
بسم الله الرحمن الرحيم

دانشکده مهندسی برق و رباتیک
گروه الکترونیک
پایان نامه کارشناسی ارشد

جستجوی کلمات در یک تصویر دست‌نوشته فارسی با استفاده از روش تعبیه مشخصه‌ها

بهاره اسدی

استاد راهنما:
دکتر علیرضا احمدی‌فرد

استاد مشاور:
مجید ایرانپور مبارکه

بهمن ۱۳۹۴

تقدیم به:

چشمه های جوشان محبت

جلوه های مهر و عطوفت الهی

لبخندهای پر مهر زندگیم

پدر و مادر عزیزم

که در تمام مراحل زندگی، به من راه و رسم درست زیستن را آموختند.

تقدیر و تشکر:

بر خود لازم می دانم مراتب سپاس و قدردانی خویش را از استاد گرامی جناب آقای دکتر علیرضا احمدی فرد که با راهنمایی دقیق و مستمر در شکل گیری این پایان نامه نقش مهمی داشتند، ابراز نمایم. بی شک نظرات و کمک‌های ایشان تأثیر بسزایی در پیشبرد هر-چه بهتر و سریعتر این پایان نامه داشته است. از استاد گرامی جناب آقای دکتر مجید ایرانپور که زحمت مشاوره این رساله را بر عهده گرفته‌اند کمال تشکر و قدردانی را دارم. همچنین از دوستان عزیزم که در ثمر رسیدن این پایان نامه یار و همراهم بودند تقدیر و تشکر می‌کنم.

تعهــد نامه

اینجانب **بهاره اسدی** دانشجوی دوره کارشناسی ارشد رشته مهندسی برق/الکترونیک دانشکده مهندسی برق و رباتیک دانشگاه صنعتی شاهرود نویسنده‌ی پایان نامه **جستجوی کلمات در یک تصویر دست‌نوشته فارسی با استفاده از روش تعبیه مشخصه‌ها** تحت راهنمایی دکتر علیرضا احمدی‌فرد متعهد می‌شوم .

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technologhy » به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ:

امضای دانشجو :

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده

جستجوی کلمه‌ی کلیدی، روشی برای استخراج اطلاعات متنی در اسناد تصویری است. در این روش کلمه‌ی پرسشی مورد نظر کاربر در اسناد تصویری، جستجو شده و مکان نمونه‌های آن کلمه در داخل اسناد مشخص می‌شود. بیشتر تحقیقات انجام شده در این زمینه بر روی متون زبان لاتین ارائه شده است. تحقیقات انجام شده در زبان فارسی و عربی بسیار کم می باشد که اکثر این تحقیقات هم بر روی اسناد چاپی هستند. در این پایان نامه الگوریتمی برای جستجوی کلمات در تصویر دست‌نوشته فارسی ارائه می‌دهیم که شامل دو مرحله است.

در مرحله‌ی اول ابتدا لازم است خطوط متن از یکدیگر جدا شده سپس تصاویر کلمات داخل متن استخراج شوند. برای انجام این مرحله مولفه‌های متصل موجود در متن استخراج شده سپس مولفه‌ی متصل که شروع کننده هر خط است مشخص می‌شود. سپس با اعمال روش نزدیک‌ترین همسایه، تمام مولفه‌های متصل هر خط بصورت یک زنجیره و به همان ترتیبی که در متن قرار دارند بدست می‌آیند. در نتیجه بعد از این مرحله می‌توان هر خط را با مولفه‌های متصلش به طور مستقل از مابقی خطوط نمایش داد. نتیجه این مرحله جداسازی خطوط از یکدیگر می‌باشد. در مرحله‌ی بعدی با در نظر گرفتن این قانون که اکثر حروف اضافه و کلمات فارسی از ترکیب یک تا هفت مولفه متصل تشکیل شده‌اند، استخراج کلمات متن انجام می‌شود. برای هر کدام از این تصاویرکلمات استخراج شده، مختصات مکانی آنها در متن نیز ذخیره می‌شود.

در نتیجه انجام مرحله‌ی اول مجموعه‌ای از کلمات کاندید، ایجاد شده که در مرحله‌ی دوم هدف یافتن نمونه‌های کلمات پرسشی در این مجموعه و مشخص کردن مکان آنها در متن می‌باشد. کلمات پرسشی ورودی به صورت تصویر و رشته‌ی متنی هستند. در این مرحله برای جستجوی کلمه‌ی پرسشی، از مشخصه‌ها با در نظر گرفتن الفبای فارسی استفاده شده است که به کارگیری این نوع مشخصه باعث می‌شود که شباهت بین کلمات مستقل از شیوه نگارش حفظ شود.

از مجموعه‌ای شامل ۵۵ کلمه پرسشی و پایگاه داده‌ای شامل ۱۰۰ سند دست‌نوشته فارسی برای ارزیابی روش استفاده شده است. متوسط دقت و فراخوانی به ترتیب برای پرسش با تصویر ۷۵/۴۵٪ و ۸۰/۴۲٪ و برای پرسش با رشته‌ی متنی ۸۰/۲۸٪ و ۷۸/۹۲٪ می‌باشد.

کلمات کلیدی

جستجوی کلمه‌ی کلیدی، اسناد دست‌نوشته‌ی فارسی، مشخصه، تعبیه برچسب، جداسازی خطوط متن، مولفه‌های متصل

لیست مقالات مستخرج

اسدی.ب، احمدی فرد.ع، ایرانپور.م، "جداسازی خطوط متن و استخراج برون خط دنباله‌ای از مولفه-های متصل در دست‌نوشته فارسی"، نهمین کنفرانس ماشین بینایی و پردازش تصویر ایران"، تهران، ۱۳۹۴.

اسدی.ب، احمدی فرد.ع، ایرانپور.م، "جستجوی کلمات در یک تصویر دست‌نوشته فارسی با استفاده از روش تعبیه مشخصه‌ها"، سومین کنفرانس بین‌المللی پژوهش‌های کاربردی در مهندسی کامپیوتر و فناوری اطلاعات "، تهران، ۱۳۹۴.

فهرست مطالب

فهرست شکل‌ها

فهرست جدول‌ها

فصل اول:

مقدمه

۱-۱ اهداف پایان‌نامه

متن همیشه یکی از منابع اطلاعاتی مهم محسوب می‌شود و امروزه حجم بسیاری از این نوع داده در کتابخانه‌های دیجیتال، کاغذ، تصاویر و ویدئو یافت می‌شود. با پیشرفت فناوری دیجیتال، روزانه هزاران سند و کتاب به صورت اسکن شده درآمده‌اند و به عنوان سند تصویری ذخیره می‌شوند. تکنولوژی مدرن امکان تولید، ذخیره و انتقال اسناد تصویری را به طور کارامد ایجاد می‌کند. تکنیک-های تشخیص و تحلیل اسناد تصویری[1] برای استخراج و در دسترس قرار دادن اطلاعات متنی مورد استفاده قرار می‌گیرند. دو رهیافت اصلی برای مشخص کردن متن مورد نظر در اسناد تصویری برای بازیابی اطلاعات وجود دارد: روش‌های تشخیص نوری نویسه‌ها (OCR)[2] [۱] و روش‌های جستجوی کلمه کلیدی[3]

نرم‌افزار OCR به بازشناسی متون موجود در اسناد تصویری پرداخته و آنها را به متون قابل جستجو و ویرایش توسط رایانه تبدیل می‌کند. این روش در مورد فونت‌های ساده خوب با پس‌زمینه تمیز عملکرد قابل قبولی دارد اما کارایی این روش با کاهش کیفیت اسناد یا در مورد اسناد دست‌نویس به شدت کاهش می‌یابد. همچنین این روش در مورد پایگاه داده‌ای که دارای تعداد زیادی سند تصویری است روشی زمان‌بر می‌باشد [۲].

بنابراین، محققین به دنبال روش‌های بازیابی بوده‌اند که اطلاعات متن را بدون تبدیل کل تصویر سند به متن جستجو می‌کند. از این رو روش‌هایی که قادر به بازیابی تصاویر حاوی یک واژه مشخص باشند یا بتوانند لغات را در یک تصویر تشخیص دهند مورد توجه قرار گرفته‌اند. جستجو کلمات ابتدا بوسیله‌ی Jones در زمینه پردازش گفتار پیشنهاد شده است [۳]. بعدها این تعریف توسط چندین محقق برای اندیس‌گذاری اسناد چاپی و دست‌نویس بکار گرفته شد. این رهیافت موقعیت کلمه انتخاب

[1] Document Image Analysis and Recognition (DIAR)
[2] Optical Character Recognition (OCR)
[3] Keyword Spotting

شده توسط کاربر را در یک سند تصویری مشخص می‌کند به عبارت دیگر، سیستم جستجوی کلمه‌ی کلیدی، کلمه موردنظر کاربر را در سند تصویری جستجو می‌کند درحالیکه توجهی به کلمات نامرتبط ندارد.

استفاده از این روش و ذخیره اسناد تصویری سبب می‌شود اطلاعات مفید و مهم که نمی‌توانند بطور مناسبی توسط OCR تشخیص داده شوند مانند گرافیک، تصاویر و ویژگیهای وابسته به نگارش مانند فونت و صفحه‌آرایی، حفظ بشوند [۴].

یک کاربرد این روش در اسناد تاریخی تصویری می‌باشد؛ به طور مثال در نظر گرفتن کلمه‌ی پرسشی Fig، به شناسایی مکان تصاویر و عنوان‌ها موجود در سند کمک می‌کند یا یک امکان جالب دیگر مشخص کردن مکان همه‌ی نمونه‌های کلمه‌ی Complain به معنای شکایت، درون نامه‌های ارسال شده برای یک شرکت است. نکته قابل ذکر دیگر اینکه گوگل[۱] و یاهو[۲] اعلام کرده‌اند که کتاب‌ها را در موتورهای جستجویشان در دسترس قرار می‌دهند [۵]. در این زمینه جستجوی کلمه ابزار ارزشمندی برای جستجوی کاربران روی محتوای این کتب خواهد بود.

بیشتر تحقیقات انجام شده در این زمینه بر روی متون زبان لاتین ارائه شده است و تعدادی از آنها در مورد زبان‌های دیگری مثل چینی وکره‌ای می‌باشند. تحقیقات انجام شده در زبان فارسی و عربی بسیار کم است که بیشتر آنها هم بر روی اسناد چاپی هستند. به همین دلیل با توجه به اهمیت مسئله و اینکه طبق بررسی‌های صورت گرفته کار قابل توجهی در زبان فارسی انجام نشده است، نیاز به انجام تحقیق در این زمینه احساس می‌شود.

[1] www.google.com
[2] www.yahoo.com

۲-۱ دسته‌بندی روش‌های جستجوی کلمه‌ی کلیدی

روش‌های جستجوی کلمه‌ی کلیدی بر مبنای نوع کلمه‌ی پرسشی به دو دسته تقسیم می‌شوند؛ کلمه مورد پرسش ممکن است یک رشته متن باشد که در این مورد به آن پرسش با رشته یا پرسش با متن[۱](QBT) گفته می‌شود، همچنین ممکن است یک تصویر باشد که در این حالت به آن پرسش با مثال[۲] (QBE) گفته می‌شود. روش‌های تشخیص کلمه کلیدی بر مبنای جستجو به روش‌های مبتنی بر قطعه‌بندی و مستقل از قطعه‌بندی تقسیم می‌شوند. در حالت اول سند مورد جستجو در سطح خط یا کلمه قطعه‌بندی شده و در حالت دوم بدون قطعه‌بندی کل سند مورد جستجو قرار می‌گیرد.

دسته‌بندی دیگری در این زمینه بر مبنای دو رهیافت کلی تطبیق قالب[۳] و مبتنی بر یادگیری است. تطبیق می‌تواند پیکسل به پیکسل یا مبتنی بر ویژگی باشد. در روش تطبیق پیکسل به پیکسل فاصله بین پیکسل‌های متناظر تصویر کلمه مورد پرسش و تصاویر کلمات داخل اسناد محاسبه می‌شود. معیار فاصله مینکوسکی مانند فاصله بلوک شهری [۶] و اقلیدسی [۷] بصورت گسترده‌ای برای این منظور مورد استفاده قرار گرفته است. در روش تطبیق مبتنی بر ویژگی تعداد ثابتی از ویژگی‌ها استخراج شده و بصورت یک بردار ارائه شده است و شباهت بین تصاویر با مقایسه بردارهای آنها تعیین می‌گردد. الگوریتم‌های مورد استفاده در این روش SC [۸] ، SLH [۹]، DTW [۱۰, ۱۱] و CORR [۱۲, ۱۳] می‌باشد. ویژگی‌های ساختاری و آماری زیادی وجود دارند که در بازیابی اسناد تصویری و سیستم‌های جستجوی کلمه مورد استفاده قرار گرفته‌اند. این ویژگی‌ها شامل پروفایل‌های افکنش[۴]،

---

[1] Query by text(QBT)

[2] Query by example(QBE)

[3] Template matching

[4] Projection profile

پروفایل بالا و پایین کلمه[1]، گذر جوهر به پس‌زمینه[2]، ارتفاع، عرض، نسبت طول به عرض[3]، ممان‌ها، میانگین، واریانس، تبدیلات موجک و فوریه، ویژگی‌های دودویی براساس گرادیان[4]، حفره‌ها[5]، تقعر، بالارونده و پایین‌رونده[6] و غیره می‌باشند [۱۴].

ارائه روش‌های جستجوی کلمات برای یک زبان ویژه با در نظر خواص بخصوص آن زبان است. متون متفاوت دارای خواص متفاوت و بنابراین دارای روش‌های متفاوتی در جستجوی کلمه می‌باشند. حروف فارسی دارای خواصی هستند که اعمال روش‌های جستجوی کلمه بر روی آنها در مقایسه با حروف انگلیسی مشکل‌تر است. این ویژگی‌های زبان فارسی شامل موارد زیر است:

۱. پیوسته نوشته شدن کلمات فارسی در اسناد دست‌نوشته و تایپی
۲. تشکیل شدن کلمات از زیر کلمات به جای حروف
۳. ساخته شدن تعداد زیادی از حروف با بیشتر از یک جزء
۴. ظاهر وابسته به محل قرارگیری حروف درون یک کلمه
۵. حضور بیش از اندازه تعداد نقاط، مکمل‌ها مثل سرکش یا آکلاد و علائم نشانه‌گذاری .

۱-۳ ساختار پایان‌نامه

در فصل دوم کارهای انجام شده در زمینه‌ی جستجوی کلمات مرور شده است. در فصل سوم به تئوری مورد نیاز در روش پیشنهادی پرداخته می‌شود. در فصل چهارم الگوریتم پیشنهادی برای

---

[1] Upper/lower word profile
[2] Background to ink transition
[3] Aspect ratio
[4] Gradient based binary features (GSC)
[5] Holes
[6] Ascenders and Descenders

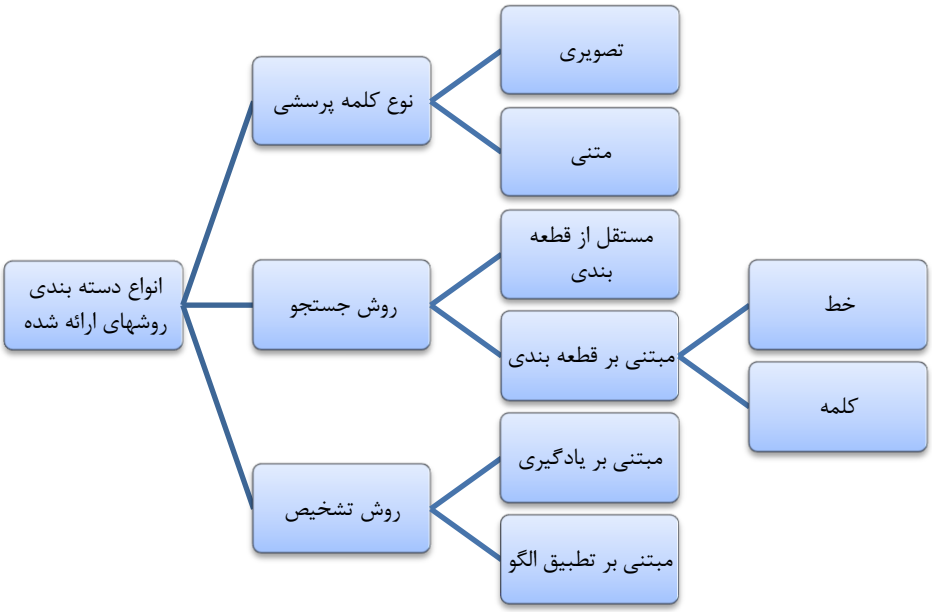
جستجوی کلمات در متون دست‌نوشته‌ی فارسی و نتایج ارزیابی آن بیان شده است. در فصل پنجم روش پیشنهادی جداسازی خطوط و استخراج برون خط دنباله‌ای از مولفه‌های متصل در متون پیوسته‌ی فارسی و نتایج ارزیابی آن ارائه شده است. فصل ششم به نتیجه‌گیری و ارائه پیشنهادات اختصاص دارد.

