بازشناسی برخط حروف مجزای دست‌نویس فارسی بر اساس تشخیص گروه اصلی بدنه با استفاده از مدل مخفی مارکوف

محمد امین مهرعلیان
دانشکده‌ی مهندسی کامپیوتر و فناوری اطلاعات
دانشگاه صنعتی امیرکبیر
mehralian@aut.ac.ir

کاظم فولادی
دانشکده‌ی مهندسی برق و کامپیوتر
دانشگاه تهران
kfouladi@ut.ac.ir

چکیده: در بازشناسی برخط حروف، داده‌های نوشته شده توسط قلم بر روی یک صفحه‌ی حساس که طبیعتاً حاوی عنصر ترتیب زمانی نیز هستند، برای بازشناسی حرف نوشته شده مورد پردازش قرار می‌گیرد. در روش ارائه شده در این مقاله، بازشناسی هر حرف فارسی، در دو مرحله صورت می‌گیرد. در مرحله‌ی اول حرف ورودی با استفاده از مدل مخفی مارکوف در قالب یکی از هجده گروه بدنه‌ی اصلی حروف، طبقه‌بندی می‌شود و سپس در مرحله‌ی دوم موقعیت، تعداد و شکل سایر حرکت‌ها مانند نقطه و سرکش (ریزحرکت‌ها)، حرف نهایی را تعیین می‌کند. به عنوان نمونه برای تشخیص حرف «ت» ابتدا گروه بدنه‌ی «ب، پ، ت، ث» تشخیص داده می‌شود و سپس وجود ریزحرکت «دونقطه» در بالای آن منجر به انتخاب «ت» از این گروه می‌شود. نتایج تجربی این کار پژوهشی که بر اساس مجموعه داده‌ی Online-TMU صورت گرفته است، متوسط نرخ بازشناسی بدنه‌ی اصلی را ۹۳٪ نشان می‌دهد و با در نظر گرفتن پس‌پردازش‌ها بر اساس ریزحرکت‌ها این نرخ به حدود ۹۶٪ می‌رسد.

واژه‌های کلیدی: بازشناسی برخط حروف فارسی، بازشناسی دست‌نوشته/ دست‌نویس، ریزحرکت، مدل مخفی مارکوف.

۱- مقدمه

«بازشناسی نوری نویسه‌ها» یا OCR یکی از زیرشاخه‌های کاربردی بسیار فعال در زمینه‌ی بازشناسی الگو است. مساله‌ی اصلی در این شاخه، بازشناسی نویسه‌ها، زیرکلمات و کلماتی است که به نوعی تصویر آنها در اختیار ما قرار دارد. حل این مساله برای ارتباط مؤثر میان انسان و ماشین نقش چشمگیری دارد و می‌تواند در خودکارسازی فرآیند پردازش اسناد مکتوب کمک شایانی نماید.

یک سیستم بازشناسی، می‌تواند برخط (on-line) یا برون‌خط (off-line) باشد. چنانچه اطلاعات زمانی از ترتیب نگارش نقاط موجود نباشد، سیستم برون‌خط خواهد بود و داده‌های آن معمولاً با اسکن کردن تصاویر متون از قبل نوشته شده به دست می‌آید. اگر ترتیب زمانی نقاط در زمان نوشتن با قلم موجود باشد، سیستم با داده‌های برخط سروکار دارد. استفاده از این سیستم در شرایطی ممکن است که متن توسط یک قلم مخصوص بر روی یک صفحه‌ی حساس نوشته شود. به طور خاص استفاده از این رویکرد در دستگاه‌هایی چون Tablet PC یا PDA بسیار طبیعی به نظر می‌رسد. در واقع در دستگاه‌هایی همچون PDA، وارد کردن متن از طریق نوشتن روی صفحه‌ی حساس، معمولاً از کار کردن با صفحه‌کلیدهای کوچک ساده‌تر به نظر می‌رسد و این رو کاربران چنین دستگاه‌هایی بعضاً مایلند از قابلیت بازشناسی دست‌نویس برای ورود اطلاعات استفاده نمایند.

در این مقاله قصد داریم به ارائه‌ی روشی جدید برای بازشناسی برخط حروف مجزای فارسی بپردازیم. در روش ارائه شده سعی شده است کمترین تعداد قید ممکن بر روی نحوه‌ی نگارش حروف قرار داده شود. اصلی‌ترین قیدی که وجود دارد، این است که کاربر بایستی ابتدا «بدنه»ی حروف را بنویسد و سپس نقطه‌گذاری آنها را انجام دهد.

در مقایسه با زبان‌های لاتین، برای بازشناسی برخط حروف فارسی و عربی، تاکنون کارهای بسیار کمتری انجام شده است. مرور مختصری بر روی مهم‌ترین کارهای انجام شده در این زمینه در بخش ۲ ارائه می‌شود.

این مقاله در هشت بخش سازمان‌دهی شده است. در بخش۲ مروری بر کارهای انجام شده در زمینه‌ی بازشناسی حروف دست‌نویس برخط فارسی/ عربی ارائه می‌شود. بخش ۲ مدلی را برای حروف دست‌نویس فارسی ارائه می‌کند و خصوصیات آن را بیان می‌کند. الگوریتم بازشناسی حروف و جزئیات آن در بخش‌های ۴، ۵ و ۶ تحت عنوان‌های پیش‌پردازش، استخراج ویژگی و طبقه‌بندی و پس‌پردازش ارائه می‌شود. در بخش ۷ به مسایل مربوط به پیاده‌سازی و نتایج تجربی پرداخته می‌شود. سرانجام بخش ۸ به خلاصه‌ی مطالب، بحث و نتیجه‌گیری اختصاص داده شده است.

۲- مروری بر کارهای انجام شده توسط دیگران

در زمینه‌ی بازشناسی دست‌نوشته‌ی برخط، برخی کارها به بازشناسی ارقام و برخی دیگر به بازشناسی حروف یا کلمات پرداخته‌اند. از آنجا

شرایط آزمایش و معیارهای ارزیابی این کارها با یکدیگر متفاوت است، مقایسه‌ی عددی آنها را بیان نمی‌کنیم و فقط روش‌های مورد استفاده در آنها را مورد توجه قرار می‌دهیم.

