فصلنامه علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

سال دهم، شماره‌های ۳۵ و ۳۶، بهار و تابستان ۱۳۹۷
صص: ۴۳- ۵۶

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

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 دریافت: ۱۳۹۸/۹/۱۸ تاریخ پذیرش: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 شیب‌خط پیشنهادی و جهت تشخیص ویژگی و تعیین رقم از دسته‌بند مدل مارکوف مخفی (HMM) استفاده‌شده است. دقت تشخیص سامانه پیشنهادی برای ارقام فارسی با دیتابیس محلی و بااعتبار سنجی متقابل ۱۰ برابری ۹۸ درصد است. سامانه پیشنهادی با نتایج چندین کار مشابه مقایسه گردید، این مقایسه‌ها نشان می‌دهند که سامانه پیشنهادی به‌صورت نسبی بهتر از سامانه‌های تحت مقایسه کار می‌کند.

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

۱- مقدمه

پیشرفت‌های عظیمی در حوزه‌های فناوری محاسباتی، ارتباطی و نمایشگرها انجام‌شده‌اند؛ 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[1] (HCI) به یک حوزه تحقیقی فعالی تبدیل گردد. اگرچه در طراحی و ساخت صفحه‌کلیدها و موشواره‌ها پیشرفت‌های زیادی انجام‌شده است؛ اما هنوز این تجهیزات، مخصوصاً در تعامل با فضاهای سه‌بعدی جهت HCI مناسب نیستند. موشواره‌ها که دارای آزادی دوبعدی هستند نمی‌توانند فضای سه‌بعدی را تقلید نمایند. استفاده از ژست‌های دست، جایگزین جذاب و طبیعی برای تجهیزات واسط دوبعدی در محیط‌های مجازی هستند. یک ژست به‌عنوان حرکت فیزیکی دست، بازو، صورت و بدن باهدف انتقال اطلاعات بامعنی تعریف می‌شود [۱]. طبق تحقیقی که انجام‌شده است، دست به‌صورت گسترده‌ای در مقایسه با سایر اجزای بدن جهت تعامل استفاده می‌شود [۲]. ژست‌های دست توسط تجهیزات تماسی و غیر تماسی قابل‌تشخیص هستند. تجهیزات غیر تماسی به خاطر طبیعی

1. Human-Computer Interaction

نویسنده عهده‌دار مکاتبات: رضا ملکی r.maleki@znu.ac.ir

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

بودن تعامل، ارجحیت بالاتری برای کاربر و محققین دارند. از مهم‌ترین کاربردهای تعامل با تجهیزات غیرتماسی نوشتار هوایی[1] است که مانند صفحه‌کلید سه‌بعدی عمل می‌کند. سامانه تشخیص نوشتار هوایی در محیط‌هایی که امکان استفاده از موشواره و صفحه‌کلید وجود ندارد مانند محیط‌های روغنی، گازی قابل‌استفاده است. سامانه تشخیص نوشتار هوایی می‌تواند کاراکترهای خیلی بیشتری نسبت به صفحه‌کلید 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 معتبری تابه‌حال گزارش نشده است؛ بنابراین، در کار حاضر تلاش و تمرکز ما روی تشخیص نوشتار هوایی ارقام فارسی صفرتا نه است و در کارهای آینده روی تشخیص حروف فارسی، اعداد و کلمات و عبارت‌های فارسی نوشته‌شده در هوا متمرکز می‌شویم.این مقاله به‌صورت زیر تنظیم‌شده است. در بخش دوم کارهای مرتبط با تشخیص نوشتار هوایی کاراکترها بیان می‌شود. در بخش سوم الگوریتم‌های استفاده‌شده 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 انتها 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 رنگ با وضوح ۲۴۰×۳۲۰ و دسته‌بند DSTW[2] برای شناسایی نوشتار هوایی ژست‌های دست ارقام انگلیسی استفاده کردند [۳]. استرن با استفاده از مفهوم MDS[3] (بخشی از ژست کاراکتر است که از بخش‌های ژست‌های دیگر متفاوت است) و دسته‌بندMDSLCS (توسعه‌یافته LCS است) ارقام انگلیسی نوشته‌شده در هوا را با حس‌گر 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 استفاده‌شده است؛ بنابراین، تشخیص ارقام در کار آنها به شرایط نوری محیط شدیداً وابسته است. در کار دیگری، تصاویر رنگ و عمق دوربین کینکت برای آشکارسازی ارقام انگلیسی صفرتا نه نوشته‌شده 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 گزارش‌شده در حوزه تشخیص نوشتار هوایی نشان دادند که استفاده از تصویر عمق بجای تصویر رنگ دقت تشخیص بالاتری را ارائه می‌دهد؛ بنابراین از روی نتایج سامانه‌های گزارش‌شده 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 دقت تشخیص ژست‌ها مؤثر است. مهم‌ترین ویژگی‌هایی که می‌توانند بردار ویژگی را از خط سیر استخراج کنند عبارت‌اند از: شیب منحنی، طول، نقاط ماکزیمم و می نیمم 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 منحصربه‌فرد است. در کار حاضر روشی پیشنهاد می‌شود که برای هر ژست دست، بردار تشخیص مؤثری را ایجاد می‌کند که این بردار به‌عنوان ورودی برای دسته‌بند HMM جهت تشخیص ژست دست استفاده می‌گردد. مشابه کارهای [۵] [۱۲-۱۳]،در سامانه پیشنهادی از دسته‌بند HMM استفاده‌شده است. HMM دارای تئوری ریاضی بسیار قوی است و درصورتی‌که خوب آموزش ببیند دارای دقت بالا 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

