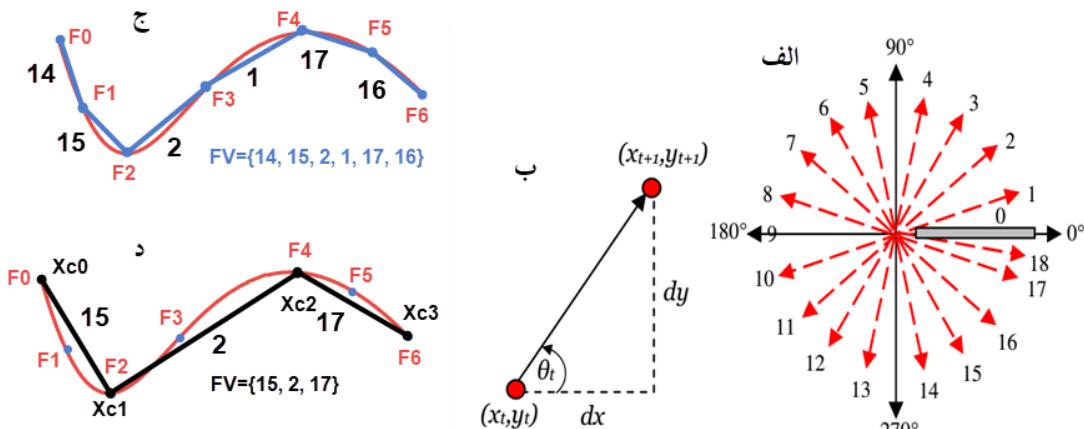
$$\begin{cases} \alpha = \tan \theta_1 = f'(x)|_{(X_1, Y_1)} = \frac{Y_1 - Y_0}{X_1 - X_0} = \frac{\Delta Y_0}{\Delta X_0} \\ \beta = \tan \theta_2 = f'(x)|_{(X_2, Y_2)} = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{\Delta Y_1}{\Delta X_1} \end{cases} \quad (4)$$

اگر علامت تائزانت‌ها در نقطه فرضی یکسان نباشند. این نقطه، به عنوان یک نقطه بحرانی عمل می‌کند. برای ارزیابی تغییرات علامت تائزانت‌ها از پارامتری بنام k استفاده می‌شود که در آن،  $k = \alpha \times \beta$  است. اگر در یک نقطه، k منفی، صفر یا بی‌نهایت باشد این نقطه، یک نقطه بحرانی خواهد بود؛ اما اگر نقطه مورد ارزیابی نقطه بحرانی نباشد یعنی  $0 < k < 1$  باشد در این حالت، پردازش بر روی نقطه بعدی انجام می‌شود؛ بنابراین،  $(X_1, Y_1) = (X_0, Y_0)$ ،  $(X_3, Y_3) = (X_2, Y_2)$  و  $(X_2, Y_2) = (X_1, Y_1)$  خواهد بود و این عمل ادامه پیدا می‌کند تا نقطه بحرانی (XC1, YC1) پیدا شود. پردازش در هر لحظه روی سه

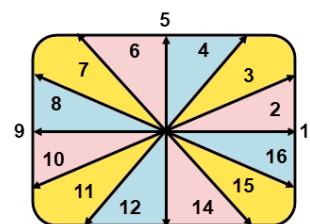
سیر و همچنین چرخش خط سیر همواره بردار ویژگی به صورت ذیل است:

$$IFV2 = \{12, 9, 7, 14\}$$

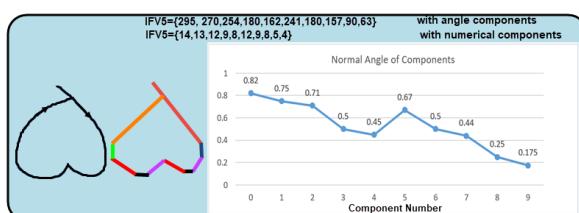
اصلاحی از IFV6 است که مستقل از چرخش خط سیر خواهد بود. شکل ۱۳ استقلال بردار ویژگی را از اندازه، انتقال و چرخش خط سیر نشان می‌دهد. در این شکل با توجه به تغییرات اندازه خط سیر، تغییرات محل شروع خط