به عنوان یکی از نخستین کارهایی که به بازشناسی برخط حروف عربی پرداخته است، می‌توان به [1] اشاره کرد که در آن از کدهای زنجیره‌ای فری‌من برای نمایش ویژگی‌های ساختاری و نیز بازنمایی سلسله‌مراتبی نویسه‌ها استفاده شده است. در [2] برای بازشناسی حروف عربی، از یک شبکه‌ی عصبی فازی استفاده شده است و نظریه‌ی «تولید حرکت»، برای تولید ویژگی‌های لازم که مبتنی بر سرعت حرکت قلم در هنگام نوشتن می‌باشد، مورد استفاده قرار گرفته است. ترکیب رویکرد فازی و ساختاری را نیز می‌توان در [3] مشاهده کرد. در این کار از ترکیب ویژگی‌های ساختاری محلی برای توصیف فشرده‌ی حروف و از یک پایگاه قواعد فازی برای طبقه‌بندی استفاده شده است.

رویکرد دیگری در [4] ارائه شده است که در آن از «نقشه‌های خودسازمان‌ده» (SOM) برای بازشناسی حروف که خود با استفاده از یک روش بازنمایی شکل خاص (مبتنی بر توزیع تجربی چند ویژگی از دنباله نقاط حرف) توصیف شده‌اند، به کار رفته است. رویکرد مشابهی در [5] با استفاده از حافظه‌های شرکت‌پذیر قابل مشاهده است. در [6] نیز یک سیستم کلی برای بازشناسی نویسه‌های عربی برخط پیشنهاد شده است. در این کار سعی شده است ویژگی‌های لازم از روی داده‌های برخط به کمک چند SOM «نقشه‌ی خودسازمان‌ده» استخراج شود تا نیازی به استخراج ویژگی‌ها با روش‌های هیوریستیک نباشد. برای بارشناسی، طبقه‌بندی‌کننده‌ی پرسپترون نسبت به MLP و برنامه‌ریزی ژنتیک قابل رقابت دیده شده است. همچنین در [7] یک بازنمایی نویسه‌های برخط عربی/ فارسی با استفاده از تقریب چندجمله‌ای مختصات ذخیره شده‌ی نقاط پیشنهاد شده است و با استفاده از این بازنمایی یک شبکه عصبی کوهونن برای طبقه‌بندی طراحی شده است.

در [8] برای بازشناسی نویسه‌های عربی، ابتدا یک گرافیتی عربی معرفی شده است که قیود لازم بر روی نحوه‌ی نوشتن حروف را با یک ضرب (stroke) نشان می‌دهد به طوری که شکل گرافیتی برای بعضی از حروف، بکلی با شکل عادی آن متفاوت است. برای طبقه‌بندی و بازشناسی گرافیتی‌ها از یک شبکه‌ی عصبی سه لایه استفاده شده است.

شبکه‌ی عصبی کوهونن گیبسی، روش دیگری است که در [9] برای بازشناسی نویسه‌های عربی استفاده شده است. در این کار بازنمایی نویسه‌ها با هیستوگرام ویژگی‌ها به عنوان یک توزیع تجربی، انجام شده است. یک ویژگی قابل توجه این کار این است که هیچ قید خاص روی نحوه‌ی نوشتن نویسه‌ها قرار داده نشده است.

کار دیگری در [10] ارائه شده است، یک سیستم همراه در کلاس درس برای یادداشت‌نویسی است که نویسه‌های عربی را شناسایی می‌کند. در این سیستم بازشناسی به کمک یک شبکه‌ی عصبی توسعه یافته با مفاهیم فازی انجام می‌شود و پارامترهای منحنی دست‌نوشته بر اساس مدل «بتا ـ الیپتیکال» استخراج می‌گردد.

رویکرد فازی علاوه بر [2] و [3] با روش‌های مختلفی در برخی کارهای دیگر نیز دیده می‌شود. در [11] بازنمایی پارامترهای دست‌نوشته با استفاده از متغیرهای زبانی در منطق فازی انجام شده است. همچنین در کار دیگری مرتبط با همین کار برای بازشناسی برخط حروف مجزای فارسی، از منطق فازی استفاده شده است. در این کار برای توصیف و تشخیص قطعات بدنه‌ی حروف، از یک الگوریتم قطعه‌بندی سلسله‌مراتبی استفاده شده است که یک طبقه‌بندی کننده‌ی فازی را به کار می‌گیرد. بازشناسی الگوی بدنه و تشخیص علایم ثانویه در دو مرحله‌ی پی در پی صورت می‌گیرد [۱]. در [12] از روش تطابق الگوی الاستیک بر روی الگوهای حروف بازنمایی شده به کمک دنباله‌ی متغیرهای زبانی فازی استفاده شده است.

در [۲] روشی برای بازشناسی حروف مجزای فارسی ارائه شده است. در این روش حروف بر اساس تعداد نقاط و تعداد حرکات در نوشتن در گروه‌هایی دسته‌بندی شده‌اند و قیدهای اندکی برای نوشتن حروف در نظر گرفته شده است. ابتدا اجزای کوچک مانند نقطه‌ها و سرکش‌ها با استفاده از یک شبکه‌ی عصبی تشخیص داده می‌شوند و با استفاده از موقعیت آنها و چند قاعده‌ی ساده، گروه حرف تعیین می‌شود. در مرحله‌ی بعدی با استفاده از نوعی فاصله‌سنجی بازشناسی نهایی حرف صورت می‌گیرد. همچنین در [۳] تعمیم این روش را با در نظر گرفتن موارد ضروری برای بازشناسی برخط زیرکلمات فارسی ارائه شده است.

رهیافت مهم دیگری که در بازشناسی برخط مورد استفاده قرار گرفته است، و ما نیز در این مقاله از آن استفاده کرده‌ایم، استفاده از «مدل مخفی مارکوف» (HMM) است. در [13] برای بازشناسی برخط کلمات دست‌نویس، روشی مبتنی بر مدل مخفی مارکوف ارائه شده است و در آن سعی شده است برای برخی از مشکلات ذاتی در بازشناسی دست‌نوشته‌های عربی مانند اتصال حروف، شکل مستقل از مکان حروف و همپوشانی زیرکلمات راه‌حل‌هایی ارائه شود. همچنین در [14] روشی را برای بازشناسی کلمات دست‌نویس عربی، با استفاده از مدل مخفی مارکوف ارائه کرده است. واحد اصلی بازشناسی، استروک‌ها (stroke) هستند که در واقع بخش‌های تشکیل‌دهنده‌ی حروف می‌باشند. برای بازشناسی استروک‌ها، از مدل مخفی مارکوف استفاده شده است. سپس از یک منطق تصمیم‌گیری برای تفسیر خروجی HMMها و تبدیل آنها به کلمات بازشناسی شده استفاده می‌شود. برای بررسی عملکرد سیستم، از داده‌های جمع‌آوری شده از شش نویسنده استفاده شده است. به عنوان نمونه‌ای دیگر، در [15] ابتدا کلمات به زیرکلمات شکسته می‌شوند. به هر ضرب از زیرکلمه، یک برچسب نسبت داده می‌شود. برای آخرین مرحله‌ی طبقه‌بندی از تعدادی HMM استفاده می‌شود.