فصل دوم:

مروری بر فعالیت‌های انجام شده

۲-۱ مقدمه

در این فصل تحقیقات انجام شده در زمینه‌ی جستجوی کلمات در متون دست‌نوشته در زبان‌های مختلف بررسی می‌شوند. بیشتر روش‌ها در این زمینه یک الگوی مشترک را دنبال می‌کند: ابتدا اسناد در یک فاز برون‌خط ذخیره و شاخص‌گذاری می‌شوند، سپس کاربر یک پرسش را مطرح می‌کند و سیستم شباهتش را با اسناد ذخیره شده ارزیابی می‌کند. بطور کلی سیستم تشخیص کلمه کلیدی کلمه موردنظر کاربر را در سند تصویری جستجو می‌کند درحالیکه توجهی به کلمات نامرتبط ندارد. چنین سیستمی یک پاسخ مثبت یا منفی نسبت به پرسش کاربر ارائه می‌دهد. با توجه به تنوع کارهای انجام‌شده روی اسناد، روش‌های موجود در جستجو کلمه را می‌توان از چند دیدگاه تقسیم‌بندی کرد. در این فصل با توجه به تقسیم‌بندی صورت گرفته، تحقیقات انجام شده درگذشته مورد بررسی قرار می‌گیرد. دسته‌بندی روش‌ها موجود در شکل ۲-۱ نشان داده شده است.



شکل ۲-۱: دسته‌بندی روش‌های موجود در این زمینه

۲-۲ تقسیم‌بندی براساس روش جستجو

در زمینه جستجو کلمات روش‌های موجود ارائه شده ممکن است سند مورد جستجو را در سطح

خط یا کلمه قطعه‌بندی کنند یا بدون قطعه‌بندی کل سند را مورد جستجو قرار دهند. در صورتی که نیاز به قطعه‌بندی کلمه وجود داشته باشد، ابتدا باید خطوط و سپس کلمات از سند جدا شود، که عمل پر خطایی است. در نوع دیگر فقط خطوط استخراج می‌شوند که باز هم مشکل خطای ناشی از آن وجود دارد ولی از نوع قبلی کمتر است. در آخرین مورد هیچ عمل قطعه‌بندی انجام نمی‌شود و از بروز این خطا پیشگیری می‌کند اما چون نیاز به بکار بردن روش‌هایی چون پنجره لغزان وجود دارد پردازش پرهزینه‌ای را نیاز خواهد داشت.

۲-۲-۱روش‌های مبتنی بر قطعه‌بندی

یک عامل مهم در روش‌های جستجوی کلمات، روش قطعه‌بندی بکار رفته برای استخراج کردن عناصر متنی از اسناد است. قطعه‌بندی می‌تواند در سطح محلی یا سراسری انجام شود. در رهیافت محلی کارهای اولیه اغلب براساس کدگذاری شکل کاراکتر[1] است. در این روش‌ها تصاویرکاراکترها، استخراج شده و با مجموعه کدهای از پیش تعریف شده تفسیر می‌شوند. محدودیت اصلی این روش‌ها حساسیت به خطای قطعه‌بندی کاراکتر است. به این دلیل که اگر یک یا چند کاراکتر بصورت شکسته نوشته شده باشند و یا کاراکترها دارای چسبندگی باشند کدگذاری آنها با اشتباه همراه خواهد بود.

برای غلبه بر این مشکل، شمای کدگذاری شکل کلمه[2] ارائه شد که به جای قطعه‌بندی هر کلمه به کاراکترهایش، تصویر کلمه بعنوان یک جزء واحد در نظر گرفته شده و ویژگی‌های آن بدون قطعه-بندی استخراج شده است. به عبارت دیگر این روش‌ها براساس کد کردن شکل کلمه است که هر تصویر کلمه بصورت یک سری از نشانه‌های وابسته به کاراکترها کد می‌شود. در بیشتر موارد مجموعه نشانه‌ها[3] یک اندازه کوچکتری نسبت به مجموعه کاراکتر زبان مرجع دارد و این کارِ تشخیص را ساده-

[1] Character Shape Coding(CSC)
[2] Word Shape Coding (WSC)
[3] Symbol Set

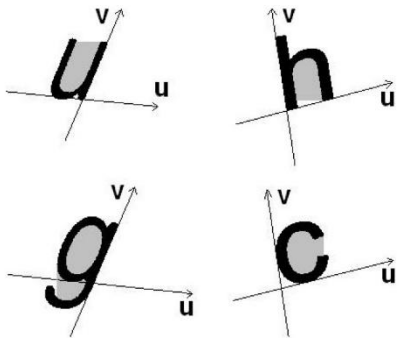
تر می‌کند. مزیت اصلی این رهیافت‌ها فرمول‌بندی ساده پرسش و عدم نیاز به فاز آموزش است. ولی بهرحال وابسته به زبان هستند و نسبت به تصاویر با کیفیت پایین شکننده هستند. این مجموعه کدها، ساختار کلی و توپولوژی هر کلمه را توصیف می‌کنند. تعداد مولفه‌ی متصل، نقاط، حفره‌ها، بالارونده‌ها، پایین‌رونده‌ها و ....مثال‌هایی از این ویژگی‌های ساختاری هستند.

تحقیق [۱۵] با روش کدگذاری شکل کاراکتر تلاش کرده تا موقعیت کلمات کلیدی در تصاویر ضبط شده بوسیله دوربین دیجیتال را مشخص کند. ابتدا خطوط و کلمات با آنالیز مولفه‌های متصل[1] استخراج می‌شوند. جهت‌های عمودی و افقی متن برای آنالیز شکل کاراکتر نیاز می‌باشند. به همین علت ابتدا جهت‌های عمودی متن با استفاده از حروف i و l و جهت‌های افقی متن با استفاده از خطوط مبنا و خطوط X تعیین می‌شوند. تصاویر کلمه با استفاده از سه ویژگی ثابت حفره کاراکترها ، مخزن آب‌ها[2] و کاراکترهای بالارونده و پایین‌رونده کد می‌شوند. این ویژگی‌ها در شکل ۲-۲ نشان داده شده-اند. این سه ویژگی ثابت، هر تصویر کاراکتر را با یک سری ارقام ۶ بُعدی کدگذاری می‌کنند. سپس هر تصویر کلمه با کنار هم قرارگرفتن کد شکل کاراکتر متناظر با آن، کدگذاری می‌شود. اولین عنصر این کد، تعداد گودال‌ها درون کاراکتر مورد بررسی را ثبت می‌کند. عنصرهای دوم تا پنجم مخازن آب سمت چپ، راست، بالا و پایین را ثبت می‌کند و آخرین عنصر اطلاعات بالارونده و پایین‌رونده را با اختصاص ۱، ۱- و ۰ به ترتیب مشخص می‌کند. به عنوان مثال کلمه‌ی the به صورت ۰۰۱۰۰۱۰۰۰۰۱۱۱۰۱۰۰۰ کد می‌شود که سه زیر دنباله‌ی ۰۰۰۱۰۰۱، ۰۰۰۱۱۰و۱۰۱۰۰۰ از کد کردن سه کاراکتر t، h وe به ترتیب تولید شده‌اند. در شکل ۲-۳ کد کاراکترهای لاتین که توسط این روش کدگذاری شده‌اند مشخص است.

در نهایت میزان مشابهت بین کد شکل کلمه‌ی پرسشی و کد شکل کلمات داخل اسناد مورد

[1] Connected Components
[2] Water reservoirs

بررسی، هم طول با کلمه‌ی پرسشی مورد نظر، با استفاده از فاصله همینگ[1] بین آنها محاسبه می‌شود. در این روش متوسط فراخوانی و دقت به ترتیب ۷۵/۷٪ و ۶۹/۰۸٪ است و بیشترین مقدار F_1 در آستانه‌ی ۰/۹، برابر با ۷۹/۹۱٪ می‌باشد.



شکل ۲-۲: وجود حفره در g و مخازن آب در g، h،u و c [۱۵]

Characters	Codes	Characters	Codes
a	110010	bd	100001
c	001000	e	101100
ft	001001	g	11010-1
h	000011	ir	000000
j	00000-1	k	001111
l	000001	m	000020
n	000010	o	100000
pq	10000-1	s	011110
uv	000100	w	000210
x	011110	y	01010-1
z	011000		

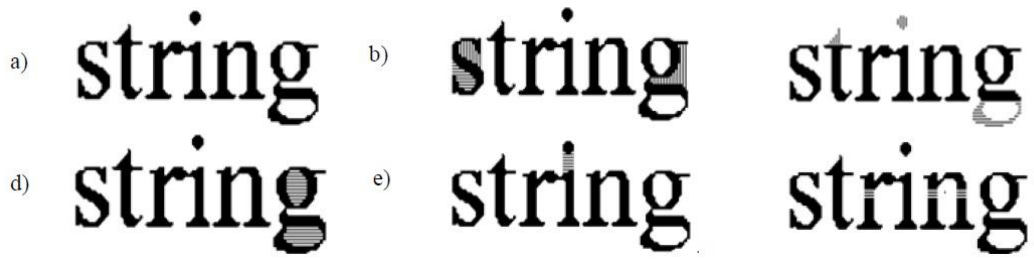
شکل ۲-۳: روش کدگذاری شکل حروف [۱۵]

در تحقیق [۱۶]، روشی برای کدگذاری شکل کلمه روی زبان لاتین، که در برابر زوائد حروف، سبک نگارشی فونت‌ها، چسبندگی بین حروف و حروف شکسته و روی هم افتاده مقاوم است ارائه شده است. در این روش ابتدا خطوط متن و کلمات، استخراج شده و سپس برای هر کلمه هفت ویژگی متفاوت تشخیص داده می‌شود. با استفاده از این ویژگی‌ها، کلمه به کد شکل کلمه تبدیل می‌شود. این

[1] Hamming distance

ویژگی‌ها شامل کاراکتر بالارونده، پایین‌رونده، حفره‌ها، تقعر شرقی، تقعر غربی، رابط i نقطه[1] و محل تقاطع خط میانی[2] می‌باشند که به ترتیب با i,c,x, o,n,l و r کد می‌شوند. این ویژگی ها در شکل ۲-۴ نشان داده شده‌اند. در این شکل b تقعر شرقی و غربی، c بالارونده و پایین‌رونده، d حفره، e رابط i نقطه و f تقاطع خط میانی را نشان می‌دهد. به طور مثال در اولین کدگذاری کاراکتر t که فقط یک ویژگی بالارونده دارد با کد l تفسیر شده است. در حالت دوم کاراکتر x با داشتن دو ویژگی تقعر غربی و شرقی بصورت xc کدگذاری می‌شود. در حالت سوم کاراکتر e با داشتن دو ویژگی حفره و تقعر شرقی به کد e نگاشت می‌شود که در این حالت دو ویژگی با هم ترکیب شده و به کد منحصربه‌فرد e اختصاص می‌یابند. بعد از فرآیند کدگذاری شکل کلمه و با داشتن کد شکل کلمه‌ی پرسشی و کد شکل کلمات داخل اسناد میزان شباهت بین آنها با برنامه‌ریزی پویا[3] محاسبه می‌شود.

این روش در مقایسه با OCR، ۱۵ برابر سریع‌تر است. از دو پایگاه داده‌ی PAT و UW1 برای ارزیابی این تحقیق استفاده شده است. دقت، فراخوانی و F_1 به ترتیب برای اولین پایگاه داده برابر با ۱۰۰٪، ۹۸/۷۸٪ و ۹۹/۳۸٪ است. بهترین مقدار F_1 برای پایگاه داده UW1 در آستانه ۰/۹ است که در این حالت دقت، فراخوانی و F_1 به ترتیب برابر با ۹۷/۵۴٪، ۸۹/۵۵٪ و ۹۳/۳۷٪ است.



شکل ۲-۴: نمایش ویژگی‌های ساختاری تحقیق [۱۶]

با توجه به این موضوع که در زبان‌هایی مانند فارسی، عربی و اردو کلمات از زیرکلمات تشکیل می‌شوند قطعه‌بندی به زیرکلمات به جای حروف مطرح است. در تحقیق [۱۷]، با تخصیص یک کد به

[1] i-dot connector
[2] Middle line intersection
[3] Dynamic Programming

هر یک از بالارونده‌ها، پایین‌رونده‌ها، حلقه‌ها، نشان‌های تشخیص بالا و پایین (مانند نقاط)، کد توصیف کننده‌ی شکل کلمات در زبان عربی بدست می‌آید. در شکل ۲-۵ کدهای حروف عربی در این تحقیق نشان داده شده است.

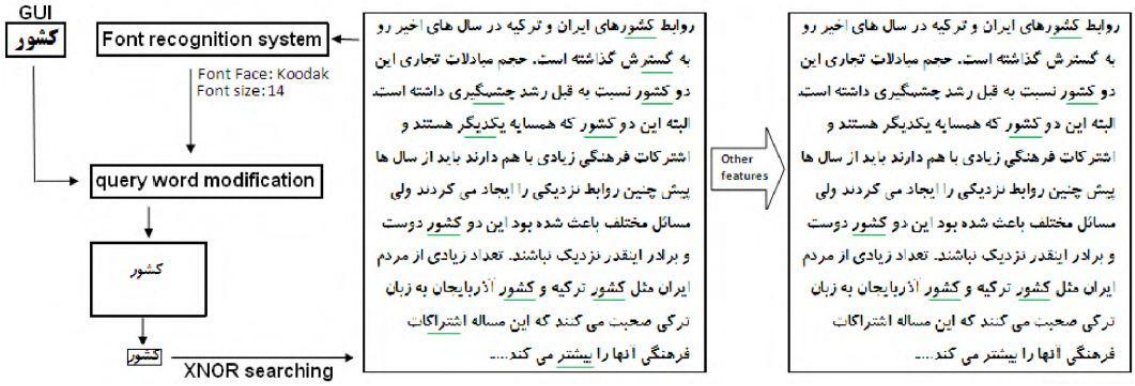
Character	*Code*	*Designation*
ا - ا - ک - ل	h	Ascender
إ	hq	Ascender +Down Diacritic
أ - أ	ph	Up Diacritic +Ascender
ل	hj	Ascender +Descender
ط	bh	Loop +Ascender
ظ	bph	Loop +Up Diacritic +Ascender
لا	hbh	Ascender +Loop+ Ascender
ك	hp	Ascender +Up Diacritic
ي	jq	Descender +Down Diacritic
غ - خ - ذ - ئ - ئ - ن	p	Up Diacritic
غ	jp	Descender +Up Diacritic
ش - ث	pp	Up Daicritic +Up Diacritic
ن - ز - خ - ئ	jp	Descender +Up Diacritic
ش	ppj	Up Diacritic + Up Diacritic + Descender
ض	bpj	Loop + Up Diacritic + Descender
ضـ - فـ - ة - غـ - قـ	bp	Loop +Up Diacritic
ق	pbj	Up Diacritic + Loop + Descender
بـ - جـ - ب	q	Down Diacritic
ح - ع - س - ر - ى	j	Descender
ج	jq	Descender + Down Diacritic
عـ - مـ - صـ - ه - ه	b	Loop
هـ	bb	Loop +Loop
ع - و - ص - م	bj	Loop +Descender
لا	hh	Ascender +Ascender
ة	pb	Up Diacritic +Loop
غ	pbj	Up Diacritic + Loop + Descender
ؤ	bjp	Loop + Descender + Up Diacritic
لأ	hbhp	Ascender + Loop + Ascender + Up Diacritic
لإ	hbqh	Ascender+Loop+Down Diacritic+ Ascender

شکل ۲-۵: کدهای حروف عربی [۱۷]

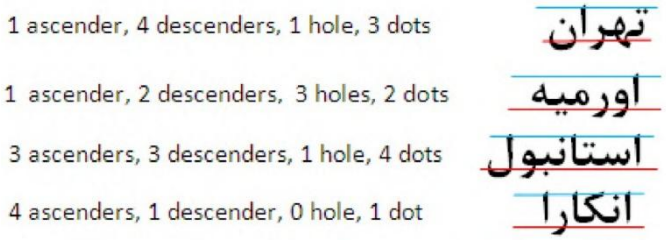
در تحقیق [۱۸]، روش قطعه‌بندی مبتنی بر جداسازی خطوط می‌باشد. این تحقیق روی اسناد چاپی فارسی ارائه شده است. در این روش، خطوط توسط پروفایل پروجکشن افقی از یکدیگر جدا شده‌اند. سپس برای جستجوی یک کلمه ابتدا نوع و اندازه فونت سندی که شامل کلمه‌ی پرسشی

مورد نظر است شناسایی شده و بر این اساس کلمه‌ی پرسشی بازنویسی می‌شود به طوری کـه کلمـه‌ی پرسشی و نمونه‌های آن در سند کاملا مشابه می‌باشـند. در مرحلـه‌ی بعـدی پنجـره‌ای شـامل تصـویر کلمه‌ی پرسشی روی تمام خطوط سند تصویری پیکسل به پیکسل حرکت کـرده و تـابع XNOR بـین تصویر کلمه‌ی پرسشی و کلمات داخل متن محاسبه می‌شود. این تابع میزان شباهت را نشان می‌دهـد. در این مرحله بعضی از کلمات متن که دارای بیشترین میزان شباهت هستند به عنوان کاندید انتخـاب می‌شوند. برای تشخیص نهایی، ویژگی‌هایی مانند تعداد حفره‌ها، تعداد حروف بالارونده و پایین‌رونـده و تعداد نقاط از کلمه‌ی پرسشی و تصاویر کاندید مرحله‌ی قبل استخراج می‌شوند. هـر تصـویر کاندیـدی که ویژگی‌های آن منتاظر با کلمه‌ی پرسشی باشد به عنوان نمونه‌ی کلمـه پرسشـی در مـتن مشـخص می‌شود.