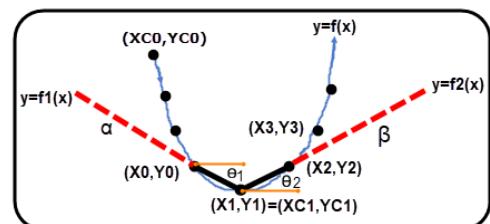
شکل ۷- مقایسه روش زنجیره کد با روش SVD پیشنهادی، (الف) سطوح کوانسیزه برای تعیین برچسب خط، (ب) روش تعیین زاویه خط دونقطه مجاور نسبت به افق، (ج) استفاده از روش زنجیره کد برای تعیین ویژگی برای خط سیر نمونه، (د) تعیین بردار ویژگی برای خط سیر نمونه با استفاده از روش پیشنهادی SVD



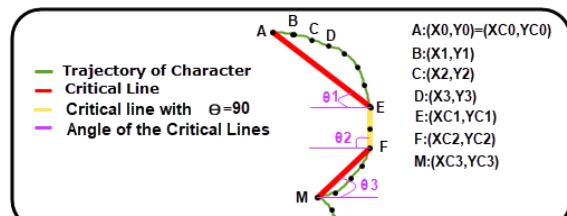
شکل ۱۰- تقریب سازی ۳۰ درجه‌ای.



شکل ۱۱- خطوط بحرانی، بردارهای ویژگی و منحنی زاویه نرمال برای رقم ۵

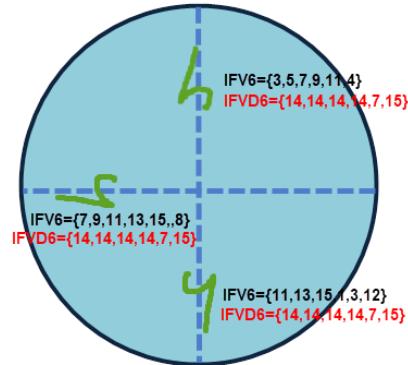


شکل ۸- خطوط گذرنده از نقاط مجاور نقطه فرضی برای تعیین نقطه بحرانی در روی خط سیر.

