دانشگاه صنعتی امیرکبیر

(پلی تکنیک تهران)

دانشکده ریاضی و علوم کامپیوتر

پایان نامه کارشناسی ارشد

رشته علوم کامپیوتر

گرایش هوش مصنوعی

مطالعه و بررسی روشهای بازیابی تصاویر کلمات دستنویس

نگارش

سمانه السادات شیرازی

استاد راهنما: دکترمحمد ابراهیم شیری

استاد مشاور: دکتر رضا عزمی

بهمن ۱۳۸۵

تقدیم به پدر و مادر عزیزم که همواره مشوق من بوده اند.

قدردانی:

با استعانت از پروردگار توانا و با یاری جستن از اساتید بزرگوارم، جناب دکتر محمد ابراهیم شیری و دکتر رضا عزمی، توانستم این پایان نامه را به پایان برسانم. از دکتر شیری، به خاطرصبرو حوصله شان در پیشرفت پایان نامه و راهنمایی‌هایشان و از دکترعزمی، به خاطر ایده‌ها، خلاقیت‌ها و کمک‌هایشان که در تمام مراحل تکمیل ساختار پایان نامه، یاری‌ام کرده اند؛ سپاسگزارم.

چکیده:

در سالهای اخیر شناسایی نوری نویسه به عنوان یکی از حوزه‌های فعال در مبحث شناسایی الگو است. اگرچه تحقیقات بر روی کلمات دستنویس فارسی (عربی) در سالهای گذشته پیشرفت بسزایی کرده است، اما در مقایسه با کلمات لاتین جای پیشرفت دارد. البته کارهای انجام شده در این زمینه حاوی اطلاعات مفید برای بالا بردن ضریب اطمینان است.

در این پایان نامه سعی بر این است که در حوزه شناسایی نوری نویسه و به خصوص کلمات دستنویس بتوان روشی برای بهبود دقت در شناسایی و بازیابی کلمات ارائه کرد. با توجه به این حقیقت که شناسایی یک تصویر به وسیله انسان‌ها بر اساس مشاهده صورت می‌گیرد، تمایز تصاویر گوناگون وابسته به قدرت بینایی انسان است. تحقیقات نشان داده است که درصد تشخیص و شناسایی یک شی در کامپیوتر با استفاده از تبدیل آن به فرکانس، برتر از شناسایی بر اساس ویژگی‌های ساختاری و ظاهری تصویر شئ می‌باشد. به همین منظور و در راستای شناسایی و بازیابی تصاویر دستنویس، از تبدیل موجک گسسته استفاده کرده و ویژگی‌های بدست آمده از این تبدیل را به عنوان معیار شناسایی و در نهایت بازیابی قرار داده ایم. سیستم شناسایی بر اساس شبکه‌های عصبی برنامه ریزی شده است. علت استفاده از شبکه عصبی، داشتن قابلیت دسته‌بندی کردن و تفکیک الگوها بر اساس ویژگی‌های مدنظر می‌باشد.

مجموعه داده‌ها شامل ۴۲ کلمه می‌باشد که توسط ۱۰۰ نفر با درجه تحصیلات متفاوت جمع‌آوری شده است. نیمی از داده‌ها را به عنوان داده‌های آموزش و نیمی دیگر را به عنوان داده‌های آزمایشی در نظر گرفته‌ایم. سیستم شناسایی، داده‌های آموزشی را با دقت ۹۵،۷۱٪ و دادهای آزمایشی را با دقت ۷۲،۸۵٪ شناسایی می‌کند. در بخش بازیابی تصاویر در مجموعه داده‌ها، سیستم پیشنهادی بطور متوسط برای ۴۲ کلمه مفروض با دقت ۹۷،۸۰٪ به بازیابی می‌پردازد. به علت اینکه در داده‌های جمع‌آوری شده یک کلمه را به صورت پیوسته در نظر گرفته‌ایم و به شناسایی حروف به طور مجزا پرداخته نشده است، نتایج بدست آمده از آزمایش سیستم پیشنهادی نشان دهنده توانایی سیستم در بازیابی و همچنین شناسایی برون خط کلمات دستنویس فارسی دارد.