این تحقیق روی ۴۴۰ سند تصویری چاپی که توسط رایانه ایجاد شده‌اند و ۲۲۴ سند تصویری اسکن شده ارزیابی شده است که به ترتیب دقت برابر با ۹۷/۵٪ و ۸۷/۶٪ و فراخوانی برابر با ۹۲/۱٪ و ۷۹/۳٪ می‌باشد. مراحل این روش در شکل ۲-۶ نشان داده شده است. در شکل ۲-۷ بعضی از کلمات فارسی و ویژگی‌های آنها نمایش داده شده‌اند.



شکل ۲-۶: مراحل روش تحقیق [۱۸]



شکل ۲-۷: کلمات فارسی و ویژگی‌های ساختاری آنها [۱۸]

۲-۲-۲ روش‌های بدون قطعه‌بندی

روش‌های مستقل از قطعه‌بندی با استفاده از روش‌های مبتنی بر پنجره لغزان سند را مورد جستجو قرار می‌دهد. تحقیق [۱۹، ۲۰] از هر دو روش مبتنی بر قطعه‌بندی و مستقل از قطعه‌بندی بر روی دست‌نوشته‌های عربی استفاده کرده است. با توجه به نتایج [۱۹] روش‌های مستقل از قطعه‌بندی دقتی در حدود ۵ تا ۱۰ درصد بهتر از روش‌های مبتنی بر قطعه‌بندی در عربی را نشان می‌دهد.

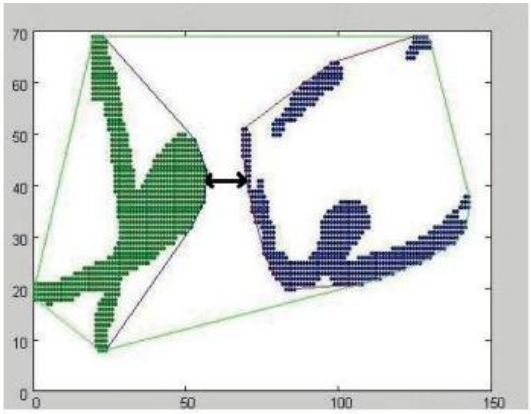
در روش مبتنی بر قطعه‌بندی در تحقیق [۱۹]، ابتدا مولفه‌های متصل برای هر خط استخراج می‌شوند سپس مولفه‌های متصلی مانند نقاط به نزدیکترین مولفه‌ی متصلی که بدنه‌ی اصلی آن می‌باشد تخصیص داده می‌شود. در مرحله‌ی بعدی این مولفه‌های متصل خوشه‌بندی می‌شوند. در زبان عربی بسیاری از کلمات با حرف "الف" آغاز می‌شوند به همین دلیل، وجود یا عدم وجود آن می‌تواند یک شاخص قوی باشد که آیا دو خوشه‌ی مجاور مربوط به زیرکلمات یک کلمه هستند یا هر کدام یک کلمه‌ی جداگانه محسوب می‌شوند. ۹ ویژگی از هر دو خوشه‌ی مجاور استخراج می‌شود. از شبکه عصبی برای تعیین اینکه فاصله بین دو خوشه‌ی مجاور، فاصله بین کلمات یا زیرکلمات است استفاده می‌شود. شکل ۲-۸ دو خوشه‌ی مجاور را نشان می‌دهد. این دو خوشه برای تشخیص اینکه مربوط به یک کلمه هستند یا دو کلمه‌ی جدا از هم، مورد آزمون قرار می‌گیرند.

این ۹ ویژگی عبارت‌اند از: طول اولین و دومین خوشه، تفاوت بین کادر محاط بر کلمه[1] دو وجود یا عدم وجود حرف "الف" در خوشه‌ی اول و در خوشه‌ی دوم، تعداد مولفه‌های متصل خوشه‌ی اول و خوشه‌ی دوم، کمترین فاصله بین بدنه‌ی محدب[2] دو خوشه، نسبت بین مجموع بدنه‌ی محدب دو خوشه به بدنه‌ی محدبی که دربرگیرنده‌ی هر دو خوشه می‌باشد.

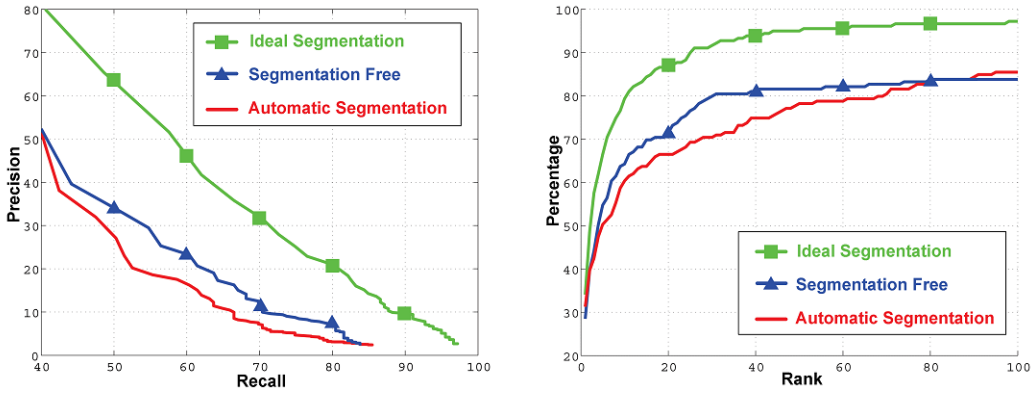
در روش مستقل از قطعه‌بندی، به جای تصویر کلمه کاندید، تصویر خط بعنوان ورودی است. خط به قطعاتی براساس الگوریتمی شبیه به الگوریتم قطعه‌بندی مبتنی بر دو یا چند حرف متصل به هم[3] در تحقیق [۲۱] تقسیم می‌شود. در واقع نقاط قطعه‌بندی روی خط مشخص می‌شوند. همه ترکیبات واقعی اجزاء متصل مجاور بعنوان نواحی بالقوه که کلمه مطلوب ممکن است نمودار شود، در نظر گرفته شده است. روش بکار رفته در پویش خط[4] یک پنجره لغزان را استفاده می‌کند، که از سمت چپ خط شروع شده و به سمت راست حرکت می‌کند. برای کلمه‌ی پرسشی داده شده یک حداقل طول و یک حداکثر طول بر اساس حداقل و حداکثر طول مجموع کاراکترها در نظر گرفته می‌شود.

الگوریتم پویش از هر نقطه قطعه‌بندی p روی هر خط شروع می‌کند. برای یک نقطه داده شده p_i اگر فضای افقی سمت چپ نقطه قطعه‌بندی وجود دارد، بعنوان یک نقطه شروع معتبر و یا بطور مشابه اگر فضای افقی سمت راست نقطه قطعه‌بندی وجود دارد، بعنوان یک نقطه پایانی معتبر درنظر گرفته می‌شود. الگوریتم، کلمات کاندید را طوری در نظر می‌گیرد که در رنج قطعات بین دو نقطه قطعه‌بندی p_i و p_j قرار داشته باشد که p_i یک نقطه شروع معتبر و p_j یک نقطه پایانی معتبر است. مقایسه نتایج بین این دو روش در شکل ۲-۹ نشان داده شده است.

[1] Bounding Box
[2] Convex Hull
[3] Ligature-based segmentation algorithm
[4] Line scanning



شکل ۲-۸: قطعه‌بندی کلمات عربی [۱۹]

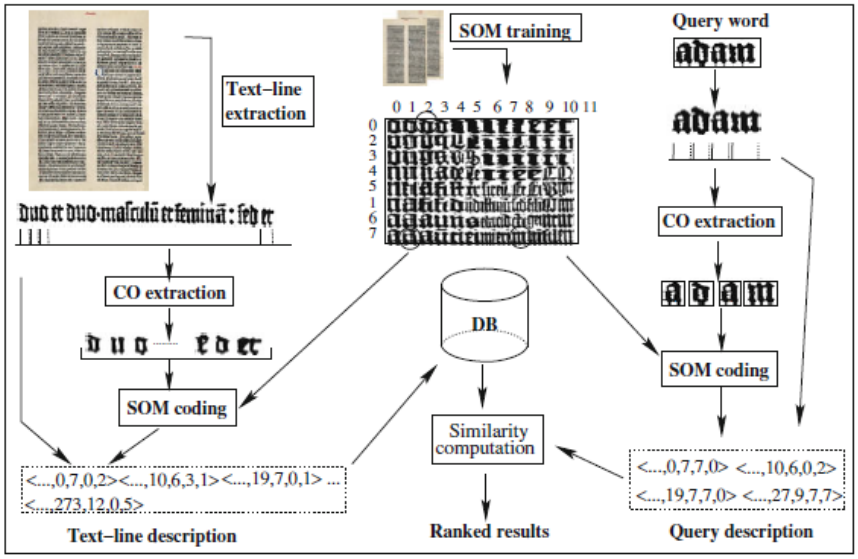


شکل ۲-۹: مقایسه دقت در دو روش مبتنی و مستقل از قطعه‌بندی در عربی [۱۹]

تحقیق [۲۲] یک روش مستقل از قطعه‌بندی خط و کلمه در استاد چاپی قدیمی را ارائه می‌کند.

استفاده از خوشه‌بندی مولفه‌های متصل، توسط الگوریتم نقشه خود سازمانده بعنوان اولین مرحله در

اندیس‌گذاری پیشنهاد شده است. بعد از آموزش توسط SOM، خطوط با در نظر گرفتن اطلاعات زیر برای هر مولفه‌ی متصل موجود در خط توصیف می‌شوند. این اطلاعات شامل شماره صفحه، سطر و ستون مولفه‌ی متصل در خط، مختصات مکانی X و Y نزدیک‌ترین نرون شبکه‌ی SOM به مولفه‌ی متصل و مختصات کادر محاط بر مولفه متصل است. سپس نمونه‌های کلمه‌ی پرسشی در سند با در نظر گرفتن یک الگوریتم پیچش زمان پویا[1](DTW) تغییر یافته که عرض کلمه را در محاسبه فاصله دربرمی‌گیرد، مشخص شده‌اند. مراحل این روش در شکل ۲-۱۰ نمایش داده شده است.

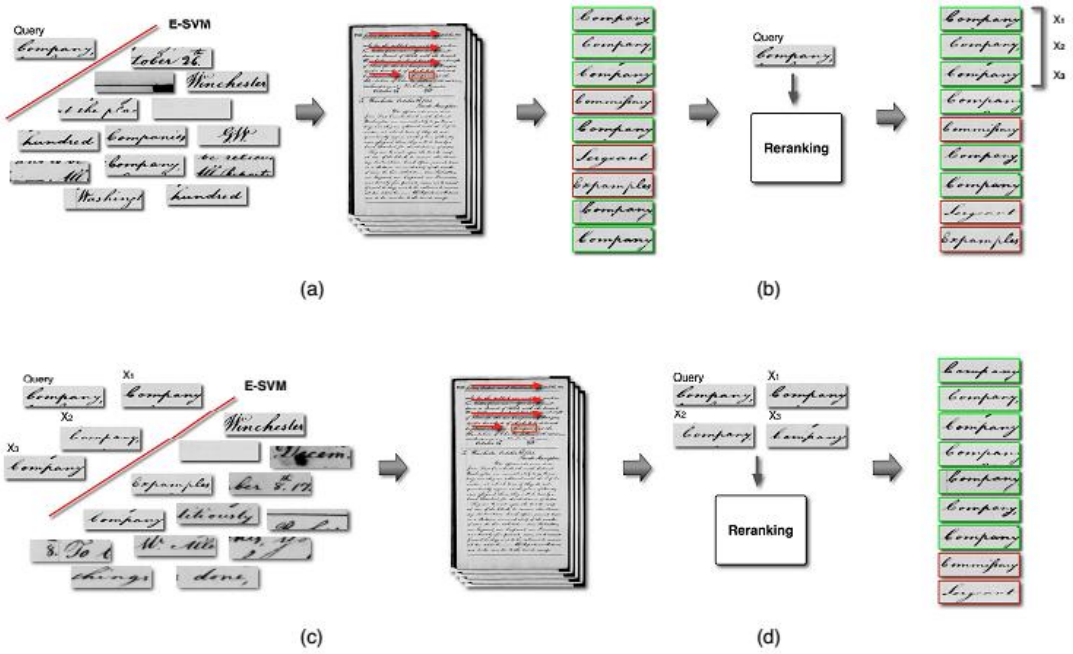


شکل ۲-۱۰: مراحل تحقیق [۲۲]

در کاری که اخیراً در [۲۳] ارائه شده است از شبکه‌ای از توصیفگرهای هیستوگرام گرادیان[2] و رهیافت پنجره لغزان برای سند و کلمه مورد جستجو استفاده شده است. برای بازنمایی بهتر کلمه مورد جستجو در روش بدون ناظر از ماشین بردار پشتیبان مثالی[3] و از یک بازنمایی با قابلیت تمایز بهتر به اسم بردار فیشر[4] [۲۴] برای رتبه‌بندی مجدد بهترین نواحی بازیابی شده، استفاده شده است. با

[1] Dynamic Time Wrapping (DTW)
[2] Histogram of Gradient (HOG)
[3] Exemplar SVM
[4] Fisher Vector

توجه به فشرده‌سازی صورت گرفته تعداد اسناد زیادی می‌توانند همزمان در حافظه رم نگهداری شوند. و با توجه به اینکه فاصله بین توصیفگرهای هیستوگرام گرادیان کوانتیزه شده می‌توانند از پیش محاسبه‌شده باشند، پنجره لغزان بسیار سریع‌تر عمل می‌کند. مراحل این سیستم در شکل ۲-۱۱ نشان داده شده است.



شکل ۲-۱۱: شمای کلی روش معرفی شده در [۲۳]

۲-۳ تقسیم‌بندی براساس روش تشخیص

روش‌های جستجوی کلمه در برخی از مقالات براساس دو رهیافت کلی تطبیق قالب و مبتنی بر یادگیری دسته‌بندی می‌شوند. روش تطبیق تصویر می‌تواند در سطح پیکسل یا ویژگی انجام شود. در روش تطبیق پیکسل به پیکسل فاصله بین پیکسل‌های متناظر تصویر کلمه‌ی پرسشی و تصاویر کلمات داخل اسناد محاسبه می‌شود. در روش تطبیق مبتنی بر ویژگی، تعدادی از ویژگی‌ها استخراج شده و بصورت یک بردار ارائه شده است و شباهت بین تصاویر با مقایسه بردارهای آنها تعیین می‌گردد.

روش‌های مبتنی بر پیکسل، یک تصویر کلمه را بوسیله ویژگی‌های سراسری مبتنی بر پیکسلِ سطح تصویر مانند خودهمبستگی شدت نور[1] و ممان‌هایی که برای نمایش هر تصویر بکار می‌روند پردازش می‌کنند. این روش‌ها برای اسناد با کیفیت پایین مناسب هستند و می‌توانند روی زبان‌های مختلف کار کنند. از طرف دیگر پردازش‌های سطح ویژگی وابسته به زبان هستند ولی از این جهت که به راحتی یک پرسش را تشکیل داده و اندیس گذاری می‌کنند و نیازی به آموزش ندارند سودمند می‌باشند. DTW در بسیاری از روش‌های مطرح شده به عنوان یک الگوریتم تطبیق قالب کارا معرفی‌شده است [۲۵-۲۷].

در بسیاری از تحقیقات انجام شده مبتنی بر یادگیری سعی شده است تا تصویر کلمه به صورت یک توالی از ویژگی‌هایی با طول متغیر استفاده شود و با روش‌هایی همچون مدل مخفی مارکوف[2] آن را رده‌بندی کنند. طول متغیر ویژگی بسیار انعطاف‌پذیرتر از بردارهای ویژگی با طول ثابت است و دیده شده که به دلیل اینکه بهتر می‌توانند با تغییرات مختلف شیوه نگارش و طول کلمه وفق پیدا کنند منجر به نتایج خوبی در کارهای جستجوی کلمه شده است [۲۸، ۲۹]. متأسفانه این روش دو ضعف اصلی دارد. اولاً، بسیاری از روش‌های با نظارت نمی‌توانند کلمات خارج از فرهنگ لغات[3] را جستجو کنند، یعنی، فقط تعداد محدودی کلمه‌ی کلیدی که در زمان آموزش شناخته شده است، می‌تواند به عنوان کلمه‌ی پرسشی بکار گرفته شود. ثانیاً، به دلیل اینکه این روش‌ها با توالی ویژگی‌ها سروکار دارند، محاسبه فاصله بین کلمات در زمان آزمون معمولاً بسیار آهسته است. در نتیجه رهیافت‌هایی مانند پنجره لغزان قابل به‌کارگیری نیستند و حتماً باید کلمات قطعه‌بندی شوند. این روش برای اسناد دارای چند نویسنده نتایج مطلوبی دارد ولی بهرحال پایگاه داده‌ای بزرگ برای آموزش چنین سیستمی مورد نیاز است. مدل مخفی مارکوف نیز یکی از پرکاربردترین رده‌بندهایی است که در چنین سیستم‌هایی بکار گرفته شده است.