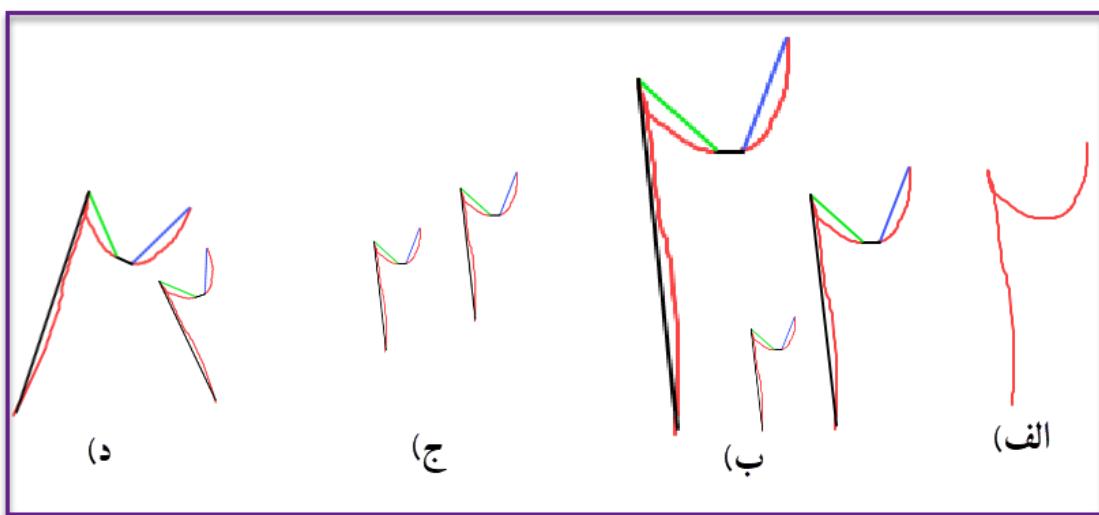


شکل ۹- سه نقطه بحرانی پیداشده روی خط سیر با روش SVD

$$\left\{ \begin{array}{lll} \Delta IFV[j] = IFV[j] - IFV[j+1] & \text{if} & \Delta IFV[j] \geq 0 \quad \& \delta \quad 0 \leq j \leq L-2 \\ IFVD[j] = \Delta IFV[j] & \text{if} & \Delta IFV[j] < 0 \quad \& \delta \quad 0 \leq j \leq L-2 \\ IFVD[j] = \Delta IFV[j] + 16 & \text{if} & (IFV[0] - IFV[L-1]) \geq 0 \\ IFVD[j] = IFV[0] - IFV[L-1] & \text{if} & |(IFV[0] - IFV[L-1])| < 0 \\ IFVD[j] = IFV[0] - IFV[L-1] + 16 & \text{if} & \end{array} \right. \quad (5)$$



شکل ۱۲- مقاوم بودن بردار ویژگی اصلاحی IFVD6 در چرخش‌های ۹۰ و ۱۸۰ درجه‌ای در جهت عقربه‌های ساعت برای رقم شش.



شکل ۱۳- مقاوم بودن بردار ویژگی به اندازه، انتقال و چرخش خط سیر، الف) خط سیر اصلی رقم ۲ ب) خطوط بحرانی و تغییر اندازه خط سیر، ج) تغییر محل خط سیر (انتقال)، د) چرخش خط سیر.

تعداد کل حالت‌ها است.

### ۴-۳- دسته‌بند HMM

- احتمال اولیه  $\pi_i$  برای هر حالت که  $i=0,1,\dots,N$  است
- ماتریس گذر  $A = \{a_{ij}\}_{N \times N}$  که  $a_{ij}$  احتمال گذر از حالت  $S_i$  به  $S_j$  است؛ که جمع ورودی‌ها نشان داده می‌شود [۹].
- مدل مارکوف مخفی (HMM) یک مدل ریاضی برای فرآیندهای استاتیکی است. HMM توسط پارامترهای زیر نشان داده می‌شود [۹].
- سری حالت‌ها  $S = \{S_1, S_2, \dots, S_N\}$

در شکل ۱۴ نشان داده شده است. چنانچه از این شکل دیده می‌شود تعداد حالتا ۸ و تعداد نمادها ۱۶ هستند و توبولوژی از نوع چپ به راست محدود است. در اینجا، برای پیدا کردن پارامترهای A، B و  $\pi$  از الگوریتم آموزش باوم-

ولش روی نمونه‌های آموزشی استفاده شده است [۹].

فرآیند کلی کار برای به دست آوردن کلاس بردار

ویژگی، به صورت ذیل است:

۱. آموزش هریک از بلوک‌های دسته‌بند هر رقم از روی نمونه‌های آموزشی جمع‌آوری شده برای آن کلاس از روی الگوریتم باوم- ولش.

۲. به دست آوردن بیشینه احتمال کلاس‌ها برای بردار مشاهده از روی ضرایب پیشرو یا پسرو.

۳. انتخاب کلاسی که بالاترین احتمال را برای بردار مشاهده نشان می‌دهد.