۳- مدل مورد نظر برای حروف دست‌نویس برخط فارسی

در زبان فارسی ۳۲ حرف وجود دارد. قسمت اصلی حروف «بدنه» نامیده می‌شود. برخی از حروف دارای علایم اضافی، مانند نقطه، سرکش یا کلاه هستند که اصطلاحاً به آنها «ریز حرکت» می‌گوییم [۲]. ویژگی مهمی که در بازشناسی حروف فارسی قابل توجه است این است که برخی از حروف بدنه‌ی یکسان دارند و تنها تفاوت آنها در ریزحرکت‌های آنهاست. با توجه به بدنه‌های مشابه، حروف فارسی را می‌توان در قالب ۱۸ گروه دسته‌بندی کرد که این گروه‌ها در جدول ۱ قابل مشاهده هستند.

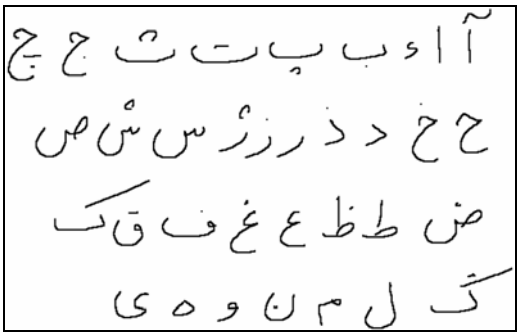
جدول ۱- گروه بندی انجام شده بر اساس بدنه حروف

گروه ۱	آ، ا	گروه ۷	ص، ض	گروه ۱۳	ل
گروه ۲	ب، پ، ت، ث	گروه ۸	ط، ظ	گروه ۱۴	م
گروه ۳	ج، چ، ح، خ	گروه ۹	ع، غ	گروه ۱۵	ن
گروه ۴	د، ذ	گروه ۱۰	ف	گروه ۱۶	و
گروه ۵	ر، ز، ژ	گروه ۱۱	ق	گروه ۱۷	ه
گروه ۶	س، ش	گروه ۱۲	ک، گ	گروه ۱۸	ی

استفاده از رویکرد برخط این امکان را به ما می‌دهد که اطلاعات ترتیب زمانی در هنگام نوشتن حروف و یا تعداد حرکات قلم را در اختیار داشته باشیم. در این کار فرض می‌کنیم که بدنه‌ی اصلی حروف در ابتدا نوشته می‌شود و سایر حرکات مانند نقاط، سرکش‌ها و . . . در مراحل بعدی گذاشته می‌شود. به عنوان نمونه حرف «چ» در شکل ۱ با سه حرکت قلم نوشته می‌شود، به این ترتیب که ابتدا بدنه‌ی اصلی «ح» و در دو مرحله‌ی بعدی نقاط قرار داده می‌شوند. به نظر می‌رسد که ذهن انسان نیز در فرآیند تشخیص حروف، ابتدا بدنه‌ی اصلی را بازشناسی می‌کند و سپس موقعیت، تعداد و شکل ریزحرکت‌ها هستند که مشخص می‌کند حرف مورد نظر چیست.

شکل ۱ - ترتیب نگارش بدنه‌ی اصلی و ریزحرکت‌ها

تنها محدودیتی که در نحوه‌ی نوشتن ریزحرکت‌ها وجود دارد، این است که «سه نقطه» می‌بایستی از ترکیب «دو نقطه» و یک «تک نقطه» ساخته شده باشد. همچنین دو سرکش «گ» باید در دو حرکت مجزای قلم نوشته شود. بعلاوه‌ی دسته‌ی «ط» و «ظ» و سرکش «ک» بایستی در حرکتی بجز حرکت بدنه‌ی اصلی نوشته شود. به عبارت دیگر حروف «ط» و «ک» باید در دو حرکت قلم و حروف «گ» و «ظ» باید در سه حرکت قلم نگاشته شوند. شکل ۲، صورت صحیح نگارش ۳۳ حالت مختلف حروف فارسی (شامل ۳۲ حرف و آ) را نشان می‌دهد که در آن هر قسمت با یک حرکت قلم نوشته شده است [۲].



شکل ۲ – نحوه‌ی نگارش صحیح برای ۳۳ حالت مختلف از حروف

۴- پیش‌پردازش

از آنجا که داده‌های برخط توسط قلم بر روی صفحه‌ی حساس نوشته می‌شوند، مختصات نقاط نمونه‌برداری شده، تعداد و فاصله‌ی آنها و همچنین ابعاد منحنی‌های ایجاد شده از یک حرف به حرف دیگر و از یک نویسنده به نویسنده‌ی دیگر می‌تواند بشدت تغییر کند. برای اینکه بتوان تاثیر این تغییرات را در فرآیند بازشناسی به حداقل رساند، باید نوعی نرمال‌سازی روی مجموعه نقاط نمونه‌برداری شده از حرکات قلم انجام شود. در این کار نرمال‌سازی فاصله‌بین نقاط نمونه‌برداری، یکسان کردن تعداد نقاط نمونه‌برداری و یکسان کردن ابعاد حرکت‌ها به عنوان مهم‌ترین موارد پیش‌پردازش انجام می‌شود.

۴-۱ نرمال‌سازی فواصل میان نقاط نمونه‌برداری

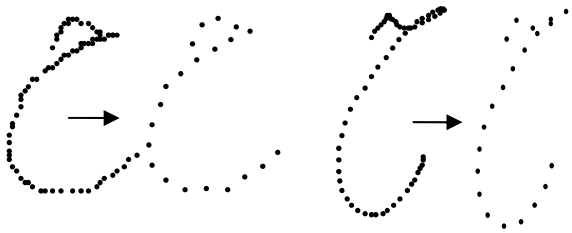
برای نرمال‌سازی فواصل میان نمونه‌ها، ابتدا میانگین فاصله‌ی نقاط متوالی در مجموعه نقاط مربوط به یک حرکت را اندازه‌گیری می‌کنیم و نقاطی را که در فاصله‌ای کمتر از دو برابر این میانگین باشند، حذف می‌کنیم. ضرورت این نرمال‌سازی از این جهت است که در بعضی بخش‌های حرف که سرعت حرکت قلم کم بوده است، تعداد نقاط نمونه‌برداری شده افزایش می‌یابد و در بخش‌هایی که سرعت حرکت قلم زیاد بوده است، تعدا نمونه‌ی کمتری مشاهده می‌شود. به علاوه، با حذف این نقاط اضافی، لرزش‌های کمتری نیز در مسیر حرکت دیده می‌شود و از این رو لرزش‌های ناخواسته در نوشتن حرف نیز تا حدی فیلتر می‌شود. شکل ۳ دو نمونه از حروف و نتیجه‌ی نرمال‌سازی فواصل میان نقاط آنها دیده می‌شود.