[1] Intensity autocorrelation
[2] Hidden Markov Models
[3] Out Of Vocabulary (OOV)

۲-۳-۱روش‌های مبتنی بر تطبیق الگو

تحقیق [۳۰] روشی در زمینه‌ی جستجوی کلمه برای بازیابی اسناد دست‌نوشته‌ی اردو ارائه داده است. اردو زبان رسمی پاکستان و گروه بزرگی از ساکنین هندوستان است که بیشتر از ۱۰۰ میلیون صحبت‌کننده دارد. متون اردو شباهت زیادی به متون عربی و فارسی دارد که کلمات آن با ترکیب‌های مختلفی از زیرکلمات که کلمه‌ی جزئی[1](PWs) نامیده شده‌اند ساخته می‌شوند. کلمات جزئی هم با ترکیب‌های متفاوتی از حروف اصلی اردو ایجاد می‌شوند. چارچوپ روش بازیابی اطلاعات تحقیق [۳۰] به دو قسمت اصلی تقسیم می‌شود: شاخص‌گذاری و بازیابی

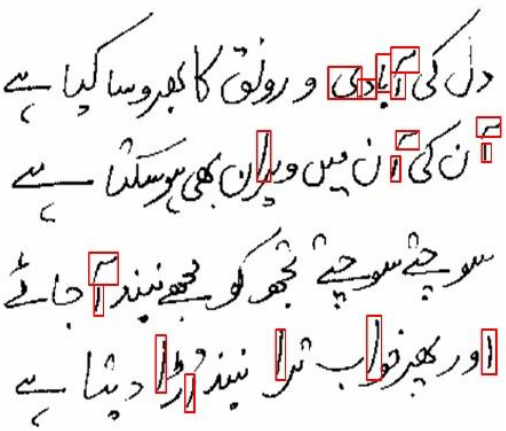
شاخص‌گذاری شامل قطعه‌بندی متن به کلمه‌های جزئی و استخراج ویژگی‌ها از هر PWs است. کلمات جزئی با یافتن اجزای متصل در یک متن باینری استخراج می‌شوند. در بازیابی، یک مقایسه‌ی چندمرحله‌ای بین PWs تصویر کلمه‌ی پرسشی و PWs موجود در سند انجام می‌شود. در مرحله‌ی شاخص‌گذاری، برای هر PWs یک مجموعه از ویژگی‌ها شامل ویژگی‌های اسکالر، بردار ویژگی‌های تعریف شده عمودی و بردار ویژگی‌های تعریف شده افقی استخراج می‌شود. ویژگی‌های اسکالر شامل نسبت طول به عرض و ناحیه محدب هر کلمه‌ی جزئی می‌باشد. ویژگی‌های تعریف شده عمودی، روی ستون یک کلمه‌ی جزئی محاسبه شده و شامل پروفایل بالا، پروفایل پایین، پروژکشن عمودی و گذر جوهر به پس‌زمینه است و ویژگی‌های تعریف شده افقی، روی سطر یک کلمه‌ی جزئی محاسبه شده و شامل پروژکشن راست و چپ است.

در مرحله‌ی بازیابی یک انطباق سه مرحله‌ای برای یافتن نمونه‌های کلمه‌ی پرسشی در پایگاه داده انجام می‌شود. اولین مرحله‌ی انطباق روی ویژگی‌های اسکالر اساسی انجام می‌شود به این صورت که تمام PWs که از نظر نسبت دید و ناحیه محدب با PWs کلمه‌ی پرسشی متفاوت هستند حذف

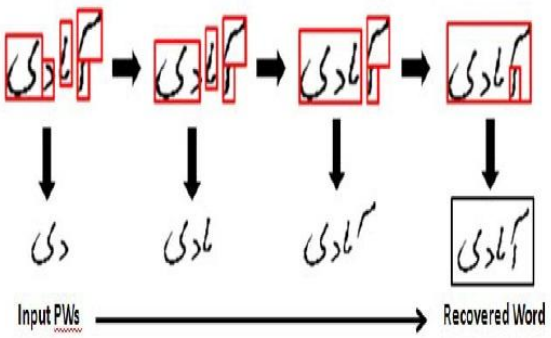
[1] Partial words(PWs)

می‌شوند. انطباق اصلی با استفاده از ویژگی‌های برداری PWs و اعمال یک DTW انجام می‌شود. در مرحله‌ی آخر انطباق، PWs که روی یک خط در متن شناسایی شده‌اند با استفاده از یک پنجره لغزان با یگدیگر ادغام شده تا زمانی که بصورت کلمه‌ی پرسشی مورد نظر ظاهر شوند.

این سیستم روی یک پایگاه داده با ۹۰ تصویر دست‌نوشته که شامل متوسط ۲۰۰ کلمه بر تصویر است ارزیابی شده است. در مرحله‌ی آزمون ۱۱۵ کلمه‌ی پرسشی انتخاب شده است. پروژکشن عمودی به عنوان موثرترین ویژگی است که در حدود ۸۴٪ انطباق در کلمات جزئی را به صورت صحیح انجام می‌دهد. بیشترین مقدار برای F_1 در مرحله‌ی انطباق اولیه برابر با ۷۲/۷٪ گزارش شده است. در این سیستم کلمات پرسشی شامل یک تا پنج PWs بر کلمه هستند که مثلا برای پنج کلمه‌ی جزئی کل زمان ۶/۳۶ ثانیه می‌باشد و برای یک کلمه‌ی جزئی ۲/۸۷ ثانیه طول می‌کشد. مراحل روش تحقیق [۳۰] در شکل ۲-۱۲ و شکل ۲-۱۳ نشان داده شده است.



شکل ۲-۱۲: تشخیص PWs در متن به ازای کلمه‌ی پرسشی مورد نظر [۳۰]



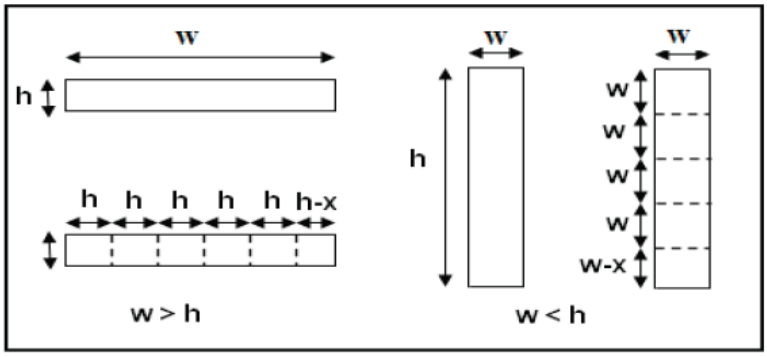
شکل ۲-۱۳: ادغام PWs با یگدیگر و یافتن نمونه‌ی کلمه‌ی پرسشی در متن [۳۰]

مرجع [۳۱]، اولین تحقیق در زمینه‌ی جستجو کلمات و شناسایی اندازه فونت روی اسناد چاپی فارسی می‌باشد. در این روش با استفاده از تعدادی ویژگی‌های مستقل از اندازه فونت مانند تعداد زیر-کلمات، نسبت طول به عرض در تصویر زیرکلمات و ویژگی‌های شبکه‌ای[1] زیرکلمات، یک کلید واژه فارسی توصیف، جستجو و یافت می‌شود. همچنین اندازه فونت تصویر سند شناسایی می‌شود. این دیدگاه روی یک پایگاه داده شامل ۵۰۰ تصویر سند فارسی ارزیابی شده است در حالی که نرخ شناسایی اندازه فونت ۹۴/۲٪، نرخ دقت ۹۲/۳٪ و نرخ فراخوانی ۷۶/۵٪ به دست آمده است. این دیدگاه با کمی انطباق برای اسناد عربی و اردو مناسب می‌باشد. در دیدگاه [۳۱] وقتی یک کاربر یک کلید واژه فارسی را از میان پایگاه داده GUI وارد می‌کند برای جستجو در یک تصویر سند، ابتدا نوع فونت سند تشخیص داده می‌شود. سپس کلمه‌ی پرسشی بر طبق نوع فونت اسناد بازنویسی و اصلاح شده است و تصویر آن ساخته می‌شود. در مرحله‌ی بعدی قاب‌های مرزی تصویر کلمه اصلاح شده ساخته شده و برخی از ویژگی‌های مستقل اندازه فونت که در بالا ذکر شد استخراج می‌شوند.

ویژگی‌های شبکه‌ای برای هر زیرکلمه با ناحیه‌بندی انطباقی استخراج می‌شود. در این روش ناحیه‌بندی، یک زیرکلمه با طول h و عرض w در بزرگترین جهت‌اش ناحیه‌بندی می‌شود برای مثال اگر h<w کادر محاط بر زیرکلمه بصورت افقی و از چپ به راست تقسیم‌بندی می‌شود به صورتی که

[1] Mesh feature

عرض هر ناحیه h است. و اگر w<h کادر محاط بصورت عمودی از بالا به پایین تقسیم‌بندی می‌شود بصورتی که ارتفاع هر ناحیه w است. و در هر ناحیه تعداد پیکسل‌های قلم محاسبه می‌شود. ناحیه-بندی انطباقی در شکل ۲-۱۴نمایش داده شده است.



شکل ۲-۱۴: ناحیه‌بندی انطباقی و استخراج ویژگی شبکه‌ای در تحقیق [۳۱]

برای انطباق و جستجوی ویژگی‌های استخراج شده کلمه مورد پرسش در سند تصویری و یافتن موارد مشابه، از انطباق چند سطحی استفاده شده است. به این صورت که در اولین مرحله تعداد زیرکلماتِ کلمه مورد پرسش و نسبت طول به عرض آنها در نظر گرفته شده و نمونه‌های یکسان در سند تصویری پیدا می‌شود. مرحله‌ی بعدی انطباق، فقط روی نمونه‌های کلمات یافت شده از مرحله‌ی قبل اعمال می‌شود. این کار عملیات محاسباتی را کاهش می‌دهد زیرا با اعمال هر نوع از ویژگی‌ها تعدادی کلمه‌ی نامرتبط کنار گذاشته شده و حوزه‌ی جستجو کوچک و کوچکتر می‌شود. بعد از این مرحله تعداد اجزاء هر زیرکلمه و ویژگی‌های شبکه‌ای اعمال شده و کلمات داخل سند تصویری که همه‌ی ویژگی‌های ذکر شده متناظر با ویژگی‌های کلمه مورد پرسش را دارا می‌باشند به عنوان جواب صحیح در نظر گرفته می‌شوند. در این تحقیق به منظور استخراج زیرکلمات از کلمه‌ی پرسشی و کلمات داخل سند تصویری از پروفایل پروژکشن عمودی استفاده شده است. در حین ارزیابی این دیدگاه مشاهده شد که دلیل اصلی خطا این دیدگاه مربوط به قطعه‌بندی خطوط متن و قطعه‌بندی کلمات به زیرکلمات است. این مساله به دلیل روی هم افتادگی قاب مرزی برخی زیرکلمات اصلی منفصل در جهت عمودی است.

در تحقیق [۳۲] یک فرآیند کامل برای مکان‌یابی کلمات در تصاویر اسکن شده‌ی کتب تاریخی دستنویس فارسی و عربی ارائه شده است. دراین فرآیند، ابتدا زیرکلمات موجود درصفحات کتاب استخراج شده و اطلاعات مکانی این زیرکلمات ذخیره می‌شود. درادامه دو نوع ویژگی شامل پروفایل‌ها و ناحیه‌بندی از تصاویر زیرکلمات استخراج می‌شود. در مرحله‌ی بعد، با در نظر گرفتن معیار شباهت بین این ویژگی‌ها بصورت ترکیبی ازمعیارهایDTW و فاصله اقلیدسی ، زیرکلمات در خوشه های هم ارز طبقه‌بندی می‌شوند. درنهایت هم، خوشه‌ها برچسب‌گذاری می‌شوند. حال اگر کاربر، کلمه‌ای را درون کتاب مورد جستجو قرار دهد، با داشتن اطلاعات مربوط به تمام زیرکلمات موجود در خوشه ها، مکانهای احتمالی وقوع آن کلمه در کتاب به کاربر ارائه می گردد. این الگوریتم بر روی یک کتاب تاریخی دست‌نویس فارسی موجود در کتابخانه‌ی مرکزی دانشگاه بیرجند مورد آزمایش قرار گرفته است.

در تحقیق [۳۳، ۳۴] با در نظر گرفتن کلمه‌ی پرسشی و تطابق آن به صورت پیکسلی با استفاده از XOR باینری کلمه‌ی پرسشی و کلمات کاندیدا قطعه‌بندی شده از سند، تصویر باینری جدیدی حاصل می‌شود که با استفاده از نگاشت فاصله اقلیدسی[۱] میزان شباهت به دست می‌آید. ویژگی‌هایی چون مساحت و نسبت طول به عرض کلمه‌ی پرسشی در این روش مورد استفاده قرارگرفته است. البته در [۳۴] با تبدیل کلمه مورد جستجو و کلمات کاندیدا نسبت به هم سعی در پیدا کردن حداقل هزینه تطبیق بر اساس مجموع مربعات اختلافات[۲] را دارد که نسبتاً نتایج بهتری را ارائه می‌کند. بزرگ‌ترین مشکل روش ارائه شده حساس بودن زیاد به تنوع شیوه نگارش و تبدیلات آفاین[۳] است. همچنین روش‌های بیان شده بسیار وابسته به نویسنده[۴] است و با تغییر نویسنده یا تنوع آن دقت نتایج به طور

[1] Euclidean Distance Mapping (EDM)
[2] Sum of Squared Differences(SSD)
[3] Affine Transform
[4] Writer-dependent

شگفت‌آوری کاهش می‌یابد. الگوریتم اسکات و لنگوت-هیگنز[1] برای بهبود دادن نتایج در شرایطی که تبدیلات اَفاین رخ داده باشد، مورد استفاده قرار گرفت [۳۳، ۳۴]. هدف از الگوریتم اسکات و لنگوت-هیگنز محاسبه یک فاصله جدید جهت بهبود نتایج تناظر بین دو مجموعه نقاط i و j است، وقتی‌که تبدیل اَفاین بین آنها وجود دارد. این روش نسبت به روش نگاشت فاصله اقلیدسی بسیار بهتر عمل می‌کرد اما به خاطر نیاز به محاسبه SVD یک ماتریس بزرگ محاسبات بسیار کند انجام می‌پذیرفت و هنوز مشکل وابستگی به نویسنده و فونت نگارش همچنان پا برجاست. (یک تبدیل اَفاین، تبدیلی است که هم‌خطی نسبت فواصل را حفظ می‌کند. یعنی همه نقاطی که در ابتدا روی یک خط قرار گرفته‌اند، پس از تبدیل نیز همچنان روی یک خط باقی می‌مانند و مثلاً وسط پاره‌خط بعد از تبدیل همچنان نقطه میانی است. در این تبدیل خطوط مستقیم، مستقیم باقی می‌مانند و موازی بودن خطوط حفظ می‌شود. انقباض، انبساط، تجانس، انعکاس، دوران و انتقال نمونه‌هایی از تبدیل اَفاین هستند.)

بعنوان یک تطبیق سطح پایین، انطباق در سطح پیکسل، روش فاصله Hausdorff برای تطبیق تصویر کلمه به دلیل سادگی و حساسیت نسبت به تغییرات مشخصه‌های تصویر استفاده می‌شود. فاصله Hausdorff شباهت بین پیکسل‌ها را با اندازه‌گیری اینکه چه مقدار دو مجموعه پیکسل دور از همدیگر هستند را ارائه می‌کند.

بعنوان مثال یک پردازش سطح پیکسلی [۳۵] تلاش کرد با استفاده از معیار فاصله وزن‌دار Hausdorff[2] کلمه مورد نظر کاربر را با تصاویر کلمات کاندیدا قطعه‌بندی شده از سند تطبیق دهد. در طول پیش‌پردازش فضای بین حروف همسایه در هردوی تصویر کلمه موردنظر و تصاویر کلمه استخراج شده از سند حذف شد. سپس تصاویر کلمه به قسمت‌های بالارونده، پایین‌رونده و ناحیه میانی تقسیم شده و وزنهای مختلفی به آنها اختصاص داده می‌شود. بعد از آن یک فاصله Hausdorff وزن‌دار برای اندازه‌گیری فاصله بین تصویر قالب و تصاویر کلمات کاندید بکار می‌رود. در صورتیکه این فاصله از یک

[1] Scott and Longuet-Higgins (SLH)
[2] Weighted Hausdorff Distance(WHD)

آستانه کمتر باشد تصویر کلمه‌ی کاندید بعنوان نمونه‌ی کلمه‌ی پرسشی در سند مشخص می‌شود. این روش به نتایج مناسبی روی اسناد چاپی چینی و لاتین دست یافته است.