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

۳- الگوریتم‌های بلوک‌های ساختاری سامانه

در حالت کلی سامانه‌های گزارش‌شده در حوزه تشخیص نوشتار هوایی از چهار بلوک اصلی تشکیل‌شده‌اند 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 گرفته‌شده است. شکل ۱ بلوک دیاگرام سامانه پیشنهادی را نشان می‌دهد؛ که در ادامه، ساختار سامانه پیشنهادی تشریح می‌گردد.

۳-۱- بلوک اخذ داده

در کار حاضر، حس گر استفاده‌شده برای تبدیل ژست دست به تصویر، دوربین کینکت XBOX360 مایکروسافت است. این حس‌گر در ابتدا توسط شرکت مایکروسافت برای کنسول بازی ساخته شده بود؛ اما محققان موفق شدند بعد از هک کردن درایورهای این حس‌گر، 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 ۴۸۰×۶۴۰ است استفاده می‌شود. برای استفاده از داده‌های عمق حس‌گر کینکت، از کلاس‌ها و متدهای خاصی که در چارچوب Kinect SDK وجود دارند استفاده شد.

۳-۲- 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 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[۱۴]. 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 به‌صورت 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 پس‌زمینه خوشه‌بند k-means تصویر عمق بکار گرفته‌شده است. خوشه‌بندی k-means روش تقریب سازی برداری است که نقاط تصویر را به چندین ناحیه اختصاص می‌دهد. هر ناحیه یک خوشه نامیده می‌شود. نقاط داخل هر خوشه برچسب‌های برابری دارند. چگونگی قرار دادن یک نقطه در داخل یک خوشه وابسته به نزدیک بودن آن نقطه به مرکز خوشه است. رابطه (۱) شرط قرار دادن نقاطی مانند X_p را در درون خوشه S_i نشان می‌دهد[۷]. در این رابطه k و m به ترتیب تعداد خوشه‌ها و مرکز خوشه‌ها هستند.

مرکز خوشه توسط متوسط گیری نقاط درون خوشه‌ها به دست می‌آید؛ که این موضوع در رابطه (۲) نشان داده‌شده است. در شروع، مرکز خوشه‌ها به‌صورت تصادفی روی تصویر انتخاب می‌شوند. سپس بروز رسانی خوشه‌ها با استفاده از

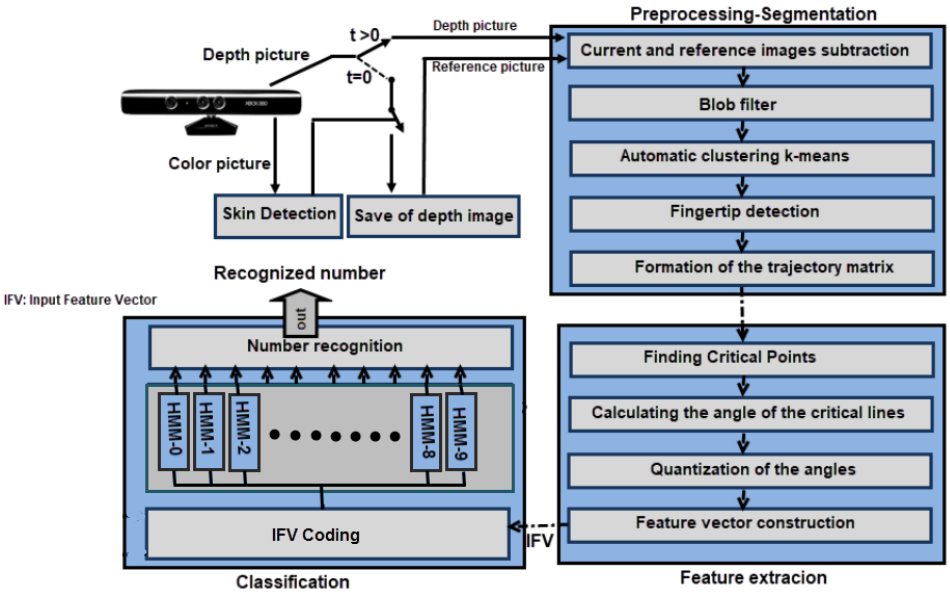
Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

