

سنة الفجر

عنوان

بررسی روش‌های تشخیص حروف فارسی

چکیده..... ۵

فصل اول: مقدمه

۱-۱- مقدمه ۷

فصل دوم: مروری بر تشخیص حروف فارسی

۱-۲- مقدمه ۱۰

۲-۲- تاریخچه مختصر خط‌شناسی ۱۱

۳-۲- مشخصات نوشتار فارسی ۱۲

۴-۲- ویژگی‌های الفبای زبان فارسی ۱۳

۵-۲- انواع شیوه‌های نوشتن ۱۴

۶-۲- برخی مشکلات شناسایی دستخط ۱۵

۷-۲- تشخیص دست خط فارسی ۱۶

۸-۲- تشخیص کاراکترهای بصری ۱۷

۹-۲- جمع‌بندی ۱۹

فصل سوم: مطالعه و ارزیابی روش‌های تشخیص حروف فارسی

۱-۳- مقدمه ۲۱

۲-۳- مطالعه روش‌های تشخیص حروف مجزای فارسی ۲۲

۱-۲-۳- بازشناسی برخط حروف فارسی بر پایه مدل مخفی مارکوف ۲۲

۲-۲-۳- تشخیص آنلاین دستنویس فارسی مبتنی بر شبکه عصبی ۲۳

۳-۲-۳- استخراج ویژگی کاراکترهای فارسی آنلاین ۲۳

۴-۲-۳- رویکرد فازی برای بازشناسی و یادگیری حروف مجزای برخط فارسی ۲۳

۵-۲-۳- بازشناسی دست نوشته فارسی با مدل مخفی مارکوف ۲۴

۶-۲-۳- بازشناسی برون خطی کلمات دست نویس فارسی ۲۴

۷-۲-۳- بازشناسی برخط حروف مجزای دست نویس فارسی بر اساس تشخیص گروه اصلی بدنه ... ۲۵

۸-۲-۳- تشخیص دستنوشته برخط فارسی به کمک ویژگی‌های مبتنی بر شکل ۲۵

- ۳-۲-۹- روشی کاربردی برای بازشناسی برخط حروف مجزای دستنویس فارسی..... ۲۶
- ۳-۲-۱۰- تشخیص دستنوشته برخط فارسی با رویکرد تجزیه ای..... ۲۷
- ۳-۲-۱۱- تشخیص حروف دست نویس فارسی با استفاده از تبدیل هاف..... ۲۸
- ۳-۲-۱۲- بازشناسی برخط حروف مجزای دستنویس فارسی بر اساس تشخیص گروه بدنه اصلی ۲۸
- ۳-۳- جمع بندی..... ۲۹

فصل چهارم نتیجه گیری و پیشنهادات

- ۴-۱- نتیجه گیری و پیشنهادات ۳۱
- مراجع ۳۳

فهرست اشکال

- شکل ۱-۲- نوشتن یک کاراکتر به شیوه‌های متفاوت ۱۵
- شکل ۲-۲- تصحیح کلمات غلط نوشته شده ۱۵
- شکل ۳-۲- کم رنگ نوشته شدن حروف ۱۶
- شکل ۴-۲- دو لبه نوشتن حروف ۱۶
- شکل ۵-۲- ضخیم نوشتن حروف ۱۶

چکیده

تشخیص حروف یکی از موضوعات جالبی است که در قرون اخیر توجه بسیاری از محققین را به خود جلب کرده است اگرچه حروف فارسی برای زبان‌هایی مانند فارسی، عربی وارد و استفاده می‌شود ولی برای توسعه یک سیستم تشخیص حروف مجزای توجه کافی به آن نشده است و به رغم فراگیری نسبی کاربرد الفبای فارسی در میان ملل مختلف قاره آسیا بررسی‌های انجام شده در خصوص یافتن روش‌هایی برای بازشناسی حروف این الفبا بسیار محدود بوده است در این سمینار روش‌های تشخیص حروف زبان فارسی مورد مطالعه و بررسی قرار گرفته شده است. اکثر روش‌های از مدل مارکوف در تشخیص حروف کمک گرفته‌اند. با در نظر گرفتن سایر روش‌ها و بررسی انجام گرفته در این سمینار به این نتیجه رسیدیم که می‌توان از الگوریتم‌های هوشمند در تشخیص حروف فارسی استفاده کرد.

کلمات کلیدی: تشخیص حروف، الفبا، دست خط، شناسایی.