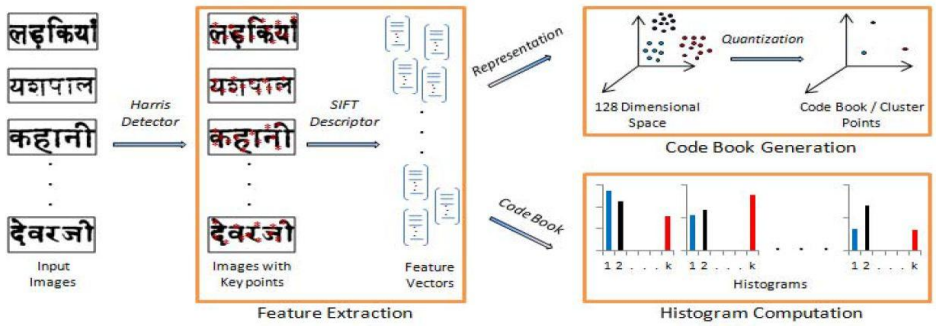
۲-۳-۲ روش‌های مبتنی بر یادگیری

یک مدل رایج در بازیابی اطلاعات برای بازنمایی اسناد رهیافت کیف بصری کلمات[1](BOVW) است [۶۳، ۳۷]. رهیافت کیف بصری کلمات سه مرحله اصلی را شامل می‌شود: در ابتدا تعداد خاصی نقاط کلیدی محلی بطور اتوماتیک بوسیله یک تشخیص‌دهنده مناسب از تصویر استخراج می‌شود. نقاط کلیدی، نقاط برجسته تصویر هستند که محتوای اطلاعاتی زیادی دارند از اینرو برای توصیف کل تصویر مناسب هستند. در مرحله دوم نقاط کلیدی، خوشه‌بندی شده و توصیف‌گرهای مشابه به یک خوشه اختصاص داده می‌شوند. هر خوشه متناظر با یک کلمه‌ی بصری[2] است که یک نمایش، از ویژگیهای به اشتراک گذاشته شده توسط توصیف‌گرهای متعلق به این خوشه است. مجموعه خوشه می‌تواند بعنوان یک فرهنگ لغت بصری تفسیر شود. در مرحله‌ی آخر هر تصویر بوسیله یک بردار شامل تعداد رخدادهای هر کلمه بصری در تصویر توصیف شده است. مهم‌ترین نقطه این رهیافت تشخیص نقاط کلیدی محلی (بعنوان مثال SIFT یا نقاط گوشه‌ای) است. در مورد روش خوشه‌بندی، K میانگین، الگوریتم K نزدیکترین همسایه و آنالیز معنایی نهفته آماری[3] و ماشین بردار پشتیبان[4](SVM) مشهورترین روش‌ها هستند [۳۸].

تحقیق [۳۹] از همین روش برای بازیابی اسناد تصویری روی نوشته‌های هندی استفاده کرده است. الگوریتم خوشه‌بندی مورد استفاده k میانگین و ویژگی‌های مورد استفاده SIFT می‌باشند. یکی از محدودیت‌های روش BOVW در مورد اسناد تصویری در نظر نگرفتن موقعیت مکانی کلمات بصری

[1] Bag Of Visual Words (BOW)
[2] Visual Word
[3] Probabilistic Latent Semantic Analysis (pLSA)
[4] Support Vector Machine (SVM)

می‌باشد به عبارت دیگر ترتیب کلمات بصری در نظر گرفته نمی‌شود. برای حل این مساله، در تحقیق [۳۹] هر تصویر کلمه در راستای ستون‌ها به سه ناحیه تقسیم می‌شود. سپس برای هر کلمه‌ی پرسشی میزان انطباق ناحیه‌ی اول آن، با تمام نواحی اول کلمات موجود در پایگاه داده محاسبه می‌شود. برای نواحی دوم و سوم هم به همین صورت است. برای هر کلمه‌ی پرسشی مجموع میزان انطباق روی کل تصویر و هر سه ناحیه، با تصاویر کلمات موجود در پایگاه داده در نظر گرفته می‌شود. به ازای هر تصویر موجود در پایگاه داده که میزان انطباق بیشتر باشد آن تصویر به عنوان نمونه‌ی کلمه پرسشی آشکار می‌شود. این روش روی پایگاه داده‌ی بزرگی شامل ۱۰۰۰۰۰ تصویر کلمه در ۴ نوع زبان مختلف هندی ارزیابی شده است که دقت بیش از ۷۵ ٪ حاصل شده است. شمای کلی این روش در شکل ۲-۱۵ مشخص شده است.



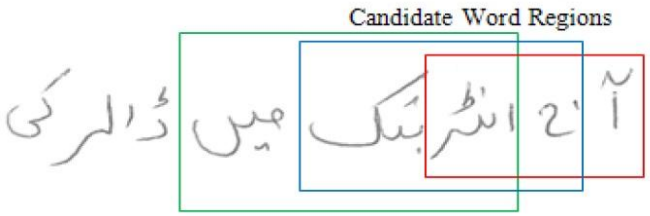
شکل ۲-۱۵: بازیابی اسناد تصویری هندی با روش BOVW [۳۹]

تحقیق [۴۰]، از ویژگی گرادیان جهت‌دار و ویژگی‌های ساختاری مانند پروفایل بالا، پروجکشن عمودی و افقی برای جستجوی کلمات در متون دست‌نوشته‌ی اردو استفاده کرده است. روش ارائه شده در این تحقیق مبتنی بر جداسازی خطوط و استخراج مولفه‌های متصل اصلی هر خط می‌باشد. با حرکت یک پنجره لغزان روی هر خط، ترکیب‌های یک تا چهارتایی از مولفه‌های متصل نزدیک به هم، به عنوان مجموعه کلمات کاندید سند ایجاد می‌شوند. مجموعه کلمات کاندید ایجاد شده توسط پنجره‌ی لغزان در شکل ۲-۱۶ نشان داده شده است. سیستم بازشناسی کلمه‌ی اردو به طور مؤثری طراحی شده و برای رده‌بندی کلمات کاندید شده به کار می‌رود. از طبقه‌بند SVM، برای کلاس‌بندی

این مجموعه استفاده شده است. به منظور افزایش کارایی سیستم، یک مرحله تایید یا عدم تایید انجام می‌شود. در واقع این مرحله برای تشخیص درست بودن یا نبودن کلاس بدست آمده برای کلمات کاندید با توجه به خروجی‌های SVM است. در این مرحله موارد زیر بررسی می‌شوند:

۱)احتمال بالا خروجی SVM ۲)تعداد نقاط ۳)تعداد پیکسل‌های سیاه ۴)فاصله‌ی اقلیدسی بین بردار ویژگی کلمه‌ی کاندید با میانگین کلاس مورد نظر

در این روش دقت ۵۰/۷۵٪، فراخوانی ۷۰/۱٪ و نرخ قطعه‌بندی صحیح ۹۲/۱۱٪ گزارش شده است.

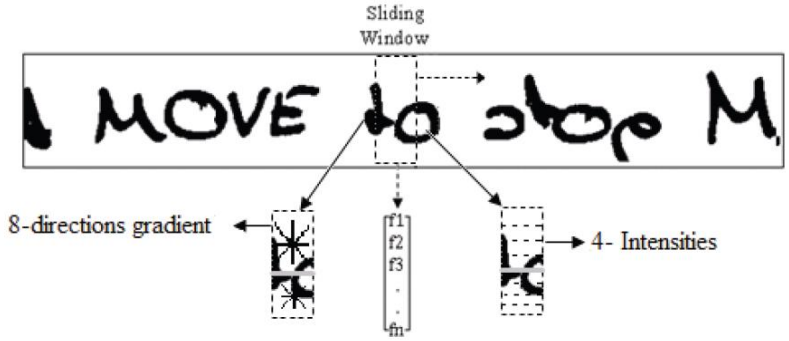


شکل ۲-۱۶: مجموعه کلمات کاندید ایجاد شده توسط پنجره‌ی لغزان [۴۰]

در تحقیق [۴۱] روش جدیدی برای تشخیص کلمات در متون دست‌نوشته بر پایه مدل‌های مخفی مارکو ارائه شده است. در این تحقیق HMM ها، با حروف، آموزش یافته و برای شبیه‌سازی هر کلمه‌ی پرسشی با هم ترکیب می‌شوند و از مدل‌های فیلتر برای ارائه بهتر تصاویر نواحی کلمات غیر-پرسشی استفاده شده است. این سیستم این قابلیت را دارد که برای کلمات بسیاری بدون نیاز به قطعه‌بندی کلمات و حروف عمل کند و قابل اعمال به زبان‌های مختلفی از جمله انگلیسی، عربی و دوانگاری است. سیستم بر روی پایگاه‌های داده از زبان‌های مختلف شامل IAM برای انگلیسی، AMA برای عربی و LAW برای دوانگاری[1] ارزیابی شده است. روی هر خط متن، پنجره‌های لغزانی به نام فریم در نظر گرفته شده است. طول این فریم‌ها ۲۰ پیکسل است و دارای ۸۵ ٪ هم‌پوشانی می‌باشند.

[1] Devanagari

برای هر فریم ویژگی‌های گرادیان و سطح روشنایی[1] استخراج می‌شوند. هر فریم بر مبنای مرکز پیکسل‌های قلم به دو بلوک ناحیه‌بندی شده و برای هر بلوک ویژگی‌های گرادیان در ۸ جهت استخراج می‌شوند. برای استخراج ویژگی‌های سطح روشنایی هر بلوک به ۴ ناحیه تقسیم شده و این ویژگی با شمارش تغییرات پیکسل‌های سیاه به سفید بدست می‌آید. این ویژگی‌ها در شکل ۲-۱۷نشان داده شده‌اند.

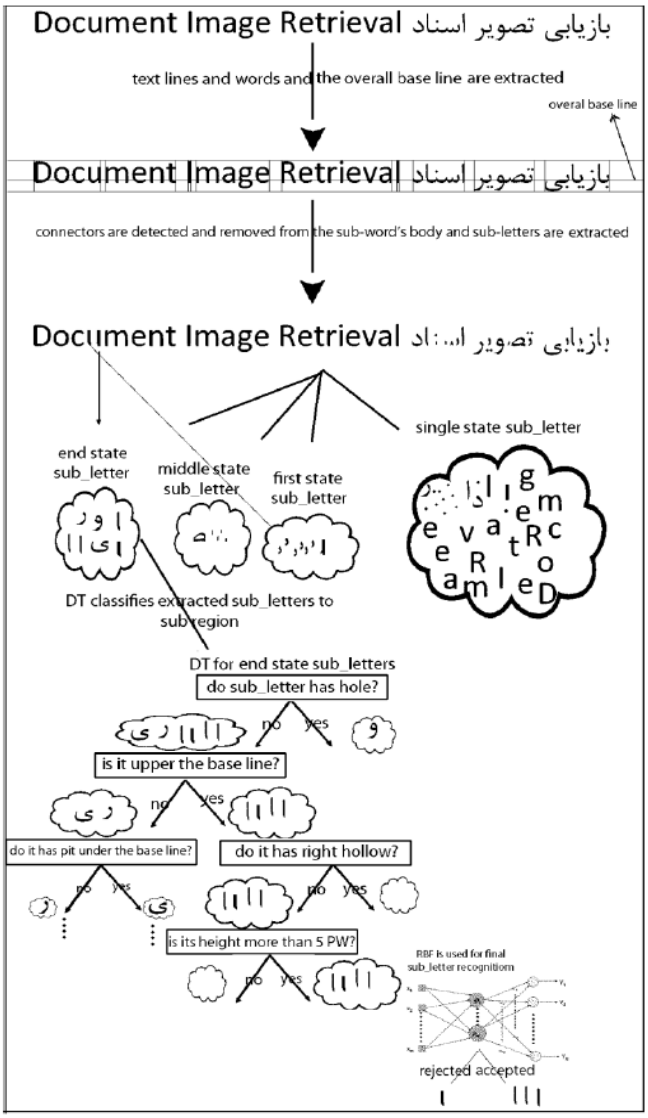


شکل ۲-۱۷: استخراج ویژگی‌های گرادیان جهت‌دار و سطح روشنایی تحقیق [۴۱]

در [۴۲] یک روش بازیابی برای اسناد فارسی/عربی مستقل از تشخیص ارائه شده است. این سیستم در متون ترکیبی فارسی/عربی و انگلیسی می‌تواند مورد استفاده قرار گیرد. متد شامل پیش-پردازش، استخراج کلمه و زیرکلمه، تشخیص و حذف متصل کننده‌های زیرحروف، تفسیر زیرحروف با کدگذاری شکل، رده‌بندی زیرحروف با استفاده از درخت تصمیم و شبکه عصبی RBF برای تشخیص زیرحرف است. این سیستم اسناد تصویری را با یک شمای جدید کدگذاری شکل زیرحروف در اسناد فارسی/عربی بازیابی می‌کند. رده‌بند مبتنی بر درخت تصمیم‌گیری فضای زیرکلمات را با استفاده از ویژگیهای شکل مکانی به تعدادی زیرناحیه تقسیم می‌کند. ویژگیهای شکل مکانی شامل طول، عرض، حفره‌ها، گودال‌ها، سوراخ‌ها، دندانه‌ها و زیرحروف بالارونده و پایین‌رونده است. در روشهای مبتنی بر قطعه‌بندی علاوه بر مسائلی مانند نویز، تنوع فونت و اندازه، خصوصا برای فارسی و نوشتار پیوسته، مشکل قطعه‌بندی نیز وجود دارد. به علت تنوع شکل، اندازه و طول حروف، تخمین نقاط قطعه‌بندی

[1] intesity

حروف پر اشتباه انجام می‌شود. مراحل این روش در شکل ۲-۱۸ نشان داده شده است.



شکل ۲-۱۸: مراحل روش تحقیق [۴۲]

۲-۴ تقسیم‌بندی براساس نوع کلمه‌ی پرسشی

در تحقیقات انجام شده در زمینه‌ی جستجوی کلمه، اکثراً کلمه‌ی پرسشی را بـه صـورت کلمـه‌ی تصویری(QBE) در نظر می‌گیرند که در بیشتر موارد از خود سند اصلی مورد جستجو بریده شده اسـت و برخی آن را به صورت متن (QBT) در نظر گرفته‌اند که کاربر آن را در کادر مربوط به صـورت متنـی

وارد می‌نماید.

روند جستجوی کلمات دست‌نوشته به صورت QBE می‌تواند به صورت یک مسئله کلاسیک بازیابی تصویر فرمول‌بندی شود. تصویر کلمه‌ی پرسشی با مجموعه‌ای از تصاویر کاندیدا در پایگاه داده مقایسه شده و سیستم شبیه‌ترین تصاویر را برمی‌گرداند یا بصورت کلی‌تر تمام کاندیدها را با توجه به میزان شباهت به تصویر کلمه‌ی پرسشی بصورت نزولی رتبه‌بندی می‌کند. از جمله معایب مهم این روش می‌توان به این مسئله اشاره کرد که کاربر برای یافتن یک نمونه از کلمه مورد جستجو و مشخص کردن محدوده آن، سند را به صورت دستی مورد جستجو قرار دهد. دو مولفه مهم سیستم QBE نمایش (تصویر توصیف‌گر) و انطباق (معیار شباهت بین توصیف‌گرها) می‌باشد.

این مسئله از QBT متمایز است چرا که در QBT ورودی رشته متنی است که باید جستجو شود. هر یک از سیستم های QBT و QBE مزایای خودشان را دارند. از طرفی سیستم های QBT تمام مزایای سیستم‌های شناسایی دستخط مثل انعطاف‌پذیری در جستجوی هر کلمه‌ی کلیدی یا امکان استفاده از حجم زیاد داده‌های برچسب‌دار در دسترس را دارند. QBE بطور خاص برای مواردی که داده‌های برچسب‌دار در دسترس نیستند یا با هزینه بالا بدست می‌آیند مناسب است. از طرف دیگر تحقیق [۴۳] نشان می‌دهد که رهیافت‌های کلمه پرسشی تصویری منحصراً بر اساس تطبیق تصویر است یعنی فاز آموزش با نظارت را نیاز ندارد و می‌تواند دقت کافی موردنیاز سناریوهای عملی را به دست آورد. با توجه به مقایسه فوق انتظار می‌رود که روش‌های مبتنی بر کلمه پرسشی متنی (بازیابی کلمه) با توجه به توانایی بازیابی فرم‌ها و نوشتارهای مختلف از کلمه ورودی، فراخوانی بهتری داشته باشد. در ادامه تعدادی از تحقیقات که شامل هر دو روش QBE و QBT هستند بیان خواهند شد.