کلمات کلیدی:

بازیابی تصاویر، شناسایی نوری نویسه، شناسایی برون خط کلمات دستنویس، موجک، شبکه عصبی

فهرست مندرجات

فصل اول: مقدمه

فصل دوم: مروری بر کارهای پیشین

فصل سوم: مرز تصویر یک شئ

فصل چهارم: تبدیل موجک

فصل پنجم: شبکه‌های عصبی

فصل ششم: ساختار سیستم پیشنهادی

فصل هفتم: نتیجه گیری و پیشنهادات

فصل اول : مقدمه

۱-۱- مقدمه

شناسایی الگوهای تصویری یکی از شاخه‌های مهم شناسایی الگو[۱] است. امضاء، حروف، اثر انگشت، اشعه X نمونه‌هایی از این الگوها هستند. این الگوها معمولا بصورت تصاویر با سطوح خاکستری[۲] یا دوسطحی[۳] (دودویی) قابل نمایش اند [۱]. یکی از مباحث جالب و قابل توجه در زمینه شناسایی الگو، شناسایی نوری نویسه[۴] می‌باشد.

سیستم‌های شناسایی نوری نویسه با حذف نقش تایپیست‌ها در فرآیند تبدیل اسناد کاغذی به قالب الکترونیکی، سرعت ورود اطلاعات به رایانه را ده‌ها برابر افزایش می‌دهند و روند انجام این فرآیند را به میزان قابل توجهی تسهیل می‌کنند. امروزه بازار مصرف سیستم‌های بازشناسی نوری نویسه، طیف بسیار وسیعی از مؤسسات (شامل مراکز نشر، دانشگاه‌ها، کتابخانه‌ها، بانک‌ها، ادارات پستی، شرکت‌های بیمه، و ...) را دربرمی‌گیرد.

در چند دهه گذشته بازشناسی الگوهای نوشتاری شامل حروف، ارقام و دیگر نمادهای متداول در اسناد نوشته‌شده به زبان‌های مختلف، توسط گروه‌های مختلفی از محققین مورد مطالعه و بررسی قرار گرفته است. نتیجه این تحقیقات منجر به پیدایش مجموعه‌ای از روش‌های سریع و تا حد زیادی مطمئن موسوم به «شناسایی نوری حروف» به منظور وارد نمودن اطلاعات موجود در اسناد، مدارک، کتاب‌ها و سایر مکتوبات تایپی و حتی دست‌نوشت به داخل رایانه شده است [۱۱].

---

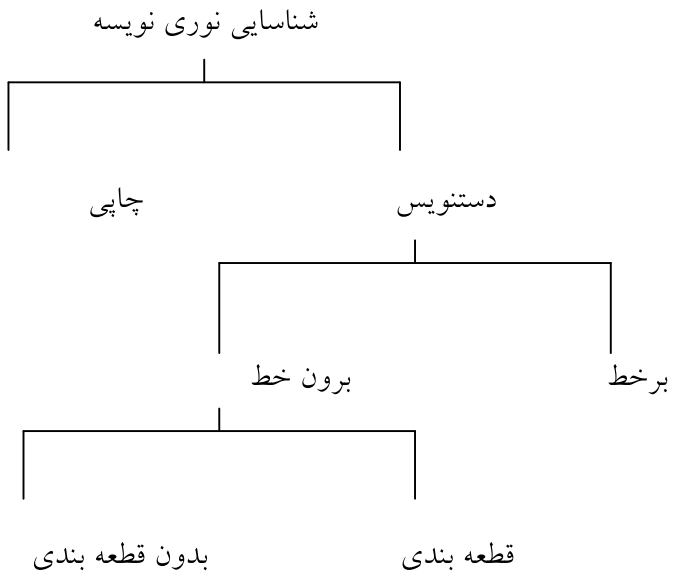
۱ Pattern Recognition
۲ Gray Level
۳ Binary
۴ Optical Character Recognition - OCR

استفاده از سیستم‌های بازشناسی نوری نویسه دارای دو مزیت عمده می‌باشد:

۱. افزایش چشمگیر سرعت دسترسی به اطلاعات. در متن بر خلاف تصویر، امکان جستجو و ویرایش وجود دارد.