شکل ۳ - دو نمونه از نرمال‌سازی فواصل میان نقاط

۲-۴ یکسان‌سازی تعداد نقاط نمونه‌برداری

یکسان‌سازی تعداد نقاط نمونه‌برداری برای سهولت استفاده از مدل مخفی مارکوف (HMM) و افزایش کارایی آن، ضروری است. در واقع با این کار همه‌ی حروف با یک تعداد مشخص از برش‌های زمانی نمونه‌برداری مجدد می‌شوند و به داده‌ی ترتیبی مناسب HMM تبدیل می‌شوند.

برای این منظور، ابتدا با استفاده از اسپیلاین‌های درجه‌ی سوم یک درون‌یابی بین هر چهار نقطه‌ی متوالی از منحنی حرف انجام می‌شود و به این ترتیب تخمینی از شکل کلی حرف صورت می‌گیرد. سپس طول منحنی به ۲۰ قسمت مساوی تقسیم می‌شود به طوری که طول منحنی بین هر دو نقطه‌ی متوالی، اندازه‌ای برابر دارد و در نتیجه ۲۱ نقطه با فواصل مساوی از روی منحنی حاصل نمونه‌برداری می‌شود. البته تعداد این نقاط برای ریزحرکت‌ها به ۱۰ نقطه کاهش می‌یابد. شکل ۴ دو مثال از انجام یکسان‌سازی تعداد نقاط نمونه‌برداری را نشان می‌دهد.



شکل ۴ - دو نمونه از یکسان‌سازی تعداد نقاط نمونه‌برداری

۳-۴ یکسان‌سازی ابعاد و مختصات نقطه‌ی شروع

از آنجا که ابعاد حرف نوشته شده توسط کاربران می‌تواند متفاوت باشد، بنابراین لازم است پیش از بازشناسی آن، یک همسان‌سازی در ابعاد صورت گیرد. برای این منظور حرف نوشته شده در چارچوب مختصاتی به ابعاد ۱۰۰×۱۰۰ قرار می‌گیرد.

بعلاوه از آنجا که نقطه‌ی شروع نگارش حرف می‌تواند هر نقطه‌ای روی صفحه باشد، نقطه‌ی شروع برای همه‌ی حروف به نقطه‌ی (0,0) منتقل می‌شود.

۵- استخراج ویژگی‌ها و طبقه‌بندی

برای انجام بازشناسی بدنه‌ی حروف و زیرحرکت‌ها با استفاده از مدل مخفی مارکوف لازم است که ویژگی‌های مناسبی را جهت بازنمایی هر یک از آنها انتخاب کنیم.

۱-۵ استخراج ویژگی‌ها

برای انتخاب ویژگی‌ها، زیرمجموعه‌ای از ویژگی‌های معرفی شده در [16] را مورد استفاده قرار می‌دهیم. برای هر نقطه‌ی نمونه مجموعه‌ی شش ویژگی زیر را محاسبه می‌کنیم:

$$f_1(x_i, y_i) = x_i \quad (۱)$$

$$f_2(x_i, y_i) = y_i \quad (۲)$$

$$f_3(x_i, y_i) = x_i - x_{i-1} (= \Delta x_i) \quad (۳)$$

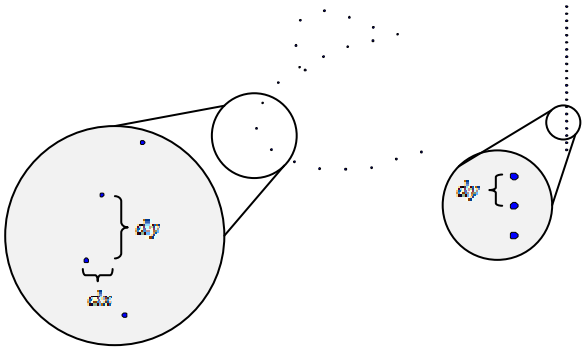
$$f_4(x_i, y_i) = y_i - y_{i-1} (= \Delta y_i) \quad (۴)$$

$$f_5(x_i, y_i) = \sin(\arctan(\Delta y_i / \Delta x_i)) \quad (۵)$$

$$f_6(x_i, y_i) = \cos(\arctan(\Delta y_i / \Delta x_i)) \quad (۶)$$

ویژگی (۱) و (۲) همان مختصات دکارتی نقطه است که در مرحله‌ی پیش‌پردازش مقادیر نرمال‌شده‌ی آن محاسبه شده است. همان طور که پیش‌تر گفته شد مقدار این ویژگی‌ها برای نقطه‌ی شروع، صفر می‌باشد. از این رو سیستم روند حرکت را برای هر حرف، از مبدا مختصات می‌بینند.

ویژگی (۳) و (۴) میزان تغییر مختصات در راستای افقی و عمودی را نشان می‌دهد. از آنجا که در مرحله‌ی یکسان‌سازی تعداد نقاط نمونه درپیش‌پردازش همگی حروف با تعداد ثابتی نقطه بازنمایی شده‌اند (که این تعداد در کار ما ۲۱ در نظر گرفته شده است)، این ویژگی‌ها نیز که متاثر از شکل و طول حروف می‌باشند، می‌توانند یکی از پارامترهای ایجاد تمایز بین بدنه‌ی حروف در نظر گرفته شوند. در شکل ۵ یک نمونه از تاثیر این ویژگی بین دو نقطه‌ی متوالی بین حروف «ا» و «ح» مشاهده می‌شود.