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

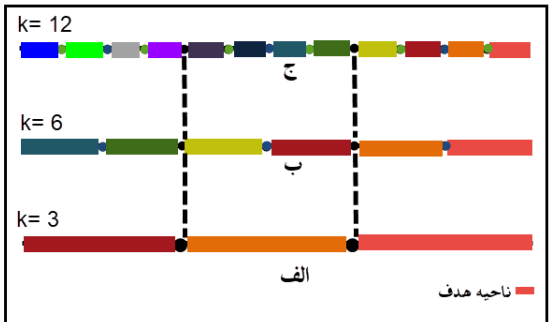
استخراج ناحیه هدف از تصویر کامل با استفاده از خوشه‌بندی k-means به‌صورت شدید وابسته به تعداد خوشه‌ها است. در فاصله نزدیک که ناحیه هدف بیشترین سطح تصویر را پوشش می‌دهد، ناحیه هدف می‌تواند با تعداد کمی از خوشه‌ها استخراج شود؛ اما زمانی که فاصله افزایش پیدا می‌کند با خوشه‌های کم نمی‌توان ناحیه هدف را استخراج کرد. شکل ۲ این موضوع را نشان می‌دهد. چنانچه از این شکل دیده می‌شود، جهت آشکارسازی صحیح ناحیه هدف باید تعداد خوشه‌ها متناسب با فاصله ناحیه هدف از دوربین باشد.

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

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



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



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

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

اگر تعداد خوشه‌ها متناسب با فاصله ناحیه هدف از دوربین نباشد امکان استخراج ناحیه هدف وجود نخواهد داشت. شکل ۳ این مطلب را نشان می دهد. دراین شکل ناحیه هدف که دست کاربر است درصورتی‌که k=7 باشد قابل‌استخراج است (شکل ۳.الف) اما اگر k=3 باشد علاوه بر دست، صورت کاربر نیز استخراج می‌شود (شکل ۳.ب) همچنین اگر k=9 انتخاب شود فقط قسمتی از دست استخراج می‌شود (شکل۳.ج). در این مقاله، برای پیاده‌سازی ارتباط بین تعداد خوشه‌ها و فاصله ناحیه هدف از دوربین، خوشه‌بندی خودکار k-means پیشنهاد می‌شود. در این روش، تعداد خوشه‌ها متناسب با فاصله کاربر از دوربین است و تعداد خوشه‌ها با دور شدن کاربر از دوربین به‌صورت خودکار افزایش پیدا می‌کند. نتیجه ۲۰ آزمایش عملی که توزیع آن در شکل ۴ آمده است، نشان داد که رابطه تعداد خوشه‌ها نسبت به نزدیک‌ترین نقطه به دوربین از رابطه (۳) به دست می‌آید. در این رابطه k، D و[.] به ترتیب تعداد خوشه‌ها، نزدیک‌ترین فاصله دوربین از کاربر برحسب متر و جزء صحیح عدد هستند. رابطه (۳) از طریق درون‌یابی خطی با روش حداقل مربعات به‌دست‌آمده است [۱۵].

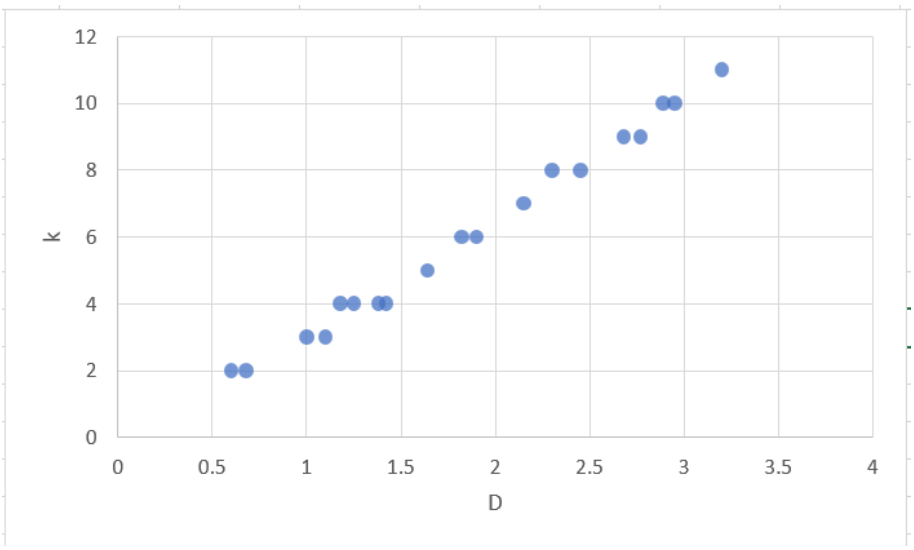
$$K = [D \times 3.56 - 0.1] \qquad (۳)$$