۲. کاهش فضای ذخیره سازی. حجم فایل متنی استخراج شده از یک تصویر، معمولاً بسیار کمتر از حجم خود فایل تصویری است.

چنین قابلیتی امکان استفاده گسترده از رایانه را در پردازش سریع حجم وسیعی از داده‌های مکتوب شرکت‌ها و مؤسسات مختلف (نظیر بانک‌ها، شرکت‌های بیمه، مؤسسات خدمات عمومی، اداره پست، و دیگر نهادهایی که سالانه با میلیون‌ها مورد پرداخت، دریافت و حسابرسی امور مشتریان خود مواجه‌اند) را فراهم می‌آورد [۱۲].



الگوهای ورودی در سیستم‌های شناسایی نوری نویسه به دو صورت دستنوشته و چاپی می‌باشند. به علت عدم وجود اعوجاج[۱] در مسیر حرکت نویسه‌های چاپی، و همچنین وجود الگوهای مرجع مشخص برای این نوع از داده‌ها، سیستم‌های شناسایی الگوهای چاپی دارای بازدهی خوبی می‌باشند. اما در زمینه متون دستنویس، به علت وابستگی ساختار نویسه به نویسنده و عدم وجود الگویی واحد یا الگوهایی مشخص در این زمینه، نتایج بدست آمده بخصوص در زمینه حروف فارسی و عربی راضی کننده نمی‌باشد و امکان پیشرفت بسیاری وجود دارد.

---

[۱] Distortion

سیستم‌های شناسایی متون دستنویس را می‌توان از لحاظ نوع الگوی ورودی به دو گروه بر خط[1] و برون خط[2] تقسیم کرد.

۱-۲- شناسایی برخط نویسه‌ها

داده‌ها در اینگونه سیستم‌ها توسط حرکت یک قلم الکترونیکی مخصوص، به عنوان وسیله نوشتاری، بر روی یک صفحه الکترونیکی به دست می‌آید. به علت اینکه این الگوها به صورت همزمان[3] به سیستم داده می‌شوند، اطلاعاتی نظیر فشار قلم، مسیر حرکت و ... علاوه بر اطلاعات تصویر قابل استفاده است؛ و به همین علت شناسایی به صورت بر خط را نسبت به برون خط آسان‌تر می‌کند.

۱-۳- شناسایی برون خط نویسه‌ها

داده‌ها در اینگونه سیستم‌ها به وسیله دوربین دیجیتالی یا اسکنر، به کامپیوتر منتقل می‌شوند و در واقع پیکسلهای مکانی الگوها بدون داشتن اطلاعات زمانی مربوطه، به سیستم برای تشخیص داده می‌شوند. شناسایی برون خط الگوها در مقایسه با شناسایی بر خط، حوزه کاربردی وسیع تری دارند. تشخیص حروف و ارقام در کتب دستنوشته قدیمی در کتابخانه ملی، آدرس‌های پستی، چکهای بانکی و غیره، که این نوع داده‌ها به صورت برون خط موجود می‌باشند.

به طور کلی سه نوع دید برای بررسی متون دستنویس وجود می‌باشد [۱۴].

۱. دانش مورفولوژی[4] (ریخت شناسی): شکل ایده آل حروف را مستقل از نویسنده، در نظر گرفته و به عنوان مرجع در نظر می‌گیریم.

۲. دانش زبان شناسی[5]: زبان خاصی را در نظر گرفته و جنبه‌های گرامری، لغوی و معنایی این زبان را به عنون مرجع در نظر می‌گیریم. این نوع نگرش بسیار موثر می‌باشد. زیرا حتی انسان هم اگر به مفهوم یک متن توجه نکند ممکن است نتواند بعضی از لغات را به صورت صحیح بخواند.

---

[1] On-line
[2] Off-line
[3] Real Time
[4] Morphology knowledge
[5] Linguistic knowledge

۳. دانش واقع گرایانه یا عملی[۱]: که در واقع شاخه‌ای از دانش زبان شناسی می‌باشد. در این نوع روش، ترتیب قرار گرفتن یک حرف را با توجه به ارتباط آن حرف به کلمه و یا عبارت، به عنوان مرجع شناخت قرار می‌دهد.