شکل ۵ – ویژگی‌های (۳) و (۴) برای دو نقطه‌ی متوالی حروف «ا» و «ح»

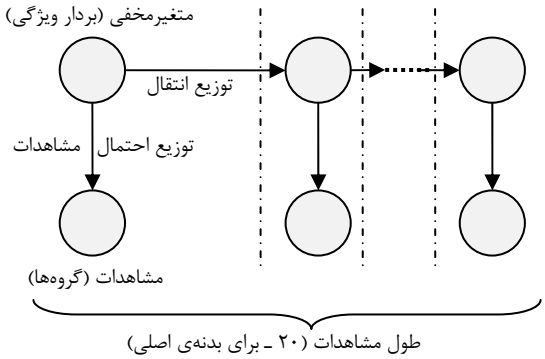
ویژگی (۵) و (۶) نیز برای بازشناسی بسیار مفید است. از آنجا که $\Delta y_i / \Delta x_i$، نسبت تغییرات عمودی به تغییرات افقی مختصات دو نقطه‌ی متوالی، شیب پاره‌خط میان آن دو را مشخص می‌کند، آرک تانژانت آن، زاویه‌ی این پاره‌خط را نسبت به محور افقی نشان می‌دهد. در نتیجه میزان انحراف این پاره‌خط از وضعیت افقی و عمودی را می‌توان با سینوس و کسینوس این زاویه نمایش داد.

مجموعه‌ی این شش ویژگی برای هر دو نقطه‌ی متوالی استخراج می‌شود و کل آنها به عنوان داده‌ای ترتیبی (۲۰ مرحله‌ای) به مدل مخفی مارکوف قابل ارائه می‌گردد.

۵-۱ طبقه‌بندی با مدل مخفی مارکوف

برای تعیین گروه اصلی بدنه‌ی هر حرف، و تعیین ریزحرکت‌ها، از مدل مخفی مارکوف (HMM) با پارامترهای زیر استفاده شده است:

- **احتمالات اولیه** (π_i): احتمالات اولیه برای هر مرحله، در ابتدا یک عدد تصادفی است که در خلال اجرای الگوریتم، با استفاده از الگوریتم k-means مقدار جدیدی پیدا می‌کند.
- **توزیع انتقال** (A) **و توزیع احتمال مشاهدات** (B): پارامترهای لازم برای این دو توزیع، بر اساس مجموعه‌ی آموزشی در فرآیند آموزش مشخص می‌شود و پس از آن برای استفاده از مدل ذخیره می‌شود. مقادیر اولیه برای این دو توزیع با استفاده از الگوریتم k-means تعیین شده است. توزیع احتمال مشاهدات، یک توزیع نرمال ترکیبی با دو مؤلفه است.
- **طول مشاهدات** (T): این پارامتر، تعداد نقاط یک مشاهده را نشان می‌دهد که مطابق توضیح ارائه شده در بخش پیش‌پردازش، برای بدنه‌ی اصلی حروف ثابت و برابر با ۲۰ و برای ریزحرکت‌ها برابر با ۱۰ می‌باشد.
- **تعداد مراحل مدل** (N): این پارامتر تعداد مراحل مدل مارکوف را نشان می‌دهد و برای هر حرف به صورت جداگانه مشخص شده است. این عدد بر اساس شکل هندسی، برای هر حرف متفاوت است. البته معیار دقیقی برای تعیین آن در نظر گرفته نشده است، اما به طور مثال برای حرف «ا» به خاطر سادگی شکل هندسی برابر با ۱۱ مرحله و برای حرفی مثل «ص» به دلیل پیچیدگی بیشتر برابر با ۱۶ مرحله در نظر گرفته شده است.
- **تعداد نمونه‌ها** (M): تعداد نمونه‌ها که در مرحله‌ی آموزش مدل استفاده می‌شود، برای هر گروه متفاوت بوده و بسته به فراوانی تکرار آنها در حرف الفبا متغیر می‌باشد. به عنوان مثال برای گروه ۳ که بدنه‌ی اصلی آنها «ح» می‌باشد، به خاطر اشتراک در چهار حرف «ج، چ، ح، خ» ۴۷۴ نمونه استخراج شده است، اما برای گروه ۱۸ که تنها شامل بدنه‌ی اصلی «ی» می‌باشد، تنها ۱۲۰ نمونه وجود دارد.
- **متغیرهای مخفی** (Q_i): متغیرهای مخفی برای HMM، همان بردار ویژگی شش‌تایی معرفی شده در زیربخش قبل است که شامل مختصات، تغییر مختصات نسبت به نقطه‌ی قبل و تغییرات زاویه نسبت به نقطه‌ی قبلی است.
- **مشاهدات** (V): مشاهدات در این مدل همان گروه‌ها هستند که براساس متغیرهای مخفی و یک توزیع احتمال مشخص می‌شوند. برای ارتباط مشاهدات با متغیرهای مخفی، نیز از یک توزیع نرمال ترکیبی با دو مؤلفه استفاده شده است که این عدد نیز صرفاً بر اساس یک آزمایش تجربی و بر مبنای آزمون و خطا انتخاب شده است تا نتیجه‌ی مطلوب حاصل شود.
- **نوع ارتباط میان مراحل:** برای ارتباط میان مراحل و حالت‌های مختلف مدل، الگوهای متعددی وجود دارد که در ساده‌ترین شکل آن، همه‌ی حالت‌ها با یکدیگر در ارتباط هستند (شکل ۶). در این کار نیز از این الگو استفاده شده است. انتخاب این نوع ارتباط نیز بر اساس یک آزمون و خطا برای یافتن الگوی مناسب انجام شده است.



شکل ۶ – مدل مخفی مارکوف و پارامترهای مورد استفاده در تعریف آن

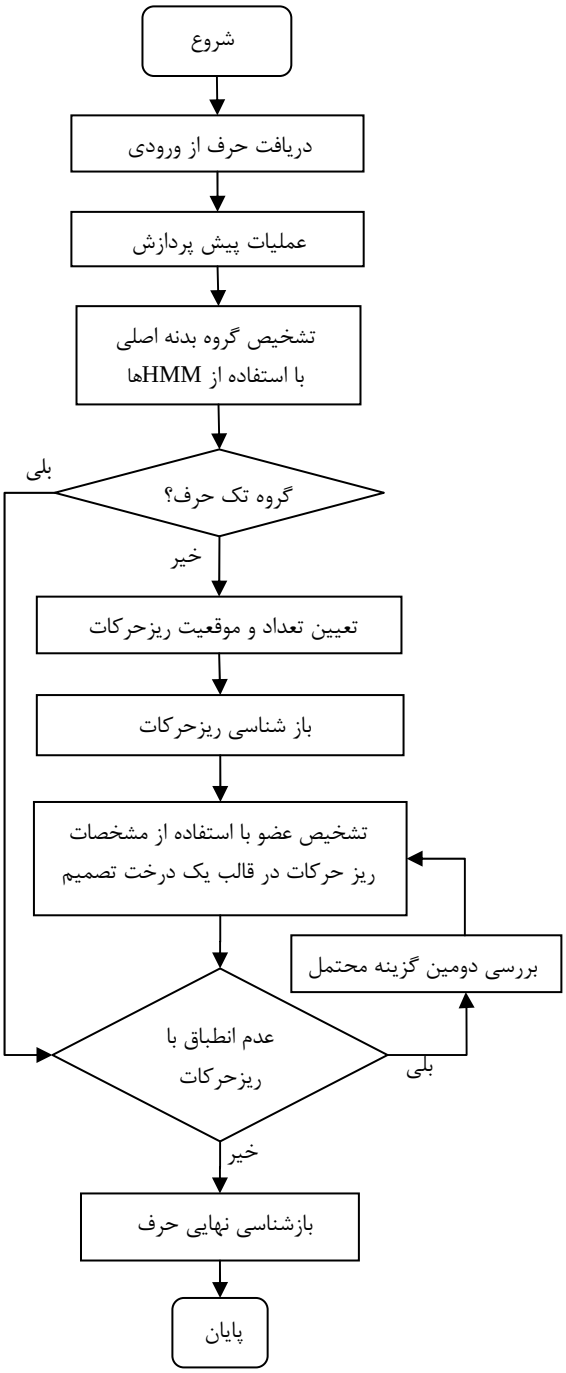
برای هر گروه از حروف، یک HMM با پارامترهای مربوط به خودش تهیه می‌شود و با استفاده از آن می‌توانیم گروه متناظر با هر حرف ورودی را تشخیص دهیم.