در مورد دسته‌بندی HMM کارهای تحقیقی زیادی گزارش شده و کدهای برنامه‌نویسی برای پیاده‌سازی این دسته‌بند در قالب مژول‌های نرم‌افزاری موجود است. در کار حاضر، برای پیاده‌سازی HMM از چارچوب Accord.net استفاده می‌شود. این چارچوب درواقع یک چارچوب یادگیری ماشین.NET است که ترکیبی از کتابخانه‌های پردازش تصویر و صوت است که در زبان C# نوشته شده است و چارچوب کاملی برای ساخت پروژه‌های بینایی ماشین با کمترین کدهای برنامه‌نویسی است. همچنین برای کاربردهای استاتیکی، پردازش سیگنال و حتی استفاده تجاری مناسب است. برای استفاده از این چارچوب کافی است که کتابخانه Statistics.DLL را به پروژه اصلی اضافه کرده و از کلاس‌ها و توابع آن استفاده کرد. استفاده از این مژول حجم کد نویسی و همچنین زمان پردازش را بسیار کاهش می‌دهد. HMM چون از خاصیت جمع استفاده می‌کند و نسبت به جابجایی مؤلفه‌ها حساس نیست

در هر ردیف از ماتریس A برابر با یک است.

- سری انتشار (مشاهده)  $\{O_1, O_2, \dots, O_T\}$  که طول مسیر ژست است.

- سری نمادهای گسسته  $\{V_1, V_2, \dots, V_M\}$  که تعداد نمادهای گسسته را بیان می‌کند.

- ماتریس مشاهده N در  $B = \{b_{im}\}_{i=1}^M$  که احتمال انتشار نماد  $V_m$  از حالت  $s_i$  است. جمع ورودی‌ها در هر ردیف ماتریس B برابر با یک است.

برای HMM سه نوع توبولوژی تعریف شده است.

۱. ارگودیگ: هر حالت به هر حالت دلخواهی می‌تواند وصل شود.

۲. مدل چپ به راست: هر حالت می‌تواند به خودش و یا به حالت‌های مستقیم وصل شود.

۳. مدل چپ به راست محدود: هر حالت می‌تواند به خودش و یا به حالت بعدی خود وصل شوند.

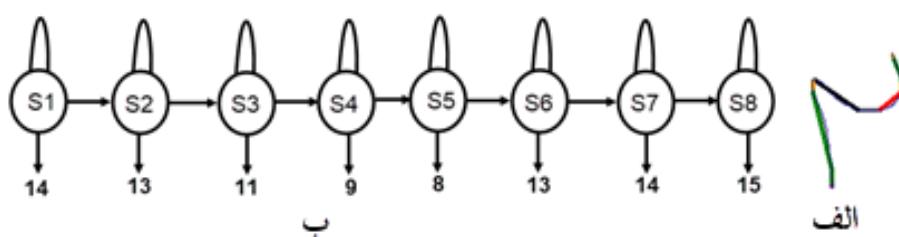
برای ساخت بلوک دسته‌بند برای هر رقم، باید تعداد حالتا و تعداد نمادهای گسسته مشخص شوند. بقیه پارامترها یعنی ماتریس گذر A، ماتریس احتمال اولیه  $\pi$  و ماتریس مشاهده B از طریق آموزش با استفاده از نمونه‌های آموزشی به دست HMM می‌آیند. انتخاب تعداد حالتا برای بلوک‌های HMM موضوعی چالشی است که هنوز جوابی قطعی برای آن ارائه نشده است. در کار حاضر، حالتا از روی بیشینه خطوط بحرانی هریک از ارقام مشخص شده‌اند. بیشینه خطوط بحرانی برای کلیه ارقام فارسی از روی نمونه‌های آموزشی مشخص می‌گردد. تعداد نمادهای گسسته برابر دامنه تغییرات مشاهدات هستند. در کار حاضر، چون از روش SVD برای استخراج بردار ویژگی استفاده می‌گردد؛ بنابراین، تعداد نمادهای گسسته ۱۶ (۱ تا ۱۶) نماد با تقریب ۳۰ درجه خواهد بود. توبولوژی HMM برای رقم فارسی دو