دو روش انتهایی را می‌توان به عنوان دانش‌های مفهومی[۲] قلمداد کنیم. کاربرد دانش مفهومی را می‌توان خواندن و شناسایی و تصحیح اشتباهات موجود در یک متن را به کمک یک فرهنگ لغات، نام برد. برای مثال، کدپستی‌ای[۳] که به صورت دستنویس نوشته شده است، را می توان با استفاده از اطلاعات موجود در آدرس تصحیح کرد. علاوه بر این دانش مفهومی می‌تواند برای تکمیل یک کلمه در زمانی که تصاویر متون دستنویس به وسیله حذف یک حرف یا وجود نویز یا اعوجاج خراب شده است، مورد استفاده قرار بگیرد و سعی در شناسایی کلمه به طور مناسب[۴] داشته باشد.

۱-۴- روشهای شناسایی برون خط کلمات

دو روش کلی برای شناسایی برون خط کلمات وجود دارد: روشهای مبتنی بر قطعه بندی و روشهای بدون قطعه بندی.

۱-۴-۱- روشهای مبتنی بر قطعه بندی

قطعه بندی مرحله‌ای بسیار مهم برای سیستم‌های شناسایی نوری نویسه، مخصوصاً برای نویسه‌هایی مانند لاتین پیوسته، فارسی و عربی (که حروف کلمات به صورت سرهم نوشته می‌شوند) می‌باشد [۴۰].

در اینگونه روشها تصویر کلمات پس از نرمال شدن و رفع نویز، به مجموعه‌ای حروف یا زیر کلمه (شبه کلمه) تقسیم می‌شوند. که روشهای زیادی برای چگونگی این تقسیم ارائه شده است.

این مرحله یکی از مراحل بسیار با اهمیت در سیستم‌های شناسایی نوری نویسه مبتنی بر قطعه بندی است. نتایج حاصل از این مرحله، مستقیماً بر روی کیفیت مرحله بازشناسی اثر می‌گذارد.

---

[۱] Pragmatic knowledge
[۲] Contextual knowledge
[۳] ZIP Code
[۴] proper

اگر قطعات جداسازی شده حروف باشند، مرحله بعدی شناسایی حروف است. ولی اگر قطعات جداسازی شده زیر حروف یا عناصر پایه باشند، مرحله بعدی معمولاً شامل شناسایی این عناصر پایه و ترکیب نتایج شناسایی در قالب روشهای آماری یا ساختاری برای تشخیص حروف و سپس شناسایی کلمه است [۱].

در این دسته روشها مسئله مهم، چگونگی قطعه بندی کردن کلمات می‌باشد به طوریکه شکستگی اعمال شده باعث از بین رفتن یک حرف نشود و یا به گونه‌ای قطعه بندی صورت گیرد که یک حرف به طور کامل از کلمه جدا شود. به علت مشکل بودن طریقه قطعه بندی در کلمات، گاهی اوقات سعی بر این است که کلمات را به شبه کلمه‌ها تقسیم کنیم. البته این روش نیز مشکلات خاص خودش را دارا می‌باشد (از قبیل تعداد شبه کلمات یا چگونگی قطعه بندی به منظور ساخت الگوهای مرجع).

۱-۴-۲- روشهای بدون قطعه بندی

کلماتی که در این دسته روشها استفاده می‌شوند به صورت کامل و بدون اعمال قطعه‌بندی استفاده می‌شوند. به طور کلی در این روش کلمات با مقایسه مجموعه کلمات مرجع و پیدا کردن بهترین تطبیق الگو، شناسایی می‌شوند. به این علت که برای تطبیق الگو، هر الگوی ورودی با الگوهای مرجع مقایسه می‌شود؛ این روش برای مجموعه داده زیاد مناسب نمی‌باشد. البته این روش دارای سرعت محاسباتی بالایی نسبت به روش قطعه بندی است اما تنها برای مجموعه داده محدود، از نظر حجم محاسباتی قابل اجرا می‌باشد.