یک مثال از خوشه‌بندی خودکار در شکل ۵ نشان داده‌شده است. چنانچه از این شکل دیده می‌شود، برای یک شخص در فاصله ۱۵۹ سانتی‌متری از دوربین تعداد خوشه‌ها طبق رابطه (۳) برابر k=[5.66-0.1]=5 خواهد بود. در اینجا، خوشه‌بندی توسط روابط (۱) و (۲) انجام‌شده و کوچک‌ترین مقدار عمق در هر خوشه به‌عنوان برچسب آن خوشه انتخاب می‌شود. چون دست نزدیک‌ترین ناحیه به دوربین است بنابراین بعد از خوشه‌بندی خودکار، خوشه با برچسب کمینه انتخاب می‌شود. این خوشه، نزدیک‌ترین خوشه 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 به‌عنوان نوک انگشت خواهد بود (شکل ۵.د). اتصال نقاط نوک انگشت پیداشده در فریم‌های متوالی، خط سیر نوک انگشت را تشکیل می‌دهد. این خط سیر درواقع 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 تفریق هر فریم از تصویر مرجع استفاده‌شده است که در آن، تصویر مرجع تصویری است که قبل از نوشتار هوایی از صحنه گرفته می‌شود. خروجی تفریق 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 استفاده‌شده است که این موضوع در شکل ۶ نشان داده‌شده است.

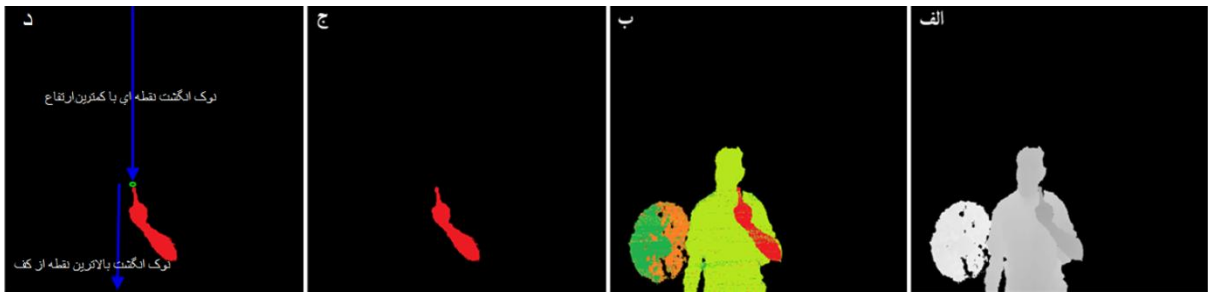


شکل ۳- نقش انتخاب k صحیح در خوشه‌بندی k-means برای استخراج ناحیه هدف، الف) خوشه‌بندی با تعداد خوشه صحیح (k=7)، ب) خوشه‌بندی با تعداد خوشه کمتر از خوشه صحیح (k=3)، ج) خوشه‌بندی با تعداد بیشتر از خوشه صحیح (k=9).

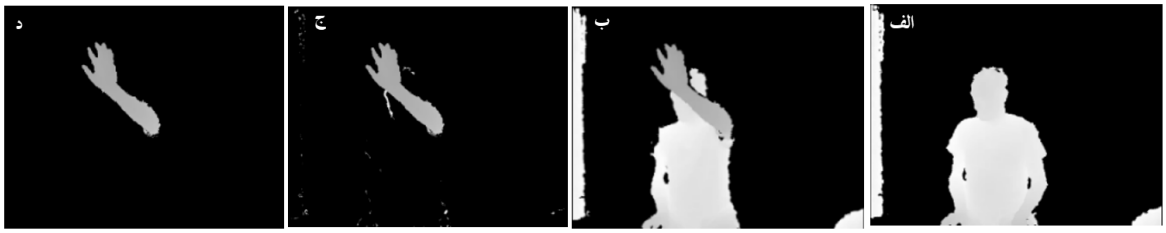
Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026



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 حباب های ریز در تصویر الف) تصویر مرجع ب) تصویر جاری، ج) تفریق تصویر جاری از تصویر مرجع، د) نتیجه اعمال فیلتر Blob.

۳-۳- استخراج بردار ویژگی

خط سیر استخراج‌شده 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.