می‌آید. در این رابطه،  $C\_IFV$  بردار ویژگی کد شده است. همچنین  $L\_sum$  و  $com9$  به ترتیب برابر  $sum = \sum_{n=0}^{L-1} n$ ، طول بردار، تعداد برچسب‌های ۹ قبل از مؤلفه مفروض و شماره مؤلفه هستند. کد کردن بردار ویژگی باعث می‌شود که تداخلی بین بردارهای ویژگی حذف شود و درنتیجه دقت تشخیص دسته‌بند HMM افزایش پیدا کند. بردارهای کد شده برای  $\{13, 7, 10\}$  و  $\{10, 7, 13\}$  از روی رابطه (۶) به ترتیب  $\{1996, 1003, 1501, 1993, 1000, 1495\}$  می‌شود تداخل بین برداری مابین آنها توسط کدینگ حذف می‌شود. ماکریم تعداد حالت برای ارقام فارسی از روی نمونه‌های آموزشی ۱۰ می‌باشد؛ بنابراین، طبق رابطه (۶) نماد یا سمبول بیشینه ۲۷۱۱ خواهد بود.

بنابراین HMM برای بردار ویژگی ورودی مثلاً  $\{13, 7, 10\}$  و  $\{10, 7, 13\}$  احتمال یکسانی را در نظر می‌گیرد (تداخل بین برداری). این امر باعث می‌شود که دقت تشخیص دسته‌بند بسیار پایین باشد. برای حل این مشکل از کد کردن بردار ویژگی استفاده کردیم که در آن، به بردار ویژگی ارقام، موقعیت مؤلفه، طول بردار و تعداد مؤلفه‌های برچسب ۹ اضافه شد. شکل ۱۵ روش کد کردن مؤلفه با مقدار ۱ اگر طول بردار ویژگی ۲ و تعداد مؤلفه‌های ۹ قبل از مؤلفه مذکور، صفر و مؤلفه با مقدار ۱ اولین باشد مؤلفه کد شده طبق شکل ۱۵ برابر ۷ خواهد بود. کد کردن برای مؤلفه با مقدار ۲ نیز همانند شکل ۱۵ است فقط نقطه شروع از ۱۶۶ خواهد بود. در حالت کلی، مؤلفه کد شده هر مؤلفه بردار ویژگی، با اعمال ویژگی‌های ذکر شده، از رابطه (۶) به دست

$$C\_IFV[i] = (sum \times 3 + 1) + 3 \times i + com9[i] + (IFV[i] - 1)$$

(۶)

 $\times 165$ 

شکل ۱۴- مدل مارکوف مخفی الف) خطوط بحرانی رقم فارسی ۲، ب) توپولوژی HMM برای رقم فارسی ۲.

		مولده ۰	مولده ۱		
		طول بردار	اعداد مولده ۹ قبل از مولده ۰	۱	۲
۱	۱	۰	۱	۲	
۲	۴	۱	۲	۳	
۳	۱۰	۴	۵	۶	۷
۴	۱۹	۱۰	۱۱	۱۲	۱۳
۵	۳۱	۱۹	۲۰	۲۱	۲۲
۶	۴۶	۳۱	۳۲	۳۳	۳۴
۷	۶۴	۴۶	۴۷	۴۸	۴۹
۸	۸۵	۶۴	۶۵	۶۶	۶۷
۹	۱۰۹	۸۵	۸۶	۸۷	۸۸
۱۰	۱۳۶	۱۰۹	۱۱۰	۱۱۱	۱۱۲

		مولده ۸	مولده ۹		
		۰	۱	۲	۰
۱۳۳	۱۳۴	۱۳۵			
۱۶۰	۱۶۱	۱۶۲	۱۶۳		
					۱۶۴
					۱۶۵

شکل ۱۵- جدول کد کردن مؤلفه بردار ویژه با مقدار ۱.