برای تعیین وضعیت ریزحرکت‌ها، موقعیت هر ریزحرکت با توجه به بدنه‌ی اصلی حرف و در سه دسته‌ی بالا، پایین و وسط تقسیم‌بندی می‌شود. نحوه‌ی دسته‌بندی به این صورت است نقطه‌ی مرکزی ارتفاع ریزحرکت استخراج می‌شود و با حداکثر ارتفاع بدنه مقایسه می‌شود. اگر از این حداکثر بالاتر بود، ریزحرکت، «بالا» در نظر گرفته می‌شود؛ در غیر این صورت با حداقل ارتفاع بدنه مقایسه می‌شود که اگر از این حداقل پایین‌تر بود، موقعیت «پایین» برای آن در نظر گرفته می‌شود و در مواردی غیر از دو مورد فوق، موقعیت «وسط» برای ریزحرکت فرض می‌شود؛ یعنی در حالتی که نقطه‌ی مرکزی ارتفاع ریزحرکت بین حداقل و حداکثر ارتفاع بدنه‌ی اصلی قرار داشته باشد. البته حالت «وسط» بسته به گروه ممکن است به یکی از دو حالت بالا و پایین تبدیل شود. برای مثال برای حرف «ج» وسط به معنی پایین و در حرف «ن» وسط به معنی بالا خواهد بود.

۶- پس‌پردازش

پس از تشخیص و بازشناسی گروه بدنه‌ی اصلی، نتایج همراه با اطلاعات مربوط به ریزحرکت‌ها، به یک درخت تصمیم داده می‌شود تا با ترکیب اطلاعات این دو بخش بررسی کند که حرف ناشناخته، متعلق به کدام‌یک از اعضای گروه‌های هجده‌گانه است و بدین ترتیب فرآیند

بازشناسی کامل حرف کامل شود. بدیهی است که این مرحله برای گروه‌های تک حرف ضرورتی ندارد.



شکل ۷ – فلوچارت الگوریتم بازشناسی حرف

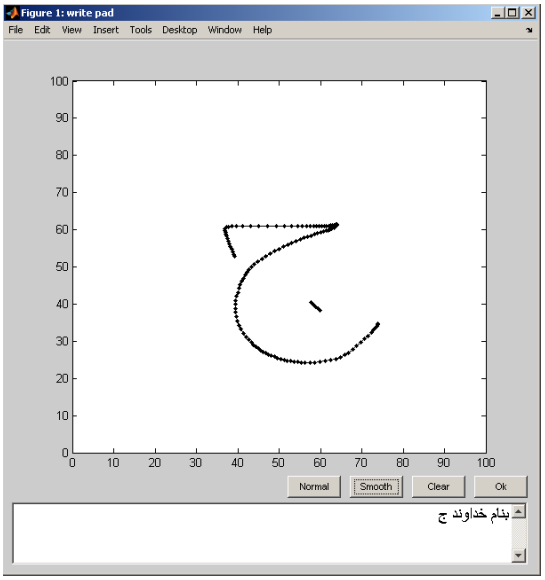
از آنجا که بازشناسی و تفکیک ریزحرکت‌ها مانند نقطه‌ها، سرکش‌ها و . . . نسبتاً دشوار است و اغلب به لحاظ ظاهری نیز با هم تداخل دارند، سعی ما بر این بوده است که از این اطلاعات در آخرین مراحل تصمیم‌گیری استفاده شود. ار این رو در ساختار درخت تصمیم، اولویت تصمیم‌گیری با تعداد و محل ریزحرکت‌هاست، نه شکل واقعی آنها. به عنوان نمونه زمانی که بدنه‌ی اصلی حرفی در قالب گروه ۱، «ا» تشخیص داده می‌شود، تنها وجود یک زیرحرکت در بالای آن کافی است تا سیستم آن را «آ» تشخیص دهد و دیگر نیازی به تشخیص شکل علامت «مد» نمی‌باشد.

موضوع دیگری که در این قسمت لحاظ شده است، این است که گاهی به خاطر عدم تطابق بین ریزحرکت‌ها و بدنه‌ی اصلی، سیستم به تشخیص خطا در بازشناسی بدنه قادر می‌شود و می‌تواند بدنه‌ی تشخیص داده شده را تصحیح کند و آن را به بدنه‌ی محتمل دیگری تبدیل نماید. برای مثال ممکن است سیستم بدنه‌ی «ل» را به همراه یک ریزحرکت در بالای آن تشخیص دهد. در چنین حالتی سیستم با توجه عدم تطابق بین آنچه تشخیص داده شده و آنچه برای حرف «ل» مورد انتظار بوده است، تشخیص بدنه‌ی خود را به «ن» تغییر می‌دهد. به عبارت دیگر بازنگری نهایی بر اساس مشخصات بسیار ساده‌ی ریزحرکت‌ها صورت می‌گیرد.

شکل ۷، فلوچارت الگوریتم بازشناسی شامل کلیه‌ی مراحل توضیح داده شده را در یک نما نشان می‌دهد.