یکی از متداول‌ترین روش‌های استخراج بردار ویژگی که در بیشتر مقالات استفاده‌شده 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،

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

این روش را به‌صورت 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 شیب[1] (SVD) استفاده می‌شود. در این روش، تانژانت‌های خطوطی که از نقاط قبلی و بعدی یک نقطه در روی خط سیر عبور می‌کنند ارزیابی می‌شوند. شکل ۸ این خطوط را برای نقطه (X1, Y1) نشان می‌دهد. در اینجا، نقاط (Xi,Yi) به ازای i=0, 1, 2 ,…,n مختصات نوک انگشت در فریم i از n فریم هستند که روی خط سیر قرار دارند. (XCj, YCj) مختصات نقطه بحرانی j روی خط سیر را نشان می‌دهند- α و β شیب یا تانژانت خطوط هستند. تانژانت‌ها از رابطه (۴) محاسبه می‌شوند.

$$\begin{cases} \alpha = \tan\theta 1 = f'(x)|_{(X1.Y1)} = \frac{Y1-Y0}{X1-X0} = \frac{\Delta Y0}{\Delta X0} \\ \beta = \tan\theta 2 = f'(x)|_{(X2.Y2)} = \frac{Y2-Y1}{X2-X1} = \frac{\Delta Y1}{\Delta X1} \end{cases} \quad (۴)$$

اگر علامت تانژانت‌ها در نقطه فرضی یکسان نباشند. این نقطه، به‌عنوان یک نقطه بحرانی عمل می‌کند. برای ارزیابی تغییرات علامت تانژانت‌ها 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 $k > 0$ باشد در این حالت، پردازش بر روی نقطه بعدی انجام می‌شود؛ بنابراین، (X1, Y1) = (X0, Y0)، (X2, Y2) = (X1, Y1) و (X3, Y3) = (X2, Y2) خواهد بود و این عمل ادامه پیدا می‌کند تا نقطه بحرانی (XC1, YC1) پیدا شود. پردازش در هر لحظه روی سه نقطه انجام می‌شود. شکل ۹، سه‌نقطه بحرانی E، F و M را که با استفاده از این روش پیداشده‌اند نشان می‌دهد. تانژانت‌های 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 به‌صورت زیر خواهد بود:

IFV= {θ1, θ2, θ3}

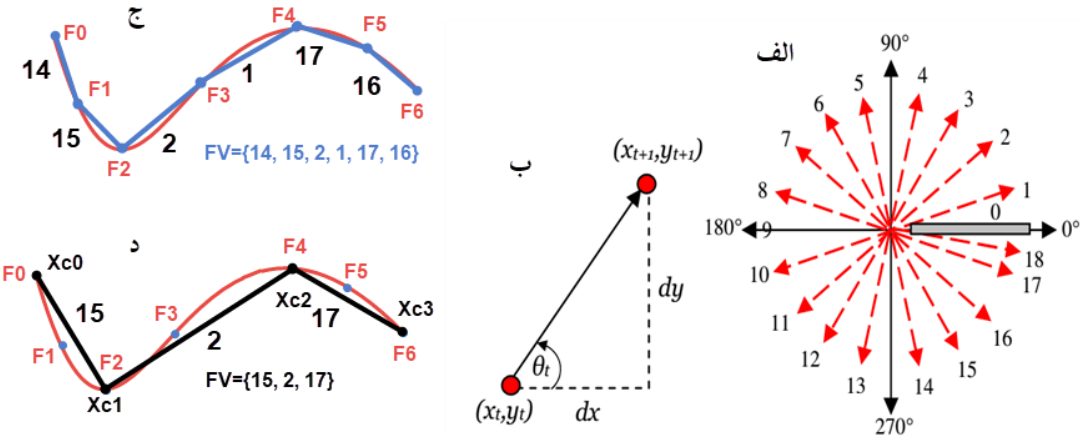
برای محدود کردن تعداد زاویه‌ها، می‌توان از کوانتایزر[2] یا تقریب کننده استفاده کرد [۸]. ساده‌ترین روش تقریب، تقسیم کردن کل زاویه ۰ تا ۳۶۰ درجه به قسمت‌های مساوی و اختصاص یک برچسب برای هریک از نواحی است. یک مثال نمونه از تقریب زاویه در شکل ۱۰ نشان داده‌شده است. در این شکل ناحیه ۳۶۰ درجه به قسمت‌های ۳۰ درجه تقسیم‌شده 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 ۳۶۰ درجه به دست می‌آید. بردار ویژگی که با روش SVD به دست می‌آید، به‌صورت کامل به اندازه و انتقال خط سیر مقاوم است. این در حالی است که به چرخش خط سیر فقط تا ۳۰ درجه مقاوم است. ارزیابی ریاضی نشان می‌دهد که برای مقاوم بودن به ۳۶۰ درجه، کافی است که مؤلفه‌های مجاور بردار ویژگی مطابق با رابطه (۵) از هم دیگر تفریق گردند. در اینجا $0 \le j < L$ است و L طول بردار IFV است. مقاوم بودن روش SVD با رابطه (۵) در قالب یک مثال در شکل ۱۲ نشان داده‌شده است. در اینجا، بردار ویژگی IFVD6 یک نسخه