A و Accord.net DLL مربوط به کتابخانه‌های Forge.net انجام شده‌اند. سخت‌افزار استفاده شده لپ‌تاپ با مشخصات Core i7-1.73GHz اینتل است. برای

#### ۴- نتایج عملی

در کار تحقیقی حاضر، کل پیاده‌سازی‌های نرم‌افزاری الگوریتم‌ها با استفاده از زبان برنامه‌نویسی C# و فایل‌های

تشخیص ارقام انگلیسی را انجام می‌دهد مقایسه گردید که این مقایسه در جدول ۲ نشان داده شده است. جدول ۲ نشان می‌دهد که سامانه پیشنهادی به صورت نسبی بهتر از سامانه‌های گزارش شده عمل می‌کند. فریم‌هایی از نوشتار هوایی برای رقم ۳ در شکل ۱۶ نشان داده شده است.

**جدول ۱- مقایسه دقت تشخیص متوسط HMM با بردار ویژگی کد شده و کد نشده.**

نوع کلاس‌بندی	دقت	مدت زمان آموزش	مدت زمان تشخیص
HMM with SVD	68	5.3 s	20 ms
HMM with Coded SVD	98	4.5 s	62 ms

اندازه‌گیری دقت دسته‌بند، پایگاه داده محلی با ۱۰۰۰ نمونه آموزشی تشکیل گردید که این نمونه‌ها توسط ۱۰ کاربر که هر رقم را به تعداد ۱۰ بار در مقابل دوربین به صورت موفقیت‌آمیز انجام داده بودند جمع‌آوری شده است. نتایج عملی روی دسته‌بند HMM با بردار ویژگی کد شده و کد نشده با اعتبارسنجی متقابل ۱۰ برابری پایگاه داده محلی در جدول ۱ نشان داده شده است. این جدول نشان می‌دهد که دقت تشخیص متوسط HMM ارقام برابر ۹۸ درصد است. که نرخ قابل قبولی برای یک سامانه تشخیص ارقام است. زمان تشخیص در HMM با بردار ویژگی کد شده به خاطر استفاده از ۲۷۱۱ سمبل، طولانی است؛ اما مدت زمان آموزش به خاطر این که سریع همگرا می‌شود پایین است. سامانه پیشنهادی با نتایج چندین سامانه مشابه که کار

**جدول ۲- مقایسه سامانه پیشنهادی با نتایج چندین کار مشابه گزارش شده.**

روش	اعتبارسنجی متقابل ۱۰ برابری	اعتبارسنجی متقابل دو برابری	اعتبارسنجی متقابل ۱۰ برابری
Kane and Khanna [10]	-	-	95.5
Feng [11]	86	-	100
Stern [4]	-	-	92.6
Proposed approach	92	-	98



**شکل ۱۶- فریم‌هایی از نوشتار هوایی رقم ۳.**

روش علاوه بر افزایش دقت شناسایی، حساسیت شناسایی به تغییرات روشنایی محیط را نیز حذف می‌کند. برای استخراج ویژگی از خط سیر در این سامانه، از روش پیشنهادی تغییرات علامت شبخط استفاده شد. مقاوم بودن در برابر انتقال، اندازه و چرخش خط سیر و تولید بردار

**۵- نتیجه‌گیری**  
برای تشخیص نوشتار هوایی ارقام فارسی با استفاده از تصویر عمق حسگر کینکت، سامانه‌ای پیشنهاد شد که در این سامانه برای استخراج دست و نوک انگشت از خوشه‌بندی خودکار k-means استفاده شد. استفاده از این

پیشنهادی ۹۸ درصد است. سامانه پیشنهادی با نتایج چندین کار گزارش شده مقایسه گردید و نتایج مقایسه نشان دادند که سامانه پیشنهادی عملکرد بهتری دارد. کارهای آینده ما روی تشخیص حروف، اعداد و کلمات فارسی با استفاده از حسگرهای سه بعدی جدید و ارزان متمرکز خواهد بود.