تحقیق [۴۴] از روش تعبیه مشخصه‌ها[1] به عنوان روشی در حوزه‌ی اسناد تصویری و بازشناسی کلمات کلیدی استفاده نموده که به نتایج مناسبی در زبان لاتین دست یافته است. در این تحقیق

[1] Attribute

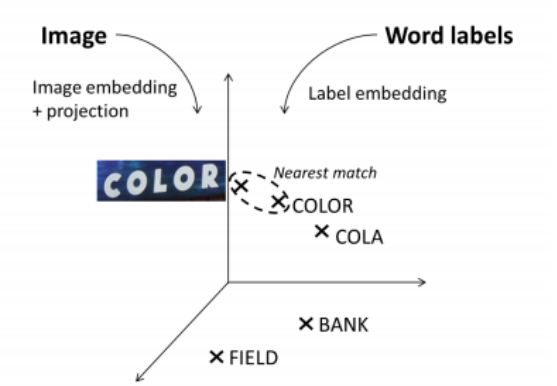
کلمه مورد جستجو به هر دو صورت متن یا تصویر می‌تواند باشد. در این روش برای هر کلمه‌ی پرسشی، تصویر آن کلمه و رشته‌ی متنی[1] مربوط به آن در یک فضای برداری مشترک تعبیه شده‌اند که در این فضا تصاویر و رشته‌ای که نشان‌دهنده‌ی کلمه پرسشی یکسان هستند نزدیک به یکدیگرند و این اجازه را می‌دهد که عملیات بازشناسی به صورت مسئله نزدیک‌ترین همسایگی انجام شود.

در این تحقیق یک نمایش رایج برای تصاویر کلمات و رشته‌های متنی متناظر با آنها معرفی شده است. برای بازنمایی رشته‌های متنی، یک هیستوگرام باینری از حروف ساخته می‌شود که هر بُعد نشان می دهد آیا رشته‌ی متنی حرف خاصی را در خود دارد یا نه. برای ایجاد تمایز بین کلماتی که دارای حروف مشابه هستند از روش هرمی استفاده شده که به جای پیداکردن حروف در کل کلمه روی نواحی مختلف کلمه تمرکز کرده است. هیستوگرام هرمی حروف[2] (PHOC) به عنوان منبع اصلی مشخصه استفاده می‌شود. برای نمایش تصاویر هم از بردارهای فیشر استفاده شده است.

این بردارهای ویژگی به همراه برچسب‌های PHOC برای آموزش مدل‌های مشخصه بر پایه SVM استفاده می‌شوند. به عبارت دیگر این تحقیق یک بازنمایی با طول ثابت و ابعاد کم از کلمه را معرفی کرده که به سرعت قابل محاسبه است به نحوی که این بازنمایی منجر به یک فضای برداری مشترک بین بازنمایی تصویر و متن کلمه می‌گردد و از یک روش کالیبره کردن به نام آنالیز همبستگی کانونی[3] (CCA) برای تصحیح مقادیر ویژگی‌ها بهره برده که منجر به اصلاح بهتر نتایج شده است. این تحقیق بر روی تصاویر کلمات بریده شده انجام شده و هدف آن دستیابی به نمونه‌ای است که در فرهنگ لغت بیشترین شباهت به کلمه‌ی مورد پرسش را دارد. یکی از موارد جالب این است که این روش توانایی بازشناسی کلمات دیده نشده را نیز تا حدی داراست.

[1] Text strings
[2] Pyramidal Histogram Of Characters(PHOC)
[3] Canonical Correlation Analysis (CCA)

روشی بر اساس تعبیه کردن رشته در زبان انگلیسی در [۴۵] پیشنهاد شده است که این بازنمـایی در یک مفهوم تعبیه کردن برچسب نه به عنوان منبعی از مشخصه‌ها اسـتفاده شـده اسـت. هـدف ایـن است که یک بازنمایی مشترک[1] برای تصویر و رشته‌ی متنی کلمه وجود داشته باشد که مشخصـه‌هـای استخراج شده از رشته و مشخصه‌های پیش‌بینی‌شده از تصویر در یک زیرفضای مشترک قرار می‌گیرنـد و می‌توانند مقایسه شوند. شمای کلی این سیستم در شکل ۲-۱۹ نمایش داده شده است.



شکل ۲-۱۹: شمای کلی سیستم [۴۵]

در [۱۲] سیستمی برای تشخیص کلمه در نوشته‌های لاتین، دوانگاری و عربی توصیف شده است. روش جستجوی کلمه در این تحقیق شامل (۱) قطعه‌بندی کلمه و خط (۲) استخراج ویژگی است. برای جداکردن خطوط از خوشه‌بندی[2] و برای جدا کردن کلمات، فاصله بین کلمات بعنوان یک مسئله رده‌بندی که آیا این فاصله بین دو کلمه است یا اجزای یک کلمه، عمل شده است. از شبکه عصبی با ویژگی‌هایی از زیرکلمات برای این رده‌بندی استفاده شده است.

ویژگی‌هایی بکار رفته، ویژگی‌های دودویی مبتنی بر گرادیان (گرادیان، سـاختاری و تعقـر) اسـت. که در مقیاسهای بزرگ، متوسط و محلی اندازه‌گیری می‌شود. ویژگی‌ها تحت یک تقسیم‌بنـدی ۴ در ۸ استخراج شده و شامل ۳۸۴ بیت ویژگی گرادیان، ۳۸۴ بیت ویژگی ساختاری و ۲۵۶ بیت ویژگی تعقـر

[1] Joint representation
[2] Clustering

می‌باشد؛ که یک بردار ویژگی دودویی به طول ۱۰۲۴ را می‌دهد. گرادیان، ویژگی‌های هندسی سطح میانه، شامل وقوع گوشه‌ها و خطوط در جهات مختلف را نشان می‌دهد. ویژگی‌های ساختاری شکل درشت‌تر کلمه را نمایش می‌دهد که وجود گوشه‌ها، خطوط قطری، عمودی و افقی در تصویر گرادیان با استفاده از ۱۲ قانون تعیین شده در [۴۶] را نشان می‌دهد. ویژگی تعقر نشان‌دهنده‌ی مشخصات هندسی و مکانی اصلی شامل جهت طاق‌نماها[1]، وجود حفره‌ها و سرکش‌های افقی یا عمودی بزرگ است. جهت شناسایی کلمه از معیار شباهت همبستگی[2] استفاده شده است. این معیار، شباهت بین دو تصویر کلمه را که شکل آنها توسط بردار ویژگی باینری نمایش داده می‌شود محاسبه می‌کند.

روش فوق روی مجموعه داده‌هایی از هر سه زبان تست شده و ادعا شده که نتایج روی متون چاپی بسیار خوب بوده است. این تحقیق در متون دست‌نوشته‌ی عربی به دقت ۸۰ ٪ رسیده است. برای حالت پرسش با متن کاربر می‌تواند کلمه‌ی عربی مورد جستجو را با حروف انگلیسی تایپ کند و الگوریتم بر اساس یک سیستم فرهنگ لغت که لغات تصویری سند را شاخص‌گذاری کرده است، کلمات عربی موردنظر را پیدا می‌کند.

[1] Bay
[2] Correlation Similarity

فصل سوم:

مباحث نظری

روش پیشنهادی این پایان نامه یک رهیافت جستجوی کلمه برای دست‌نوشته‌های زبان فارسی است که هدف آن یافتن یک کلمه پرسشی در یک مجموعه داده شامل اسناد تصویری است. به همین منظور از روش پیشنهاد شده توسط Almazan و همکارانش استفاده شده است[۴۴]. در این فصل روش تحقیق [۴۴] که اساس این پایان نامه نیز می‌باشد به تفضیل بیان خواهد شد. همچنین ابتدا به تئوری مباحث استفاده شده در این روش که شامل استخراج ویژگی از کلمات در سند، تشکیل بردار فیشر و .... است پرداخته می‌شود.

۳-۱ استخراج ویژگی‌های *SIFT*

یکی از مهم‌ترین مراحل در سیستم‌های بازشناسی الگو، مرحله استخراج ویژگی می‌باشد. هدف از استخراج ویژگی فراهم کردن توصیفگرهای مناسب برای بیان یک تصویر می‌باشد. این توصیفگرها در قالب یک بردار ویژگی در مرحله کلاسه‌بندی و بازشناسی به کار می‌رود.

استخراج نقاط کلیدی از تصویر یک شی که نماینده‌های خوبی برای توصیف آن باشند و در نماهای مختلف دارای پایداری بوده و بازشناسی مناسبی را تحقق بخشند از چالش‌های بزرگ در حوزه بینایی ماشین است. کالیبره کردن دوربین، ثبت تصاویر، ناوبری ربات و تشخیص اشیا تنها تعدادی از کاربردهای این ویژگی‌ها می‌باشند. به طور مثال در کاربرد بازشناسی شئ، استفاده از این نقاط کلیدی را میتوان طی سه مرحله اصلی زیر انجام داد:

۱) یافتن نقاط کلیدی: این مرحله با جستجو کردن در تصویر و یافتن نقاط منحصربه‌فردی که در تبدیل‌های هندسی صفحه و یا تغییر روشنایی به عنوان نقاط منحصربه‌فرد باقی بمانند، انجام می‌شود.

۲) توصیف نقاط کلیدی: توصیف این نقاط باید به گونه‌ای باشد که در مقابل نویز محیط، تغییرات هندسی و نورپردازی، توصیف یکسانی را از نقاط کلیدی داشته و متمایز و غیرحساس باشند.

۳) مرحله‌ی آخر تطبیق این نقاط بین تصاویر مختلف شئ یا صحنه است. عموما در این مرحله از

روش‌های محاسبه‌ی فاصله، میان بردارهای ویژگی که در مرحله‌ی قبل بوجود آمدند، استفاده می‌شود.

در مقایسه‌ای که میان روش‌های مختلف توصیف ویژگی‌ها انجام شده است؛ SIFT، بارزترین توصیف را از شی ارائه داده است[۴۷]. این ویژگی‌ها نسبت به تغییر مقیاس، چرخش، بزرگنمایی و همچنین تا حدودی نسبت به تغییرات روشنایی و نویز، ثابت می‌باشند[۴۸].

الگوریتم SIFT بوسیله‌ی David Lowe در سال ۱۹۹۹ ابداع گردید. SIFT یک روش برای آشکارسازی و استخراج ویژگی‌های مستقل و مشخص از تصاویر می‌باشد که می‌تواند برای تناظریابی بین تصاویر و یا شئ با یک تصویر مورد استفاده قرار گیرد. الگوریتم SIFT غالبا برای کاربردهایی نظیر تشخیص شئ، تناظریابی تصاویر، ردیابی و ساختن منظر سه‌بعدی، بازیابی شئ در بانک‌های اطلاعاتی چند رسانه‌ای (مولتی مدیا)، ربات‌های خودکار محلی و ..... مورد استفاده قرار می‌گیرد [۴۹].

در ادامه به طور اجمالی مراحل اصلی آشکار کردن و استخراج ویژگی‌های SIFT که در سه مرحله خلاصه می‌شود ارائه می‌گردد[۵۰].

۳-۱-۱ یافتن نقاط کلیدی

یافتن نقاط کلیدی، اصلی‌ترین مرحله در این روش می‌باشد که با استفاده از مفاهیم هرم تصویری[۱] و فضای مقیاس[۲] انجام می‌پذیرد. بوسیله هرم تصویری یک تصویر در درجه وضوح مختلف نمایش داده می‌شود. با توجه به مفهوم فضای مقیاس که در آن تصویر در مقیاس‌های مختلف ساخته می‌شود میتوان مطابق شکل ۳-۱ ابتدا تصویر را در چند مقیاس مختلف تشکیل داد. سپس برای هر مقیاس تصویری، پنج تصویر (اکتاو[۳]) با انحراف معیار مختلف بوجود آورد که در نتیجه به تعداد مقیاس‌های مختلف دسته‌های پنج تصویری وجود دارد. دسته‌های پنج تایی با انحراف معیارهای

[1] Image Pyramid
[2] Scale Space
[3] Octave

مختلف توسط فیلترگوسین و ضرب کرنل گوسین در تصویر اصلی بوجود آمده‌اند.

سپس همان‌گونه که در شکل ۳-۱ ملاحظه می‌گردد، هر دو تصویر مجاور در داخل دسته‌های پنج تایی را با اختلاف تفاضلی براساس جبر ماتریسی از هم کم کرده و تصاویر جدید براساس فیلتر اختلاف تفاضلی(DOG) ایجاد می‌شود.

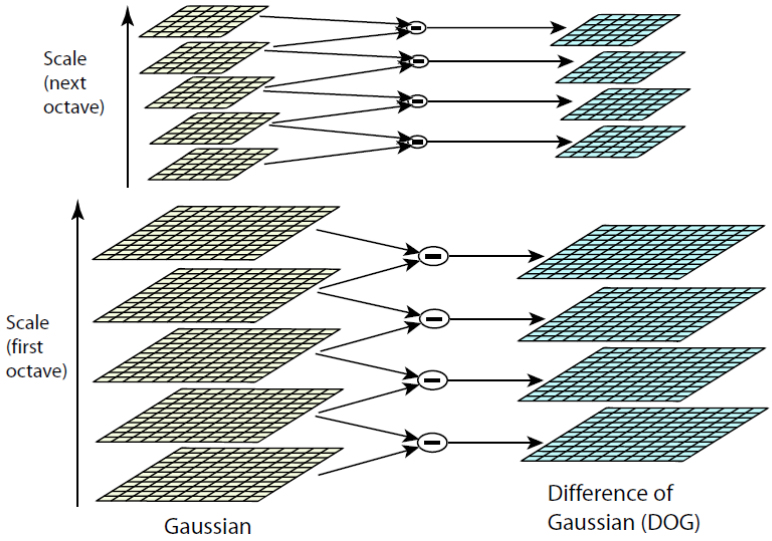
در روابط (۳-۱) تا (۳-۳) معادلات حاکم بر این مرحله آورده شده است. در این روابط I تصویر اصلی می‌باشد و G تابع کرنل گوسین است که با ضرب تلفیقی(کانولوشن) ، تصویر نرم شده L را تولید می‌نماید. سپس با تفاضل دو تصویر حاصل شده‌ی L در هر دسته پنج تایی با تصویر مجاورش که اختلاف آنها در انحراف معیارσ می‌باشد، تصویر D براساس اختلاف تفاضلی گوسین ایجاد می‌گردد.

$$L(x,y,\sigma) = I(x,y) * G(x,y,\sigma) \quad (۳-۱)$$

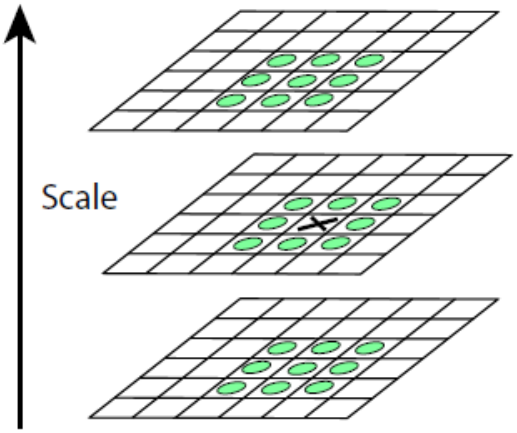
$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(\frac{x^2+y^2}{2\sigma^2})} \quad (۳-۲)$$

$$D(x,y,\sigma) = \big(G(x,y,k\sigma) - G(x,y,\sigma)\big) * I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma) \quad (۳-۳)$$

مرحله‌ی بعد یافتن نقاط ماکسیمم یا مینیمم در هر اکتاو است. در این مرحله اندازه سطح خاکستری هر پیکسل با اندازه ۸ پیکسل مجاور آن مقایسه می‌گردد و همچنین با ۹ پیکسل در تصاویر مجاور بالایی و پایینی (که از لحاظ σ با هم اختلاف داشتند و از تابع اختلاف گوسین حاصل شده بود) مقایسه می‌گردد اگر مقدار این پیکسل از تمام ۲۶ پیکسل همسایه بیشتر یا کمتر بود به عنوان یک ویژگی کاندید انتخاب شده و به آن نقطه کلیدی گفته می‌شود. (شکل ۳-۲)



شکل ۳-۱: ایجاد هرم DOG [۵۰]



شکل ۳-۲: آشکارسازی اکسترمم‌ها در هرمDOG [۵۰]

۳-۱-۲ تعیین موقعیت مکانی هر نقطه اصلی

در این مرحله به طور دقیق موقعیت هر نقطه اصلی از لحاظ مختصات تعیین می‌گردد. سپس دو عمل زیر انجام می پذیرد:

۱) نقاط ناپایدار که روشنایی پایینی دارند، حذف می‌شود.

۲) نقاط مربوط به لبه‌ها نیز حذف می‌گردد چرا که موجب ناپایداری می‌شوند.