۱-۵- روش پیشنهادی در این پایان نامه

همانطور که گفته شد، به علت کاربرد فراوان شناسایی نوری نویسه، در زندگی روزمره، سالهای اخیر تحقیقات بسیاری در این زمینه صورت گرفته است. متاسفانه تحقیقات و مطالعات انجام شده در زمینه دستنوشته‌های فارسی و عربی (به خصوص در حوزه کلمات) بسیار نوپا بوده و جای کار بسیاری دارد [۱و۴و۸و۱۱]. در این پایان نامه ابتدا یک سیستم، به منظور شناسایی تصاویر کلمات دستنویس فارسی ارائه شده است و سپس به بررسی تئوری تصدیق[۱] می‌پردازیم و در پایان مسئله بازیابی کلمات را مورد بررسی قرار می‌دهیم و ایده‌ای برای برآورد دقت بازیابی پیشنهاد می‌کنیم. روش پیشنهادی، روش بدون قطعه بندی می‌باشد به عبارت دیگر به شناسایی و بازیابی یک کلمه به

---

[۱] Verification Theory

طر پیوسته می‌پردازیم. داده‌ها در اندازه ۱۲۸×۱۲۸ پیکسل و با دقت $^{\text{dpi}}$۳۰۰ و به صورت خاکستری اسکن شده و مورد استفاده قرار می‌گیرند. فرض بر این است که الگوهای ورودی، تصویر یک کلمه باشند و خطوط زمینه نیز وجود نداشته باشد.

به منظور شناسایی تصاویر کلمات، مرز[۱] تصویر هر کلمه را پس از مرحله پیش پردازش، استخراج کرده و نقاط استخراج شده را بر اساس پیکسلهای مکانی شان، به صورت زوج (x_k, y_k) ذخیره می‌کنیم. مجموعه پیکسلهای مرزی را به صورت دو مجموعه مجزای $\{x_k\}, \{y_k\}$ در نظر گرفته و به سیگنال تبدیل می‌کنیم. ضرایب بدست آمده از سیگنال‌های خروجی را به عنوان ویژگی هر تصویر در نظر می‌گیریم. ویژگی‌های بدست آمده را به منظور تکمیل عمل شناسایی، به یک نوع شبکه عصبی چند لایه پرسپترون[۲] به نام شبکه عصبی پیش آبشار[۳] می‌دهیم. ساختار این شبکه در برنامه متلب[۴] موجود می‌باشد. شبکه را توسط یکی از الگوریتم‌های پس انتشار خطا آموزش می‌دهیم. نتایج بدست آمده حاکی از قدرت بالای موجک‌ها در شناسایی الگو است.

ساختار سیستم بازیابی کلمات، مبتنی بر سیستم شناسایی عمل می کند. به منظور بازیابی یک کلمه خاص، در مجموعه داده‌ها به جستجو می‌پردازیم. ابتدا مرز تصویر هر کلمه را استخراج کرده و سپس مجموعه پیکسلهای مرزی را به صورت دو مجموعه مجزا به سیگنال تبدیل می‌کنیم. ضرایب بدست آمده از این تبدیل را به عنوان خصوصیات هر تصویر در نظر می‌گیریم. در پایان از یک شبکه عصبی به منظور جستجوی کلمه مورد نظر در تصاویر کلمات مورد جستجو استفاده می‌کنیم. تنها تفاوت سیستم بازیابی ارائه شده با سیستم شناسایی پیشنهادی، در مولفه‌های استفاده شده در شبکه عصبی می‌باشد.

داده‌های استفاده شده در این سیستم، شامل ۴۲ کلمه می‌باشند که از هر کلمه ۱۰۰ نمونه تهیه شده است. نمونه‌های بدست آمده، توسط افراد مختلف و با درجه تحصیلات متفاوت نوشته شده است. در مجموع از ۴۲۰۰ نمونه بدست آمده، ۲۱۰۰ نمونه را به عنوان نمونه‌های آموزشی و ۲۱۰۰ نمونه را به عنوان نمونه‌های آزمایش در نظر می‌گیریم.