ویژگی منحصر به فرد با ابعاد کم از مهم‌ترین مزایای روش پیشنهادی در استخراج بردار ویژگی است. برای شناسایی خط سیر از روی بردار ویژگی از دسته‌بند HMM استفاده گردید. دقیق تر تشخیص متوسط در HMM به خاطر تداخل بین برداری نمونه‌های آموزشی بسیار پایین است؛ بنابراین، برای افزایش آن از کدینگ برداری روی نمونه‌ها استفاده شد. نتایج عملی نشان می‌دهند که دقیق تر تشخیص متوسط سامانه

## منابع

7. Mackay, D. 2003. Information Theory, Inference and Learning Algorithms. Cambridge University Press. pp. 284–292. ISBN 0-521-64298-1.MR 2012999.
8. Liu, N., Lovell, B. C., Kootsookos, P. J. 2003. Evaluation of HMM training algorithms for letter hand gesture recognition. Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology.
9. Rabiner, LR. 1989. A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition. Proc. of the IEEE, Vol.77, No.2, pp:257—286
10. Kane, L., Khanna, P. 2017. Vision-Based Mid-Air Unistroke Character Input Using Polar Signatures. IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS.
11. Based Mid-Air Unistroke Character Input Using Polar Signatures. IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS.
11. Feng, Z., Xu, S., Zhang, X., Jin, L., Ye, Z. 2012. Real-time Fingertip Tracking and Detection using Kinect Depth Sensor for a New Writing-in-the Air System. The 4th International Conference on Internet Multimedia Computing and Service (ICIMCS), China.
12. Elmezain, M., Alhamadi, A., Appenrodt, J., Michaelis, B. 2008. A
1. Mitra, S., Acharya, T. 2007. Gesture recognition: a survey. IEEE Trans Syst Man Cybern (SMC)Part C Appl Rev 37(3), pp:311–324.
2. Karam, M. 2006. A framework for research and design of gesture-based human computer interactions. PhD Thesis, University of Southampton.
3. Stefan, A., Athitsos, V., Alon, J., Sclaroff, S. 2008. Translation and scale invariant gesture recognition in complex scenes. in Proc. 1st ACM Int. Conf. PErvasive Technol. Related Assist. Environ., Art. no. 7.
4. Stern, H., Shmueli, M., Berman, S. 2013. Most discriminating segment Longest common subsequence (MDSLCS) algorithm for dynamic hand gesture classification. Pattern Recognit. Lett., vol. 34, no. 15, pp:1980–1989.
5. Elmezain, M., AlHamadi, A., Michaelis, B. 2009. Hand trajectory-based gesture spotting and recognition using HMM. In using HMM. In Proc. 16th IEEE Int.Conf. Image Process., pp: 3577–3580.
6. Doliotis, P., Stefan, A., McMurrrough, C., Eckhard, D., Athitsos, V. 2011. Comparing gesture recognition accuracy using color and depth information. in Proc. 4th ACM Int. Conf. Pervasive Technol. Related Assist. Environ., Art. no. 20.

International Symposium on Signal Processing and Information Technology.

14. Liu, F., Du, B., Wang, Q., Wang, Y., Zeng, W. 2017. Hand Gesture Recognition Using via Deterministic Learning. 29th Chinese Control and Decision Conference (CCDC)

۱۵. رضایی و ذهابی، ۱۳۸۹، اندازه‌گیری الکترونیکی، انتشارات دانش نگار، تهران

Hidden Markov Model-Based Isolated and Meaningful Hand Gesture Recognition.  
PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY.

13. Liu, N., Lovell, B. C., Kootsookos, P. J. 2003. Evaluation of HMM training algorithms for letter hand gesture recognition. Proceedings of the 3rd IEEE