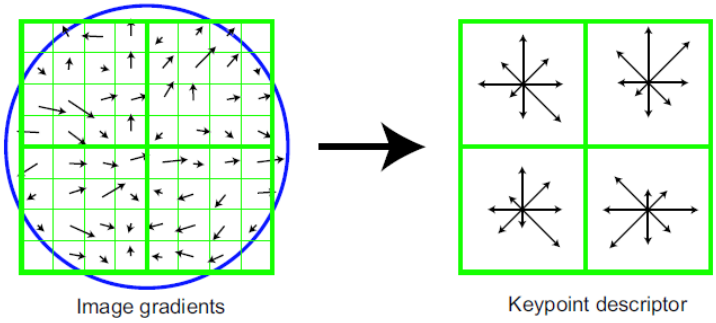
سپس دامنه‌ی گرادیان و جهت آن برای هر نقطه اصلی با توجه به روابـط (۳-۴) و (۳-۵) تعیـین می‌گردد.

$$M(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \qquad (۳-۴)$$

$$\theta(x,y) = \tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)) \qquad (۳-۵)$$

۳-۱-۳ توصیفگر نقاط کلیدی

در این مرحله بردار ویژگی اصلی ایجاد خواهد شد. در ابتدا دامنه‌ی گرادیان و جهت آن در اطراف نقطه‌ی کلیدی محاسبه می شود. مرجع [۵۰] در آزمایشات خود از یک شبکه‌ی ۱۶×۱۶ در اطراف هر نقطه‌ی کلیدی استفاده نموده است که این شبکه به بلوک‌های ۴×۴ تقسیم‌بندی شده و از هر بلوک یک هیستوگرام ۸ جهته استخراج می‌شود. در واقع در هر بلوک، برای هر یک از جهت‌ها دامنه‌ی گرادیان‌ها با هم جمع می‌شوند. بنابراین طول بردار ویژگی ۱۲۸= ۴×۴×۸ عنصر برای هر نقطه‌ی کلیدی خواهد بود. شکل ۳-۳ این مرحله را به ازای آرایه‌ی ۲×۲ نشان می‌دهد. طبیعت محلی و چند مقیاسی بودن ویژگی‌ها، آنها را نسبت به نویز، شلوغی و انسداد، غیر حساس می‌سازد.



شکل ۳-۳: هیستوگرام گرادیان جهت‌دار [۵۰]

۳-۲ بردار فیشر

بردار فیشر یک روش بازنمایی تصویر است که از تجمیع ویژگی‌های محلی بدست می‌آید. بردار فیشر را می‌توان به صورت کیف کلماتی(BOW) در نظر گرفت که آمارهای مراتب بالاتر را کد می‌کند. با استفاده از این نمایش ویژگی‌های محلی با یک بردار در ابعاد بالاتر کد می‌شوند. سرعت محاسبات بالا، عملکرد مناسب در کار با طبقه‌بندهای خطی، فشرده‌سازی با حداقل کاهش دقت از مزایای استفاده از بردارهای فیشر است. در بخش بعدی روش استخراج بردارهای فیشر ارائه می‌گردد [۲۴].

۳-۲-۱ استخراج بردار فیشر تصاویر

مجموعه‌ی $X=\{x_t, t=1,\dots T\}$ به صورت T توصیفگر محلی D بُعدی استخراج شده از یک تصویر تعریف می‌شود. (به عنوان مثال مجموعه توصیفگرهای SIFT تصویر)

ایده‌ی اصلی بردار فیشر مدل کردن داده‌ها با محاسبه‌ی گرادیان درستنمایی[1] داده‌ها نسبت به پارامترهای مدل (میانگین و کوواریانس)می‌باشد. با فرض استقلال نمونه‌ها داریم:

$$G_\lambda^X = \frac{1}{T}\sum_{t=1}^{T} \nabla_\lambda \log U_\lambda(x) \qquad (۶-۳)$$

در رابطه‌ی بالا U_λ تابع چگالی احتمال توصیفگر x است که به صورت مخلوطی از چگالی احتمال-های گوسی[2](GMM) انتخاب شده است (روابط (۳-۷) و(۳-۸)).

$$U_\lambda(x) = \sum_{k=1}^{K} W_k\, u_k(x) \qquad W_k \geq 0 \quad \sum_{k=1}^{K} W_k = 1 \qquad (۷-۳)$$

$$u_k(x) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_k|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(X-\mu_k)'\Sigma_k^{-1}(X-\mu_k)\} \qquad (۸-۳)$$

[1] Gradient of the likelihood
[2] Gaussian Mixture Model

پارامترهای GMM بصورت $\lambda=\{w_k, \mu_k, \Sigma_k, k=1\dots K\}$ است. w وزن توابع گوسی در مخلوط، μ میانگین گوسی‌ها، Σ کوواریانس(قطری) گوسی‌ها و K تعداد گوسی‌ها در هر GMM است. برای بدست آوردن بردار فیشر مجموعه داده‌های X، ابتدا GMM روی تعدادی زیادی از تصاویر آموزش دیده و پارامترهای آن توسط الگوریتم EM محاسبه می‌شوند. بعد از آموزش GMM بردار فیشر مجموعه‌ی X برای گوسیk اُم توسط روابط زیر بدست می‌آیند:

$$\mathcal{G}_{\mu,k}^{X} = \frac{1}{T\sqrt{W_k}} \sum_{t=1}^{T} \gamma_t(k) \left(\frac{x_t - \mu_k}{\sigma_k}\right) \quad (۹-۳)$$

$$\mathcal{G}_{\sigma,k}^{X} = \frac{1}{T\sqrt{W_k}} \sum_{t=1}^{T} \gamma_t(k) \frac{1}{\sqrt{2}} \left[\frac{(x_t - \mu_k)^2}{\sigma_k^2} - 1\right] \quad (۱۰-۳)$$

$$\gamma_t(k) = \frac{W_k u_k(x_t)}{\sum_{j=1}^{k} W_j u_j(x_t)} \quad (۱۱-۳)$$

$\mathcal{G}_{\mu,k}^{X}$ و $\mathcal{G}_{\sigma,k}^{X}$ بردارهای D بُعدی هستند و بردار فیشر نهایی با ترکیب این بردارها به ازای تمام گوسی‌ها حاصل می‌شود.

$$FV = \left[\mathcal{G}_{\mu,1}, \mathcal{G}_{\sigma,1}, \dots\dots, \mathcal{G}_{\mu,K}, \mathcal{G}_{\sigma,K}\right] \quad \in 2KD \quad (۱۲-۳)$$

۳-۳ روش پیشنهادی *Almazan*

نویسندگان مختلف در متون دست‌نویس شیوه نگارشی کاملاً متنوعی دارند طوری که کلمات یکسان می‌تواند کاملاً متفاوت در نظر گرفته شوند. این تغییرات زیاد در شیوه نگارش باعث می‌شود که این دسته مسائل از گونه‌های دیگر بسیار پیچیده‌تر و سنگین‌تر باشد. علت این پیچیدگی دو عامل تأثیرگذار در تفاوت بین کلمات است: شیوه نگارش و حروف تشکیل‌دهنده. اما هدف این است که فقط تفاوت بین حروف باعث متفاوت بودن کلاس شود و تفاوت شیوه نگارشی تأثیری نداشته باشد.

همان طور که در فصل دوم نیز بیان شد Almazan و همکـارانش در تحقیـق [۴۴] مشخصـه‌ای برای کلمات در زبان لاتین معرفی نموده‌اند. در این پایان نامـه از همـین مشخصـه بـا در نظـر گـرفتن خصوصیات زبان فارسی استفاده می‌شود. به کارگیری این نوع مشخصه باعث می‌شود که شـباهت بـین کلمات مستقل از شیوه نگارش حفظ شود. برای ارزیـابی ایـن الگـوریتم از دو پایگـاه داده فارسـا بـرای مرحله‌ی آموزش و پایگاه داده متن دست نوشـته‌ی فارسـی[1](FHT)[۵۱] بـرای آزمـون اسـتفاده شـده است. پایگاه داده‌ی فارسا شامل ۱۳۵۰۰تصویر دست‌نوشته از کلمات مختلف می‌باشد. بـرای هـر کلمـه ۴۵ نمونه از دست‌نوشته‌های افراد مختلف وجود دارد. پایگاه داده‌ی FHT شامل ۱۰۰۰ تصـویر دسـت-نوشته‌ی فارسی از متون، در ۱۳ موضوع مختلف می‌باشد. در این روش پرسش می‌توانـد بـه دو صـورت تصویر و رشته‌ی متنی کلمه باشد. در ادامه مراحل روش تحقیـق [۴۴] روی زبـان فارسـی بـه تفضـیل بیان خواهد شد.

۳-۳-۱ نمایش کلمات بر پایه‌ی مشخصه‌ها

مشخصه‌ها خواص معنایی هستند که می‌توانند تصاویر و گروه‌ها را توصیف کنند و اخیرا برای وظایف کلاس‌بندی و بازیابی تصاویر شهرت فراوانی را کسب نموده‌اند. به طور کلی مشخصه‌ها این ظرفیت را دارند که به راحتی به ما اجازه توصیف، مقایسه و دسته‌بندی اشیاء را می‌دهند. استفاده از مشخصه‌ها منجر به روشی می‌شود که از دانش بین کلاس‌ها بهره می‌برد. در واقع اشتراک‌گذاری اطلاعات بین کلمات برای یادگیری مناسب بسیار مهم است و استفاده از مشخصه‌ها یکی از روش‌های دستیابی به این هدف است.

مشخصه‌ها در تشخیص کلمات باید به صورتی تعریف شوند که قادر به تمیز دادن کلمات از یگدیگر باشند و همچنین به نوع نگارش وابسته نباشند. یکی از این مشخصه‌های پیشنهاد شده

[1] Farsi Handwritten Text(FHT)

هیستوگرام هرمی حروف (PHOC) است. با استفاده از این روش رشته‌های متنی مربوط به کلمات در داخل یک فضای باینری به نام فضای برچسب تعبیه می‌شوند. ورودی سیستم در این مرحله رشته‌های متنی مربوط به تصاویر کلمات است که در داخل فایل Notepad تایپ شده‌اند.

در این روش یک هیستوگرام باینری از حروف ساخته می‌شود. با در نظر گرفتن الفبای فارسی، هیستوگرام ۳۲ بُعد دارد که به ترتیب حروف الفبای فارسی از "الف" شروع و به "ی" ختم می‌شود. هر بُعد نشان دهنده‌ی این است که آیا رشته‌ی متنی حرف خاصی را در خود دارد یا نه. در واقع اگر کلمه شامل یک حرف خاص بود در بُعد مربوط به آن در داخل هیستوگرام یک و در غیر این صورت صفر قرار می‌گیرد.

در این حالت کلماتی مانند "سالم" و "سلام" نمایش مشابه‌ای دارند. به همین دلیل برای ایجاد تمایز بین کلماتی که دارای حروف مشابه هستند از روش هرمی استفاده شده که به جای پیدا کردن حروف در کل کلمه روی نواحی مختلف کلمه تمرکز نموده است. در این روش برای هر کلمه سطوح ۲، ۳، ۴ و ۵ در نظر گرفته شده است. در سطح ۲ کلمه به دو بخش، در سطح ۳ کلمه به سه بخش و به همین ترتیب برای سطوح بعدی تقسیم می‌شود. برای هر بخش ۳۲ بُعد در نظر گرفته می‌شود. به طور مثال در سطح ۲ مشخصه‌هایی به این صورت که آیا کلمه یک حرف خاص را در بخش اول یا در بخش دوم دارد تعریف می‌شوند. در صورت وجود این مشخصه در بُعد مربوط به آن یک قرار می‌گیرد. مشخصه‌ها با این تعریف تمیز دهنده‌ی کلمات هستند.

در نهایت این هیستوگرام دارای ۴۴۸=۳۲×(۲+۳+۴+۵) بُعد برای حروف است که هر بُعد یک مشخصه است. بطور مثال وجود یک در بُعد اول این هیستوگرام معرف مشخصه‌ای به این صورت است که اگر کلمه‌ای بصورت دو بخشی در نظر گرفته شود حرف"الف" را در بخش اول خود داراست.

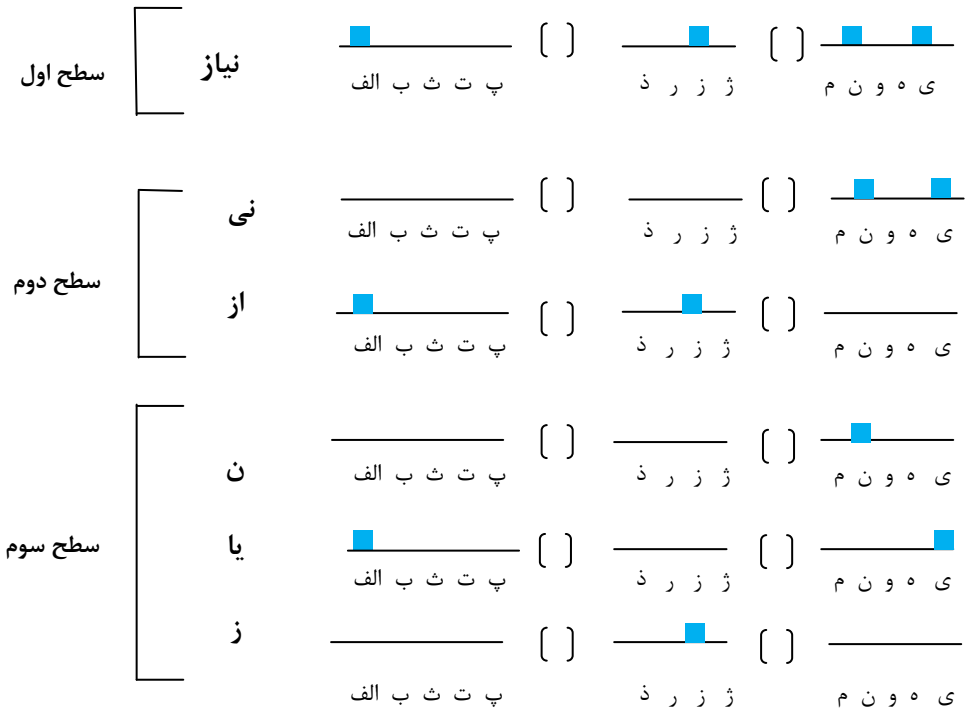
در ادامه ۵۰ تا از رایج‌ترین بایگرام‌های فارسی که هر کدام دارای دو حرف هستند در سطح ۲ اضافه شده است که با اضافه شدن ۲۰۶ بُعد اضافی در نهایت این هیستوگرام دارای ابعاد ۶۵۴ می‌شود.

شکل ۳-۴ هیستوگرام هرمی را برای کلمه‌ی" نیاز" نشان می‌دهد.

برای تشخیص اینکه هر حرف از کلمه‌ی مورد نظر به کدام بخش از هرم اختصاص دارد از روابط (۳-۱۴) و (۳-۱۵) استفاده شده است. محل قرار گرفتن c اُمین حرف در کلمه‌ای به طول m به صورت بازه‌ی محاسبه شده توسط رابطه‌ی (۳-۱۴) تعریف می‌شود. بازه‌ی در نظر گرفته شده برای بخش rاُم در سطح l هم توسط همین رابطه مشخص می‌شود. سپس در صورت برقراری رابطه‌ی (۳-۱۵) ، c اُمین حرف به بخش r در سطح l اختصاص دارد.

$$Occ(c,m) = [\frac{c-1}{m}, \frac{c}{m}] \qquad (۳-۱۴)$$

$$\frac{|Occ(c,m) \cap Occ(r,l)|}{|Occ(c,m)|} \geq 0.5 \; where |[a\text{-}b]| = b\text{-}a \qquad (۳-۱۵)$$



شکل ۳-۴: هیستوگرام هرمی حروف در سطوح ۱، ۲ و۳

۳-۳-۲ استخراج بردارهای فیشر

در مرحله‌ی قبل رشته‌های متنی مربوط به تصاویر پایگاه داده فارسا با استفاده از PHOC در فضای برچسب تعبیه شده‌اند. در مرحله بعدی برای آموزش مشخصه‌ها، ابتدا تصاویر کلمات پایگاه داده‌ی فارسا به بردارهای ویژگی کدگذاری می‌شوند. سپس این بردارهای ویژگی به همراه برچسب‌های PHOC متناظر برای آموزش مشخصه‌ها به کار می‌روند. از بردارهای فیشر به عنوان نمایش پایه تصویر استفاده شده است و بردارهای ویژگی برای هر تصویر استخراج می‌شوند. در ادامه روش استخراج این بردارها بیان می‌شود.

نرمال‌سازی ابعاد تصاویر یکی از مهمترین مراحل پیش‌پردازش است. اگر عرض تصویر w و ارتفاع آن h باشد نسبت ابعاد تصویر برابر است با:

$$R = \frac{h}{\mathrm{w}} \qquad (۱۶-۳)$$

در این حالت برای تمام تصاویر ارتفاع برابر با ۸۰ پیکسل در نظر گرفته شده و با توجه به رابطه‌ی (۳-۱۶) و با فرض ثابت ماندن نسبت طول به عرض، طول جدید هر تصویر بدست می‌آید.