1.Slope Variations Detection
2.Quanrizer

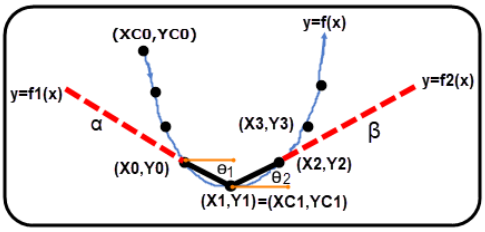
Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

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 سیر و همچنین چرخش خط سیر همواره بردار ویژگی به‌صورت ذیل است:

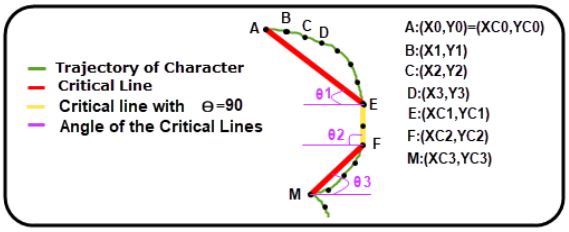
IFV2= {12, 9, 7, 14}



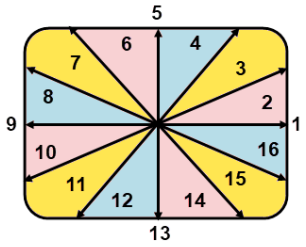
شکل ۷- مقایسه روش زنجیره کد با روش SVD پیشنهادی، الف) سطوح کوانتیزه برای تعیین برچسب خط، ب) روش تعیین زاویه خط دونقطه مجاور نسبت به افق، ج) استفاده از روش زنجیره کد برای تعیین ویژگی برای خط سیر نمونه، د) تعیین بردار ویژگی برای خط سیر نمونه با استفاده از روش پیشنهادی SVD



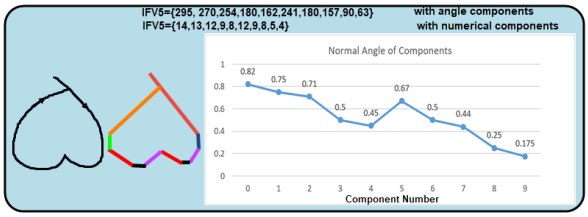
شکل ۸- خطوط گذرنده از نقاط مجاور نقطه فرضی برای تعیین نقطه بحرانی در روی خط سیر.



شکل ۹- سه‌نقطه بحرانی پیداشده روی خط سیر با روش SVD.



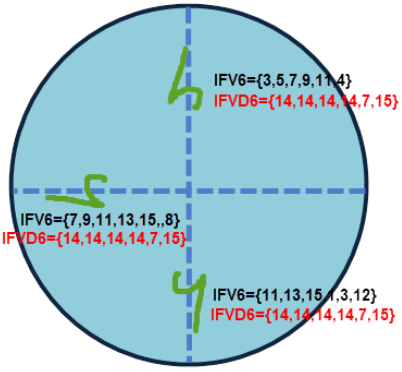
شکل ۱۰- تقریب سازی ۳۰ درجه‌ای.



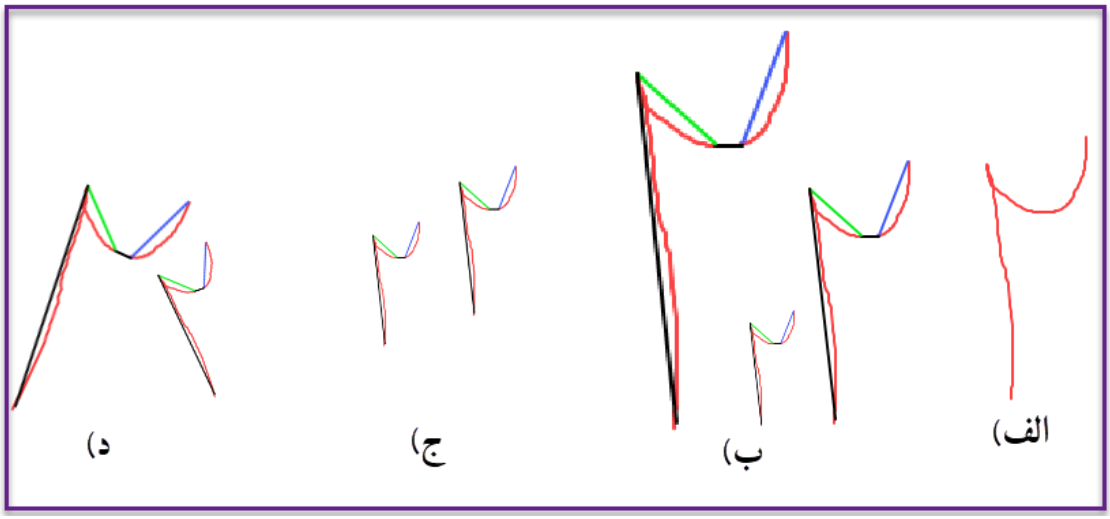
شکل ۱۱- خطوط بحرانی، بردارهای ویژگی و منحنی زاویه نرمال برای رقم ۵.