۱-۶- ساختار نوشتاری پایان نامه

---

[۱] Contour
[۲] Multi-Layer Perceptron (MLP)
[۳] Cascade-Forward
[۴] Matlab

در فصل دوم، مختصری از کارهای انجام شده در حوزه شناسایی نوری نویسه را بررسی می‌کنیم. کارهای انجام شده در این حوزه شامل سه مرحله پیش پردازش، پردازش و پس پردازش می‌باشد که برخی از سیستم‌های پیشنهادی در این فصل تنها به یکی از این سه مرحله پرداخته‌اند.

در فصل سوم، مرز تصویر و راههای متداول استخراج مرز را بررسی کرده و در پایان روش استفاده شده در این پایان نامه را مورد بررسی قرار داده ایم.

برای بدست آوردن ویژگی در الگوهای تصویری، از ضرایب موجک‌ها[1] استفاده کردیم. در فصل چهارم، مختصر توضیحی در رابطه با موجک‌ها و موجک‌های متداول، داده شده و همچنین موجک استفاده شده در این پایان نامه را بررسی می‌کنیم.

در فصل پنجم، ساختار کلی شبکه عصبی و چند نمونه کار انجام شده در این حوزه با استفاده از شبکه‌های عصبی را بازگو کرده و سپس ساختار شبکه عصبی مورد استفاده را بیان و در مورد پارامترهای استفاده شده در آن به شرح و بررسی می‌پردازیم.

در فصل ششم، ساختار سیستم پیشنهادی را به طور کامل توضیح داده شده است. به طور کلی سیستم پیشنهادی دارای سه قسمت می‌باشد که در این فصل به طور کامل بررسی شده‌اند.

در فصل هفتم، به نتیجه گیری پرداخته ایم و پیشنهاداتی برای کارهای آتی نیز ارائه شده است.

---

[1] Wavelet Coefficients

فصل دوم : مروری بر کارهای پیشین

۲-۱- مقدمه

سیستم‌های بازشناسی نوری نویسه را می‌توان به سه مرحله پیش پردازش[۱]، پردازش[۲] و پس پردازش[۳] تقسیم بندی کرد. بررسی هر کدام از این مراحل دقت و حوصله خاص خودش را می‌طلبد و همچنین کار و مطالعه بر روی هر کدام، حائز اهمیت می‌باشد. سعی بر این است که در این فصل یک شمای کلی از کارهای انجام شده در این زمینه را نشان بدهیم.

به طور کلی هدف از اعمال مرحله پیش پردازش، آماده سازی داده‌های خام (اولیه) برای پردازش می‌باشد. خروجی این مرحله، داده‌های هستند که به منظور استخراج هر چه راحتتر ویژگی‌ها[۴] در مرحله پردازش، آماده شده‌اند. نکته قابل توجه در اینجا این است که در حالتی که سیستم شناسایی به صورت برخط عمل می‌کند، مرحله پیش پردازش اهمیت خود را از دست می‌دهد.

داده‌های خام یا به صورت کلمات مجزا، یک جمله، یک خط از کلمات، یک پاراگراف و یا یک صفحه خطوط می‌باشند. در حالت کلی، فرض بر این است که یک صفحه از خطوط به عنوان داده اولیه به سیستم داده شود در اینصورت ابتدا خطوط درون صفحه شناسایی می‌شود که روشهای متداولی برای این کار وجود دارد. به عنوان مثال، با استفاده از هیستوگرام افقی کل صفحه، مقدار متوسط هیستوگرام را برای هر سطر محاسبه کرده و سپس سطرهایی که میزان هیستوگرامشان از

---

[۱] preprocessing
[۲] processing
[۳] post processing
[۴] features

مقدار متوسط کمتر بود به عنوان فاصله مابین خطوط در نظر گرفته می‌شوند. بنابر این صفحه به خطوط شکسته می‌شود. سپس از هیستوگرام عمودی برای جدا کردن کلمات نوشته شده در هر خط استفاده می‌شود (که این مرحله نیز شامل روشهای متنوع می‌باشد).

بعد از استخراج هر کلمه از کل تصویر دریافتی، تصویر کلمه، پیش پردازش می‌شود.