۷- پیاده‌سازی و نتایج تجربی

برای پیاده‌سازی برنامه‌ی بازشناسی، از نرم‌افزار MATLAB استفاده شده است. برای استفاده از HMM نیز یک جعبه ابزار آماده‌ی تحت MATLAB که توسط دانشگاه MIT بر روی اینترنت منتشر شده است، به کار گرفته شده است. شکل ۸ نمایی از این نرم‌افزار را نشان می‌دهد. همان طور که ملاحظه می‌شود، در این نرم‌افزار باید هر حرف را بر روی صفحه‌ی سفید موجود نوشت و سیستم نتیجه‌ی بازشناسی را در کنار حروف قبلی در کادر متنی زیرین نمایش می‌دهد.



شکل ۸ – نمایی از نرم‌افزار بازشناسی برخط حروف فارسی در MATLAB

مجموعه داده‌ی Online-TMU که توسط بخش مهندسی برق دانشگاه تربیت مدرس تهیه شده است [۴]، به عنوان مجموعه‌ی داده‌ی آموزشی و آزمایشی مورد استفاده قرار گرفته است که البته با توجه به مجزا بودن حروف در سیستم مورد بررسی ما، تنها بخش حروف مجزای آن استفاده شده است. در این مجموعه داده، برای هر حرف به طور میانگین حدود ۱۲۰ نمونه جمع‌آوری شده است. یکی از مشکلاتی که در استفاده از این پایگاه وجود داشت، عدم تفکیک بخش‌های مختلف یک حرف مانند بدنه و نقطه‌ها بود. در واقع اطلاعات حرف صرفاً یک آرایه از مختصات x و y بود که نقطه‌ی شروع و پایان هر حرکت قلم در آن مشخص نشده بود. از این جهت برای استفاده از آنها لازم بود، حرکت‌های مختلف در این داده‌ها جداسازی شود. بدین منظور، ابتدا میانگینی از فواصل هر دو نقطه‌ی متوالی در حروف گرفته شد و سپس فرض شد که اگر بین دو نقطه‌ی متوالی سه برابر این میانگین فاصله افتاده باشد، نقطه‌ی اول پایان حرکت قبلی و نقطه‌ی دوم شروع حرکت بعدی خواهد بود. این عملیات با دقت بیش از ۹۰ درصد یک مجموعه داده‌ی جدید شامل بدنه‌ی حروف و زیرحرکت‌ها را ایجاد کرد. طبیعتاً تعداد نمونه‌های بسته به میزان اشتراک حروف در بدنه‌ی اصلی شده است. مثلاً از حروف «ج، چ، ح، خ» تا ۴۷۴ بدنه به صورت «ح» استخراج شده است و تعداد نمونه‌های بدنه‌ی اصلی برای حرف «ی» که منحصر بفرد بوده است، به بیش از ۱۲۰ نمونه نمی‌رسد.

در مرحله‌ی آموزش از ۷۰ درصد نمونه‌هایی که از هر بدنه در اختیار بود، استفاده شد و مدل‌های HMM آموزش داده شدند و پارامترهای آنها ذخیره شد. برای مرحله‌ی آزمون، هر گروه به صورت جداگانه با ۳۰ درصد باقی‌مانده از نمونه‌ها مورد آزمایش قرار گرفتند. نتیجه‌ی مرحله‌ی آزمون در ماتریس پراکنش موجود در جدول ۲ نشان داده شده است. همان طور که مشاهده می‌شود، بازشناسی با میانگینی برابر با ۹۳٪ درست صورت می‌گیرد. البته در مورد گروه‌هایی چون گروه ۱۲ این تشخیص به دلیل شباهت زیاد بدنه‌ی اصلی حروف واقع در آن با گروه‌های دیگر، تداخل بیشتری مشاهده می‌شود. برای بهبود شرایط می‌توان چنین گروه‌هایی را با هم ادغام نمود و تشخیص بدنه‌ی داخل آنها را بر اساس مشخصات دیگر انجام داد. از این رو دو گروه ۱۲ و ۲ (یعنی بدنه‌های «ک» و «ب») با یکدیگر ادغام خواهند شد.

نرخ بازشناسی کلی سیستم برای حروف پس انجام پس‌پردازش به حدود ۹۶٪ درصد می‌رسد. جدول ۳ نرخ بازشناسی کلی این سیستم را با دو سیستم دیگر که آنها نیز برای ارزیابی از مجموعه داده‌ی Online-TMU استفاده کرده‌اند نشان می‌دهد. مشاهده می‌شود که در صورت یکسان بودن شرایط آزمایش‌ها، سیستم پیشنهادی ما نرخ بازشناسی صحیح بالاتری را ارائه می‌کند.

جدول ۳- مقایسه‌ی نرخ بازشناسی صحیح این سیستم با دو کار دیگر بر اساس مجموعه داده‌ی Online-TMU

سیستم ارائه شده در [۱]	٪۹۰/۲۷
سیستم ارائه شده در [۲]	٪۹۳/۳
سیستم پیشنهادی ما	٪۹۶

جدول۲- ماتریس پراکنش (Confusion) برای بازشناسی گروه‌ها

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	%
1	71	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.99
2	0	125	0	0	0	1	0	0	0	5	0	12	0	0	1	0	0	0	0.87
3	0	0	141	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0.99
4	0	0	0	69	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0.96
5	1	1	0	2	102	0	0	0	0	1	0	2	0	0	0	0	0	0	0.94
6	0	0	0	0	0	71	0	0	0	0	0	0	0	0	0	0	0	1	0.99
7	0	0	0	0	0	1	69	0	0	1	1	0	0	0	0	0	0	0	0.96
8	0	0	0	0	0	0	1	64	0	0	0	0	0	0	0	1	0	0	0.97
9	0	0	1	0	0	0	0	1	69	0	0	0	0	0	0	0	0	0	0.97
10	0	1	0	0	0	0	1	0	0	31	2	1	0	0	0	0	0	0	0.86
11	0	0	0	0	0	1	2	0	0	0	32	0	0	0	1	0	0	0	0.89
12	0	13	0	1	1	1	0	0	0	0	0	51	0	0	4	0	0	0	0.72
13	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0	0	1	0	0.97
14	0	0	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	1.00
15	0	0	0	0	0	0	0	0	0	0	1	1	2	0	30	0	1	1	0.83
16	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	35	1	0	0.95
17	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	33	0	0.94
18	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	36	0.97
میانگین																			0.93

۸ - خلاصه، بحث و نتیجه‌گیری

در این مقاله، روشی ساده برای بازشناسی برخط حروف فارسی ارائه شده است که مبنای آن، بازشناسی بدنه‌ی اصلی حروف با استفاده از HMM می‌باشد. بدین ترتیب که بدنه‌ی اصلی حرف با استفاده از مدل مخفی مارکوف شناسایی می‌شود و سپس با استفاده از ریزحرکات (قسمت‌های فرعی حروف شامل مد، سرکش و نقاط)، حرف به طور کامل بازشناسی می‌شود. مراحل پیش‌پردازش و پس‌پردازش در رسیدن به نتیجه‌ی مطلوب نقش مهمی را بازی می‌کنند.