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

$$\begin{cases} \Delta IFV[j]=IFV[j]-IFV[j+1] & \\ IFVD[j]=\Delta IFV[j] & \text{if} \quad \Delta IFV[j]\geq 0 \;\&\delta\; 0\leq j\leq L-2 \\ IFVD[j]=\Delta IFV[j]+16 & \text{if} \quad \Delta IFV[j]<0 \;\&\delta\; 0\leq j\leq L-2 \\ IFVD[j]=IFV[0]-IFV[L-1] & \text{if} \quad (IFV[0]-IFV[L-1])\geq 0 \\ IFVD[j]=IFV[0]-IFV[L-1]+16 & \text{if} \quad (IFV[0]-IFV[L-1])<0 \end{cases} \tag{5}$$



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.

۴-۳- دسته‌بند HMM

مدل مارکوف مخفی (HMM) یک مدل ریاضی برای فرآیندهای استاتیکی است. HMM توسط پارامترهای زیر نشان داده می‌شود[۹].

- سری حالت‌ها $S=\{s_1, s_2,\ldots, s_N\}$ که در آن N تعداد کل حالت‌ها است.
- احتمال اولیه π_i برای هر حالت که $i=0,1,\ldots,N$ است و $\pi_i = P(s_i)$
- ماتریس گذر $A=\{a_{ij}\}$، $N\times N$ که a_{ij} احتمال گذر از حالت S_i به S_J که $1\leq i \leq N, J$ است؛ که جمع ورودی‌ها

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

در هر ردیف از ماتریس A برابر با یک است.

- سری انتشار (مشاهده) $O= \{o_1,o_2, \ldots, o_T\}$ که T طول مسیر ژست است.
- سری نمادهای گسسته $V= \{v_1, v_2, \ldots., v_M\}$ که M تعداد نمادهای گسسته را بیان می‌کند.
- ماتریس مشاهده N در M، $B=\{ b_{im} \}$ که b_{im} 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 توپولوژی تعریف‌شده 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 حالت‌ها و تعداد نمادهای گسسته مشخص شوند. بقیه پارامترها یعنی ماتریس گذر A، ماتریس احتمال اولیه π و ماتریس مشاهده B از طریق آموزش با استفاده از نمونه‌های آموزشی به دست می‌آیند. انتخاب تعداد حالت‌ها برای بلوک‌های HMM موضوعی چالشی است که هنوز جوابی قطعی برای آن ارائه نشده است. در کار حاضر، حالت‌ها از روی بیشینه خطوط بحرانی هریک از ارقام مشخص‌شده‌اند. بیشینه خطوط بحرانی برای کلیه ارقام فارسی از روی نمونه‌های آموزشی مشخص می‌گردند. تعداد نمادهای گسسته برابر دامنه تغییرات مشاهدات هستند. درکار 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. توپولوژی HMM برای رقم فارسی دو در شکل ۱۴ نشان داده‌شده است. چنانچه از این شکل دیده می‌شود تعداد حالت‌ها ۸ و تعداد نمادها ۱۶ هستند و توپولوژی از نوع چپ به راست محدود است. در اینجا، برای پیدا کردن پارامترهای A، B و π از الگوریتم آموزش باوم– ولش روی نمونه های آموزشی استفاده‌شده است [۹].

فرآیند کلی کار برای به دست آوردن کلاس بردار ویژگی، به‌صورت ذیل است:

۱. آموزش هریک از بلوک‌های دسته‌بند هر رقم از روی نمونه‌های آموزشی جمع‌آوری‌شده 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.

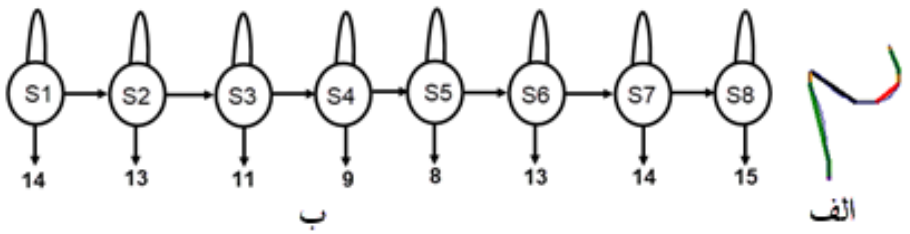
در مورد دسته‌بندی HMM کارهای تحقیقی زیادی گزارش شده و کدهای برنامه‌نویسی برای پیاده‌سازی این دسته‌بند در قالب ماژول‌های نرم‌افزاری موجود است. در کار حاضر، برای پیاده‌سازی HMM از چارچوب Accord.net استفاده می‌شود. این چارچوب درواقع یک چارچوب یادگیری ماشینNET. است که ترکیبی از کتابخانه‌های پردازش تصویر و صوت است که در زبان C# نوشته‌شده 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 ماژول 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