فصل اول

مقدمه

خط‌شناسی عبارت است از مطالعه و تحلیل دستخط یک نفر به منظور گردآوری اطلاعات درباره ویژگی‌های مختلف شخصیتی او. بیشتر دانشمندان، خط‌شناسی را به عنوان یک "شبه علم" محسوب می‌دارند. اما تحلیل دستخط برای تشخیص صحت امضاء در مستندات مختلف به عنوان یک شناسه قانونی به کار می‌رود. عبارت خط‌شناسی همچنین به بررسی علمی دستخط برای تشخیص جعلی بودن یا نبودن اطلاق می‌گردد و شهادت متخصصان خط‌شناسی در این مورد در دادگاه‌ها پذیرفته می‌شود. خط‌شناسان معمولاً به یک نمونه خط که در شرایط عادی و با جوهر نوشته شده باشد نیاز دارند. نمونه خط نباید نام، جنسیت، ملیت، دین و چپ دست یا راست دست بودن نویسنده را آشکار سازد. هیچ دو دستخطی، مثل اثر انگشت، دقیقاً یکسان نیستند. دستخط هر کس، یگانه و منحصر به فرد است. الگوهای نوشتاری هیچکس با کس دیگر به هیچوجه مطابقت ندارد. خط‌شناسان معمولاً برای تحلیل دستخط، عوامل بسیاری را از قبیل نوع حرکت قلم، اندازه‌ها، شکل‌بندی‌ها، انحناها، فاصله بین حروف، حروف بزرگ و کوچک، یکنواختی امضاها، زاویه‌ها، فشار قلم، تقاطع‌ها، شیب‌ها، حاشیه‌ها، اتصالات، فاصله بین کلمات، خطوط مبنا، ارتفاع و پهنای حروف و غیره در نظر می‌گیرند که نشانگر جنبه‌های مرتبط با ویژگی‌های شخصیتی نویسنده می‌باشد. معمولاً دستخط‌ها دارای دو نوع مشخصات عمومی و خصوصی هستند. مشخصات عمومی ویژگی‌هایی از دستخط هستند که به شکل کلی دستخط و زوایای دستخط مربوط می‌شود. این مشخصات عبارتست از طول خطوط اصلی، زاویه خطوط اصلی، فاصله بین خطوط اصلی و چینش از سمت راست و چپ و غیره. مشخصات خصوصی یک دستخط ویژگی‌هایی مربوط به حروف و کلمات آن دستخط می‌باشند این ویژگی‌ها عبارتند از میزان چسبندگی، طول نویسه‌ها، کشیدگی، فاصله بین حروف، جهت حروف و غیره که در شناخت عادات نوشتاری مفید هستند [1]. شناسایی دستخط عبارتست از تعیین حروف یا کلمات موجود در تصویر دیجیتال که توسط انسان نوشته شده است. این در ارتباط بین انسان و ماشین بسیار اهمیت دارد و می‌تواند به پردازش اتوماتیک مدارک دستخط کمک کند. شناسایی دستخط زیر شاخه‌های از OCR است [2]. شناسایی حروفات دستنویس فارسی یکی از موضوعات تحقیقاتی در زمینه شناسایی الگو، یادگیری ماشین، پردازش تصویر، بینایی ماشین و داده کاوی است. کاربردهای زیادی از جمله شناسایی پلاک اتومبیل، استخراج کلمات کلیدی از تصویر و استفاده از آن برای حاشیه نویسی و تشخیص کدپستی و چک‌های بانکی و سیستم ورود نمره در دانشگاه‌ها وجود دارد. در شناسایی دستخط با مشکلاتی از جمله

تفاوت در نوع دستخط، تفاوت در فشار نویسنده به قلم، استفاده از قلم ضخیم یا نازک روبرو هستیم. روش‌های شناسایی حروف می‌تواند بر مبنای دو معیار: فرآیند اکتساب داده (برخط و برون از خط) و نوع متن (نوشته شده توسط ماشین یا نوشته شده توسط دست) دسته‌بندی شود. دو روش را برای تشخیص حروف وجود دارد یکی روش تحلیل که متن به حروفات تقسیم می‌شود و سپس حروفات شناسایی شده و سپس از نتایج تولید شده یک لیست از کلمات ممکن تولید می‌گردد که نیاز به لغتنامه بزرگی دارد که بسیار پرهزینه است و دیگری رویکرد سراسری است که کل متن برای شناسایی استفاده می‌گردد [3].