روش پیش پردازش تصویر کلمه به طور کلی شامل مراحل زیر می‌باشد:

۱. کاهش نویز. نویز ایجادشده بواسطه دستگاه‌های اسکنر نوری منجر به ایجاد نقطه‌های لک مانند[۱] ، قطعه خط‌های گسسته ، اتصال بین خطوط، فضاهای خالی در خطوط متن، پرشدن حفره‌های موجود در تصویر برخی حروف، و ... می‌گردد. همچنین اعوجاج‌های[۲] مختلف شامل تغییرات محلی، منحنی‌شدن گوشه‌های حروف، تغییر شکل یا خوردگی حروف را نیز باید در نظر داشت. قبل از مرحله بازشناسی حروف، لازم است که این نقایص برطرف شوند. مهم‌ترین دلیل برای کاهش نویز، کم‌کردن خطا در مراحل قطعه‌بندی و بخصوص بازشناسی می‌باشد. کاهش نویز همچنین سبب کم‌شدن اندازه فایل تصویر می‌شود که به نوبه خود، کاهش زمان مورد نیاز برای پردازش‌ها و ذخیره‌سازی‌های آینده را در پی خواهد داشت [۴۱].

۲. تصحیح کج شدگی کل کلمه[۳]: گاهی اوقات به علت کج قرار دادن برگه در اسکنر، ممکن است تصاویر کلمات به صورت کج وارد سیستم شناسایی بشوند که ایجاد اختلال در عمل شناسایی می‌کنند. در مواردی ممکن است که نویسنده به صورت کج (بدون در نظر گرفتن خط زمینه) کلمه را نوشته باشد که در این صورت نیز باید به وسیله الگوریتم‌های مناسب و با کمک خط زمینه، به طور تقریبی زاویه چرخش را پیدا کرده و کلمه را به حالت متعادل برگردانیم.

۳. تصحیح کج شدگی حروف[۴]: کج شدگی در حروف به طور رایج در دستنوشته‌های عادی وجود دارد. در رسم الخط‌های خاص (مانند نستعلیق در فارسی)، به علت وجود قوانین مربوطه، کج شدگی در حروف بندرت دیده می‌شود. اما در دستنوشته‌های عادی (مانند لاتین پیوسته)، کج شدگی حروف به وضوح قابل مشاهده است. که روشهایی در این زمینه برای

---

[۱] Salt & Pepper
[۲] Distortion
[۳] Skew
[۴] Slant

رفع کج شدگی وجود دارد. در نویسه‌های روزمره فارسی، متاسفانه اعمال این مرحله بسیار مشکل و حتی می‌توان گفت که غیر ممکن می‌باشد.

۴. حذف خط زمینه: در برخی زبانها، دستنوشته‌ها در بالای خط زمینه نوشته می‌شوند که جدا کردن این خطوط کار مشکلی نمی‌باشد. اما در دستنوشته‌های فارسی به خصوص دستنوشته‌های غیر نستعلیق (یعنی رسم الخطهایی بدون وجود قوانین مشخص)، جدا کردن خطوط زمینه کار مشکلی می‌باشد.

۵. نازک سازی[۱] کلمات: در بعضی از سیستم‌ها، در مرحله پیش پردازش گاهی با یافتن ضخامت قلم، ضخامت کل تصویر کلمه را نرمال می‌کنند.

۶. آستانه گیری[۲]: در اغلب سیستم‌های شناسایی نویسه، به طور معمول تصاویر کلمات را با استفاده از آستانه گیری از کیفیت خاکستری به دودویی منتقل می‌کنند.

همانطور که پیش تر گفته شد، برخی از سیستم‌های ارائه شده تنها به بررسی یک مرحله از سه مرحله اصلی شناسایی نوری نویسه می‌پردازند. از جمله کارهای انجام شده در این زمینه، می‌توان به مقاله [۱۵] اشاره کرد که الگوریتمی برای شناسایی و حذف خطوط زمینه ارائه کرده است. مقاله شامل سه الگوریتم برای سه نوع خط زمینه می‌باشد.

۱. خطوط زمینه مستقیم

۲. خطوطی که پائین تر از نیمه کلمه کشیده شده باشند.

۳. خطوطی که به صورت مورب باشند.