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

بنابراین HMM برای بردار ویژگی ورودی مثلاً {۱۳و ۷و ۱۰} و {۱۰و ۱۳و ۷} 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 را نشان می‌دهد؛ بنابراین، برای 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 ذکرشده، از رابطه (۶) به دست می‌آید. در این رابطه، C_IFV بردار ویژگی کد شده است. همچنین sum، L، com9 و i 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 {۱۳و ۷و ۱۰} و {۱۰و ۱۳و ۷} از روی رابطه (۶) به ترتیب {۱۹۹۶و ۱۰۰۳و ۱۴۹۵} و {۱۵۰۱و ۱۹۹۳و ۱۰۰۰}هستند و چنانچه دیده می‌شود تداخل بین برداری مابین آنها توسط کدینگ حذف می‌شود. ماکزیمم تعداد حالت برای ارقام فارسی از روی نمونه‌های آموزشی ۱۰ می‌باشد؛ بنابراین، طبق رابطه (۶) نماد یا سمبل بیشینه ۲۷۱۱ خواهد بود.

$$C_IFV[i]= (sum\times 3+1) +3\times i+com9[i]+(IFV[i]-1) \quad (۶)$$

$$\times 165$$



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

تعداد مولفه 9 قبل از مولفه	طول بردار	مولفه 0			مولفه 1			...	مولفه 8			مولفه 9		
		0	1	2	0	1	2		0	1	2	0	1	2
	1	1	2	3										
	2	4	5	6	7	8	9							
	3	10	11	12	13	14	15							
	4	19	20	21	22	23	24							
	5	31	32	33	34	35	36							
	6	46	47	48	49	50	51							
	7	64	65	66	67	68	69							
	8	85	86	87	88	89	90							
	9	109	110	111	112	113	114		133	134	135			
	10	136	137	138	139	140	141		160	161	162	163	164	165

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

۴- نتایج عملی

درکار تحقیقی حاضر، کل پیاده‌سازی‌های نرم‌افزاری الگوریتم‌ها با استفاده از زبان برنامه‌نویسی C# و فایل‌های DLL مربوط به کتابخانه‌های Accord.net و A Forge.net انجام شده‌اند. سخت‌افزار استفاده‌شده لپ‌تاپ با مشخصات Core i7-1.73GHz اینتل است. برای

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

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 به‌صورت موفقیت‌آمیز انجام داده بودند جمع‌آوری‌شده 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 داده‌شده 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 مدت‌زمان آموزش به خاطر این‌که سریع همگرا می‌شود پایین است. سامانه پیشنهادی با نتایج چندین سامانه مشابه که کار تشخیص ارقام انگلیسی را انجام می‌دهد مقایسه گردید که این مقایسه در جدول ۲ نشان داده‌شده است. جدول ۲ نشان می‌دهد که سامانه پیشنهادی به‌صورت نسبی بهتر از سامانه‌های گزارش‌شده عمل می‌کند. فریم‌هایی از نوشتار هوایی برای رقم ۳ در شکل ۱۶ نشان داده‌شده است.

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

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

جدول ۲- مقایسه سامانه پیشنهادی با نتایج چندین کار مشابه گزارش‌شده.

روش	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



شکل۱۶- فریم‌هایی از نوشتار هوایی رقم ۳.

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 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 شیب‌خط استفاده شد. مقاوم بودن در برابر انتقال، اندازه و چرخش خط سیر و تولید بردار

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026

ویژگی منحصربه‌فرد با ابعاد کـم از مهـم‌ترین مزایای روش پیشـنهادی در استـخراج بردار ویژگی است. برای شناسـایی خط سیر از روی بردار ویژگی از دسته‌بند HMM استفاده گردید. دقت 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 دقت تشخیص متوسط سامانه پیشنهادی ۹۸ درصد است. سامانه پیشنهادی با نتایج چندین کار گزارش‌شده 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 سه‌بعدی جدید و ارزان متمرکز خواهد بود.

منابع

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., McMurrough, 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.

7. Mackay, D. 2003. Information Theory, Inference and Learning Algorithms. Cambridge University Press. pp. 284–292. ISBN 0-52 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

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 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

Downloaded from jor.iranaict.ir at 1:12 +0330 on Monday November 9th 2026