آنچه بر مبنای نتایج تجربی به نظر می‌رسد، این است که تشخیص بدنه‌ی اصلی حروف با روش HMM نتایج مطلوبی را در بر دارد، اما نهایتاً این ریزحرکت‌ها هستند که شکل نهایی حروف را مشخص می‌کنند. بنابراین، هر قدر هم که سیستم در تشخیص بدنه‌ی حروف دقیق عمل کرده باشد، بدون بازشناسی صحیح ریزحرکات، نتایج مطلوب نخواهد بود. البته این مشکل در مورد گروه‌هایی که اختلاف اعضای آنها در نقاط است، مانند گروه‌های «ب، پ، ت، ث» و «ج، چ، ح، خ» بیشتر مطرح است. اما با وجود مشکلات ذکر شده در تشخیص ریزحرکت‌ها احتمالاً بهتر است که علاوه بر تشخیص اولیه، از یک درخت تصمیم پیچیده‌تر نیز استفاده شود. به عنوان نمونه هنگام تشخیص بدنه‌ی حرف «ک» با ریزحرکت «دو نقطه» سیستم با توجه به میزان اعتبار هر یک از این دو تشخیص تصمیم بگیرد که حرف نهایی «ک» یا «ت» بوده است. بنا به دلایلی چون تعدد شکل ریزحرکت‌ها (مانند نوشتن سه نقطه در قالب یک، دو یا سه حرکت) و عدم انحصار در شکل (مانند شباهت مد و دو نقطه) استفاده از HMM برای بازشناسی آنها چندان کارامد به نظر نمی‌رسد.

مراجع

[۱] سلیمانی باغشاه، م.، باقری شورکی، س.، کسائی، ش.، "**بازشناسی و یادگیری حروف مجزای برخط فارسی به روش فازی**"، مجموعه مقالات چهاردهمین کنفرانس مهندسی برق ایران، دانشگاه صنعتی امیرکبیر، تهران، ۱۳۸۵.

[۲] رضوی، س. م.، کبیر، ا.، "**بازشناسی برخط حروف مجزای فارسی**"، مجموعه مقالات ششمین کنفرانس سیستم‌های هوشمند، دانشگاه شهید باهنر، کرمان، آذر ۱۳۸۳.

[۳] رضوی، س. م.، کبیر، ا.، **روشی ساده برای بازشناسی برخط زیر-کلمات فارسی**، نشریه‌ی مهندسی برق و مهندسی کامپیوتر ایران، سال ۳، شماره ۲، صص. ۶۳-۷۲، ۱۳۸۴.

[۴] رضوی، س. م.، کبیر، ا.، "**یک پایگاه داده برای بازشناسی دست‌نوشته‌های برخط فارسی**"، مجموعه مقالات ششمین کنفرانس سیستم‌های هوشمند، دانشگاه شهید باهنر، کرمان، آذر ۱۳۸۳.

[1] El-Wakil, S. A., "On-line Recognition of Handwritten Isolated Arabic Characters", *Pattern Recognition*, Vol. 22, 1989.

[2] Alimi, A. M., "Neuro-Fuzzy Approach to Recognize Arabic Handwritten Characters", In Proceedings of *IEEE International Conference on Neural Networks*, Piscataway, NJ, United States, pp. 1397-1400, 1997.

[3] Bouslama, F., "Structural and Fuzzy Techniques in the Recognition of Online Arabic Characters", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 13, pp. 1027-1040, 1999.

[4] Mezghani, N., Mitiche, A., Cheriet, M., "A New Representation of Character Shape and Its Use in On-line Character Recognition by a Self Organizing Map", In Proceedings of *IEEE Conference on Image Processing, ICIP*, pp. 2123-2126, 2004.

[5] Mezghani, N., Mitiche, A., Cheriet, M., "A New Representation of Shape and Its Use for High Performance in On-line Arabic Character Recognition by an Associative Memory", *International Journal on Document Analysis and Recognition*, Vol. 7, pp. 201-210, 2005.

[6] Klassen, T. J., Heywood, M. I., "Towards the on-line recognition of Arabic characters", In Proceedings of *International Joint Conference on Neural Networks*, Honolulu, HI, pp. 1900-1905, 2002.

[7] Nourouzian, E., Mezghani, N., Mitiche, A., Johnston, R., "Online Persian/Arabic Character Recognition by Polynomial Representation and a Kohonen Network", In Proceedings of *IEEE International conference on pattern recognition*, ICPR, Singapour, 2006.

[8] Al-Ghoneim, A., "Online Arabic character recognition for handheld devices", In Proceedings of *International Conference on Image Processing, Computer Vision, and Pattern Recognition*, IPCV, pp. 15-22, 2008.

[9] Mezghani, N., Mitiche, A., "A Gibbsian Kohonen Network for Online Arabic Character Recognition", *Lecture Notes in Computer Science*, LCNS 5359, pp. 493-500, 2008.

[10] Kherallah, A., Alimi, A. M. "New Lecture Support Based on On-line Arabic Handwriting Recognition", In Proceedings of *International Conference on Innovations in Information Technology*, IIT, pp. 673-677, 2008.

[11] Baghshah, M. S., Shouraki, S. B., Kasaei, S., "A Novel Fuzzy Approach to Recognition of Online Persian Handwriting", In Proceedings of *International Conference on Intelligent Systems Design and Applications*, ISDA, pp. 228-273, 2005.

[12] Halavati, R., Shouraki, S.B., "Recognition of Persian Online Handwriting Using Elastic Fuzzy Pattern Recognition", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 21, No. 3, pp. 491-513, 2007.

[13] Biadsy, F., El-Sana, J., Habash, N., "Online Arabic handwriting recognition using Hidden Markov Models", In Proceedings of 10th *International Workshop on Frontiers in Handwriting Recognition*, IWFHR, France, 2006.

[14] Al-Habian, G., Assaleh, K., "Online Arabic Handwriting Recognition Using Continuous Gaussian Mixture HMMS", In Proceedings of *International Conference on Intelligent and Advanced Systems*, 2007.

[15] Faradji, F., Faez, K., Mousavi, M. H., "An HMM-based online recognition system for farsi handwritten words", In Proceedings of *International Conference on Intelligent and Advanced Systems*, ICIAS, pp. 1187-1192, 2007.

[16] Biem, A., "Minimum classification error training for online handwriting recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 7, 2006.