تعداد ۳۱۷ کلمه، از مجموعه داده‌های کلمات پیوسته CEDAR[3] انتخاب شده است و برای آزمایش الگوریتم‌ها استفاده شده است. روش پیشنهادی به میزان ۹۷،۱۶٪ خطوط زمینه را به درستی حذف کرده و ۹۶،۱۲٪ کج شدگی کل کلمه را به درستی شناسایی کرده است.

اما در برخی سیستم‌ها به طور همزمان هر سه مرحله شناسایی نویسه نیز در نظر گرفته می‌شود. به عنوان مثال در مقاله [۱۶]، ساختار کلی کار به این صورت می‌باشد. در مرحله

---

۱ Thin
۲ Threshold
۳ Center of Excellence for Document Analysis and Recognition

پیش‌پردازش، ابتدا تصویر کلمه رفع نویز شده و سپس باینری می‌شود (به وسیله یک حد آستانه) و بعد از آن خطوط در کل صفحه شناسایی شده و صفحه به خطوط مجزا تقسیم شده است. سپس هر خط را به کلمات تقسیم کرده‌اند. در مرحله شناسایی، هر کلمه به حروف مجزا قطعه بندی می‌شود و با پیدا کردن خصوصیات هر حرف، به شناسایی می‌پردازد. مقاله مفروض روش پیشنهادیی نیز برای شناسایی بهتر کلمات در مرحله پس پردازش، ارائه کرده است.

داده‌ها شامل ۷۰۰۰ کلمه در اندازه ۱۲۸×۱۲۸ پیکسل بوده که توسط یک نفر نوشته شده است. این کلمات در ۲۵ صفحه بدون وجود علائم انشایی نوشته شده‌اند. روش پیشنهادی ۸۷٪ خطوط را به درستی به کلمات تجزیه کرده و در مرحله جداسازی حروف در کلمات الگوریتم ۷۴،۱۰٪ از ۱۲۰۱ کلمه را به درستی قطعه بندی کرده است.

همانطور که اشاره شد، پس از مرحله پیش پردازش به شناسایی کلمه می‌پردازیم. برای شناسایی کلمه، با توجه به نوع داده‌ها (برخط یا برون خط) روشهای متفاوتی وجود دارد. به علت وجود زبانهای مختلف در دنیای امروز و همچنین وجود کلمات فراوان در هر زبان، شناسایی هر کلمه در هر زبان و با هر رسم الخط، کاری دشوار و حتی می‌توان گفت که با استفاده از علم امروز غیر ممکن می‌باشد. به همین علت، محدودیت‌هایی بر روی سیستم‌های پیشنهادی گذاشته می‌شود تا حجم محاسباتی و پیچیدگی سیستم کاهش یابد. این محدودیت‌ها شامل بررسی دستنوشته‌ها در یک زبان خاص، بررسی دستنوسته‌ها با رسم الخط‌های معین (مانند نستعلیق)، بررسی کلمات با تعداد داده‌های محدود و حتی بررسی کلمات با تعداد نویسنده محدود و غیره می‌باشد.

در برخی از سیستم‌های ارائه شده، مرحله دیگری نیز پس از پردازش تصویر کلمه و به منظور بالا بردن قدرت شناسایی کلمه استفاده می‌شود که این مرحله را پس پردازش می‌نامند. از جمله روشهای پیشنهادی در این مرحله، تطبیق الگویی الگوی شناسایی شده با فرهنگ لغات، اجرای مجدد مرحله پردازش با دادهایی با ضریب اطمینان بالاتر از مجموعه شناسایی شده و غیره می‌باشد.

تحقیقات انجام شده در زمینه حروف لاتین ناپیوسته در مقایسه با دستنوشته‌های پیوسته لاتین، از دقت بالایی برخوردار می‌باشند. اگرچه دستنوشته‌های لاتین پیوسته (که دارای شباهت‌هایی با دستنوشته‌های فارسی و عربی می‌باشد)، به علت پیوسته بودن حروف به یکدیگر و مشکل بودن قطعه بندی کردن کلمات به حروف در روش مبتنی بر قطعه بندی، به سختی قابل شناسایی می‌باشند، اما کارهای ارائه شده با بازدهی بالایی به بهبود بازشناسی کمک کرده‌اند. در واقع می‌توان گفت در