تصاویر پایگاه داده‌ی فارسا، بصورت خاکستری می‌باشند که پیکسل‌های مربوط به پس زمینه[1] دارای سطح یک هستند و پیکسل‌های مربوط به کلمه یا پیش زمینه[2] بصورت خاکستری می‌باشند. برای رفع نویز، ابتدا آنها را به تصاویر باینری تبدیل کرده و سپس با اعمال فیلتر گوسی دوباره به تصاویر خاکستری تبدیل می‌شوند.

برای آموزش مدل GMM، ۲۵۰ تصویر به صورت تصادفی و غیر تکراری از تصاویر پایگاه داده‌ی

[1] Background
[2] Foreground

فارسا انتخاب می‌شوند. سپس توصیفگرهای متراکم تبدیل ویژگی مستقل از تغییر مقیاس[1] از هر تصویر استخراج می‌شود. دلیل استفاده از DSIFT به جای SIFT این است که با صرف مدت زمان کمتری می‌تواند تعداد ویژگی‌های بیشتری استخراج کند[۵۲]. تفاوت بین این دو الگوریتم در این است که در DSIFT مانند SIFT نیازی به پیدا کردن نقاط کلیدی براساس فیلتر اختلاف تفاضلی (DOG) نمی‌باشد و با در نظر گرفتن یک گام مشخص در جهت سطرها و ستون‌ها در مرکز هر بلوک یک توصیفگر استخراج می‌گردد. بردار ویژگی با ابعاد ۱۲۸×n برای هر تصویر ایجاد می‌شود. در واقع n برابر با مجموع تمام نقاطی است که در آن ویژگی استخراج شده است. همانطور که بیان شد این نقاط گره‌های یک شبکه منظم را تشکیل می‌دهند.

یکی از مشکلات روش‌های کدگذاری مثل فیشر این است که به طور واضح موقعیت مکانی ویژگی‌ها را که برای توصیف کلمه بسیار مهم است کدگذاری نمی‌کنند. پیش‌بینی صحیح مشخصه‌ها در مکان مناسب اهمیت خیلی زیادی دارد چرا که شدیداً با عملکرد نهایی در عملیات تشخیص و بازشناسی همبستگی دارد. به همین دلیل برای آموزش GMM، به ۱۲۸ ویژگی DSIFT استخراج شده، مختصات مکانی نقاط آنها نیز اضافه می‌شود و ابعاد ویژگی‌ها به ۱۳۰ می‌رسد.

از ویژگی‌ها کل تصویر بصورت یکجا برای آموزش GMM استفاده نمی‌شود بلکه ابتدا تصویر به ۱۲ ناحیه تقسیم شده و ویژگی‌های هر ناحیه به صورت مستقل از هم در نظر گرفته می‌شوند و GMM روی هر ناحیه با توجه به ویژگی‌های همان ناحیه آموزش داده می‌شود. در نظر گرفتن مختصات مکانی ویژگی‌ها و آموزش GMM اختصاص داده شده به هر ناحیه نتایج بهتری در پیش‌بینی مشخصه‌ها در محل صحیح دارد. از ۱۶ گاوسی در هر GMM استفاده شده است که پارامترهای آن بصورت $g=\{w_k, \mu_k, \Sigma_k, k=1\dots K\}$ است. w وزن گوسی‌ها در مخلوط، μ میانگین گوسی‌ها، Σ

[1] Dense Scale Invariant Feature Transform(DSIFT)

کوواریانس(قطری)گوسی‌ها و k تعداد گاوسی‌ها است. در نهایت GMM بـا در نظـر گـرفتن ۱۲ ناحیـه دارای ۱۹۲=۱۶×۱۲گاوسی می‌باشد. برای ایجاد ۱۲ ناحیه روی تصویر مراحل زیر انجام می‌شود:

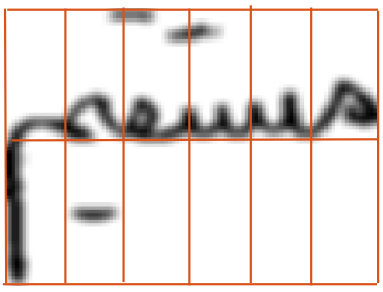
۱) به دلیل اینکه ابعاد تصاویر با هم متفاوت است مختصات مکانی نقاط باید نرمال شوند. مختصات x و y با روابط (۳-۱۷) بین ۰/۵- و ۰/۵ قرار می‌گیرند. h و w به ترتیب ارتفاع و عرض تصویر هستند.

$$Y_{new} = (\frac{Y_{old}}{\mathrm{h}})\text{-}0.5 \quad , \quad X_{new} = (\frac{X_{old}}{\mathrm{w}})\text{-}0.5 \qquad (۳\text{-}۱۷)$$

در جهت محور x شش بازه و در جهت محور y دو بازه بصورت زیر در نظـر گرفتـه مـی‌شـود. ایـن بازه‌ها برای تمام تصاویر ثابت فرض شده‌اند.

Intervals the axial direction X = [-0.5,-0.33], [-0.33,-0.166], [-0.166,0], [0,0.166], [0.166,0.33], [0.33,0.5]
Intervals the axial direction Y = [-0.5,0] , [0,0.5]

۲) با توجه به این بازه‌ها، ۱۲ ناحیه روی تصویر ایجاد شده که با در نظر گرفتن مختصـات مکـانی ویژگی‌ها، هر یک از آنها به یکی از این بازه‌ها اختصاص پیدا می‌کند. در شکل ۳-۵ این ناحیه-ها برای کلمه‌ی "مستقیم" نشان داده شده است.



شکل ۳-۵: ایجاد ۱۲ ناحیه با اندازه‌ی ثابت بر روی تصویر کلمه‌ی مستقیم

همان‌طور که بیان شد برای آموزش GMM از ۲۵۰ کلمه استفاده می‌شود. مراحل این روش به این صورت است که ابتدا برای تمام ۲۵۰ کلمه، ناحیه‌بندی انجام شده و تمام ویژگی‌های مربوط به

ناحیه‌ی اول آنها، در کنار یگدیگر قرار گرفته و به همین ترتیب ویژگی‌های مربوط به ناحیه دوم تا دوازدهم به صورت جداگانه برای این ۲۵۰ کلمه در یک مجموعه قرار می‌گیرند و هر بار GMM با ویژگی‌های مربوط به هر ناحیه آموزش می‌بینند.

قبل از آموزش GMM، ابعاد ویژگی‌ها با روش تحلیل مولفه‌های اصلی(PCA)[1] از ۱۲۸ به ۶۲ کاهش پیدا می‌کند. با اضافه کردن دو مختصات مکانی ابعاد ویژگی‌ها به ۶۴ می‌رسد. ابعاد پارامترهای GMM برای وزن برابر با ۱×۱۹۲، میانگین ۱۹۲×۶۴ و کوواریانس ۱۹۲×۶۴ می‌باشد.

بعد از آموزش مدل GMM با روش توضیح داده شده در بالا، استخراج بردارهای فیشر برای هر یک از تصاویر پایگاه داده‌ی فارسا طی مراحل زیر انجام می‌شود:

۱) ویژگی‌های SIFT بصورت متراکم از هر تصویر استخراج شده و نرمال می‌شوند.
۲) مختصات نقاط کلیدی بین ۰/۵- و ۰/۵ نرمال می‌شوند.
۳) با روش PCA ابعاد ویژگی‌ها از ۱۲۸ به ۶۲ کاهش می‌یابد.
۴) مختصات مکانی به ویژگی‌ها اضافه شده و ابعاد ۶۴ می‌شود.
۵) با توجه به ویژگی‌ها و پارامترهای مدل GMM، بردار فیشر برای هر تصویر بصورت ستونی با ابعاد ۱×۲۴۵۷۶ استخراج می‌شود.
۶) در انتها ماتریس ویژگی‌های فیشر دارای ابعاد ۱۳۵۰۰×۲۴۵۷۶ می‌باشد. هر ستون از این ماتریس نشان‌دهنده‌ی یک نمونه است.

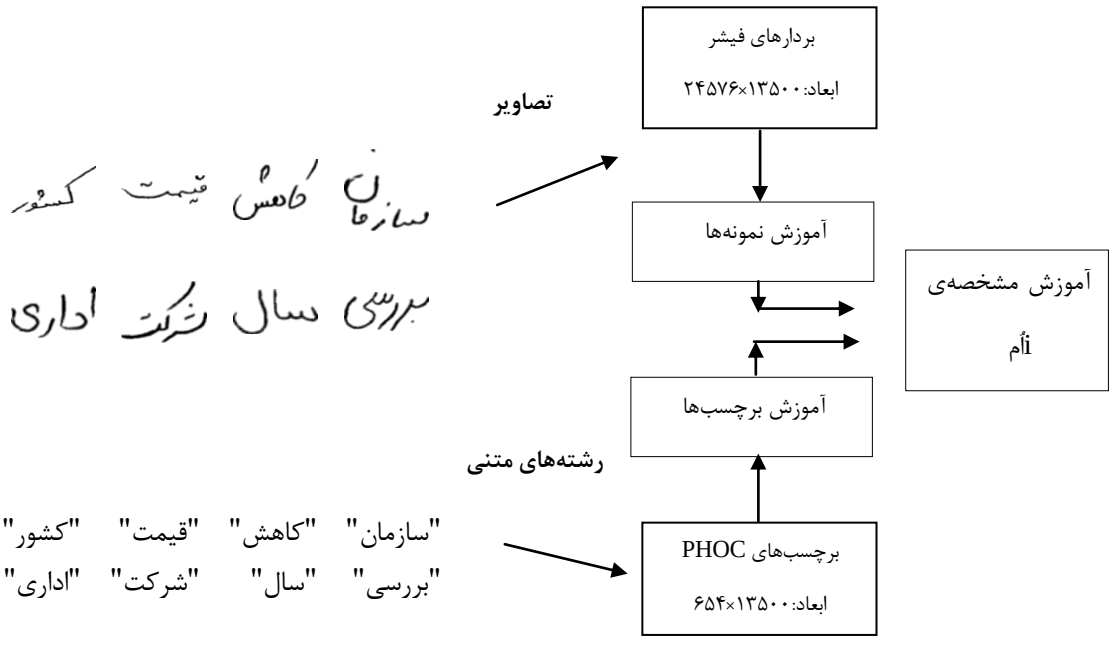
با توجه به توضیحات ارائه شده در بخش ۳-۲ ابعاد بردار فیشر طبق رابطه‌ی زیر محاسبه شده است:

[1] Principal Component Analysis(PCA)

۲۴۵۷۶=ابعاد بردار فیشر =۲× (ابعاد ویژگی‌ها +۲) × تعداد نواحی X × تعداد نواحی Y × تعداد گوسی‌ها

۳-۳-۳ آموزش فضای مشخصه

با در نظر گرفتن الفبای فارسی و استفاده از روش PHOC، ۶۵۴ مشخصه وجود دارد. برای آموزش این مشخصه‌ها، ابتدا برای تمام تصاویر موجود در پایگاه داده‌ی فارسا، رشته‌های متنی متناظر ایجاد می‌شوند. سپس با روش بیان شده در بخش ۳-۳-۱ این رشته‌های متنی تبدیل به برچسب‌های PHOC می‌شوند و ماتریسی با ابعاد ۶۵۴×۱۳۵۰۰ ایجاد می‌شود. هر سطر این ماتریس نشان‌دهنده‌ی یک مشخصه و هر ستون آن نشان‌دهنده‌ی یک کلمه است. با در نظر گرفتن این ماتریس و ماتریس بردارهای فیشر تصاویر، آموزش مشخصه‌ها صورت می‌گیرد. برای آموزش از SVM خطی استفاده شده است. مراحل این روش در شکل ۳-۶ نشان داده شده است.



شکل ۳-۶: فرآیند آموزش مدل مشخصه‌ی iأم با SVM خطی

یکی از مشکلات آموزش فضای مشخصه در این مرحله و فضای برداری مشترک در مرحله‌ی بعدی، نیاز به داشتن پایگاه داده‌ی بزرگ می‌باشد. این موضوع به این دلیل است که داده‌هایی که برای

آموزش فضای برداری مشترک استفاده می‌شوند باید از داده‌هایی که بـرای یـادگیری فضـای مشخصـه استفاده می‌شوند متفاوت باشند. علت این مساله این است که اگر همان داده‌های مـورد اسـتفاده بـرای یادگیری فضای مشخصه را برای آموزش فضای برداری مشترک استفاده کرد، رتبه‌های SVM به شدت بیش‌برازش[1] خواهند بود (بیشتر آنها خیلی نزدیک به ۱ یا -۱ خواهنـد شـد). بنـابراین فضـای بـرداری مشترک آموزش داده شده با استفاده از این داده‌ها شدیدا بایاس خواهند داشت که این موضـوع باعـث افت نتایج می شود.

برای غلبه بر این مشکل از آموزش توسط روش bagging استفاده می‌کنیم، داده‌های آموزش به بخش‌های مختلف آموزش و اعتبارسنجی[2] تقسیم بندی شده‌اند. یک مشخصه با استفاده از داده‌های آموزشی آن بخش آموزش داده می‌شود و برای رتبه‌دهی، داده‌های اعتبارسنجی استفاده می‌شوند. بنابراین رتبه‌های داده‌های اعتبارسنجی بدون بایاس است. از همین داده‌ها برای آموزش فضای برداری مشترک در گام بعد استفاده می‌شود.

در این روش، آموزش هر مشخصه در دو مرحله[3] و هر مرحله در پنج تکرار[4] انجام می‌شود. داده-های آموزش و اعتبارسنجی هر تکرار از هم مستقل هستند. به طور کلی در هر تکرار یک قسمت از داده‌های آموزش به عنوان مجموعه‌ی اعتبارسنجی قرار می‌گیرد به طوری که در پایان پنج تکرار تمام داده‌های مجموعه‌ی آموزش یکبار به عنوان داده‌ی اعتبارسنجی رتبه‌دهی شده‌اند. برای آموزش در هر مرحله داده‌ها بصورت تصادفی مرتب می‌شوند.

در انتهای پردازش، یک مدل کلی با میانگین‌گیری همه مدل‌های محلی حاصل می‌شود. با نرمال کردن رتبه مشخصه هر نمونه به وسیله تعداد دفعاتی که آن نمونه رتبه‌دهی شده است رتبه‌های بدون

[1] Over fit
[2] Validation
[3] Pass
[4] Iteration

بایاس مجموعه یادگیری ایجاد می‌شود که می‌تواند بدون مشکلی برای آموزش فضای برداری مشترک استفاده شود. به طور کلی مراحل آموزش مشخصه‌ی iاُم به صورت زیر است:

مرحله‌ی اول: با توجه به سطر iاُم ماتریس برچسب PHOC، نمونه‌های مثبت و منفی برای این مشخصه تعیین می‌شوند. (نمونه‌های مثبت، کلماتی هستند که دارای این مشخصه می‌باشند در واقع برای این نمونه‌ها در سطر iاُم یک قرار گرفته و برای نمونه‌های منفی که دارای این مشخصه نیستند در سطر iاُم صفر قرار دارد)

مرحله‌ی دوم: مجموعه نمونه‌های مثبت و منفی بصورت تصادفی مرتب می‌شوند.

مرحله‌ی سوم: درهر تکرار ، ۸۰ درصد اول مجموعه نمونه‌های مثبت و منفی برای مرحله‌ی آموزش و ۲۰ درصد باقیمانده برای مجموعه‌ی اعتبارسنجی در نظر گرفته می‌شوند.

مرحله‌ی چهارم: در مجموعه‌های آموزش و اعتبارسنجی ابتدا نمونه‌های مثبت و سپس نمونه-های منفی قرار می‌گیرد.

مرحله‌ی پنجم: برای مجموعه‌ی آموزش و اعتبارسنجی، ماتریس‌های بردارهای فیشر و برچسب با توجه به مراحل بالا ایجاد می‌شوند. (بردار برچسب به این صورت ایجاد می‌شود که برای تمام نمونه-های مثبت ۱ و برای نمونه‌های منفی ۱- در نظر گرفته می‌شود)

مرحله‌ی ششم: آموزش توسط SVM خطی صورت می‌گیرد. پارامترهای ورودی به طبقه‌بند SVM، ماتریس بردارهای فیشر و بردار برچسب مجموعه‌ی آموزش و خروجی این طبقه‌بند، بردار ستونی وزن و مقدار بایاس به ازای مشخصه‌ی iاُم است.

مرحله‌ی هفتم: به ازای بردار وزن نهایی بدست آمده از مرحله‌ی قبل رتبه مشخصه برای داده-های اعتبارسنجی محاسبه می‌شود. بردار ستونی وزن و مقدار بایاس در هر تکرار ذخیره می‌شود.

رتبه مشخصه‌ها (Sc) برای داده‌های اعتبارسنجی از رابطه‌ی (۳-۱۸) محاسبه می‌شود. در این رابطه FV_{val} ماتریس بردارهای فیشر برای مجموعه‌ی اعتبار سنجی و W بردار وزن می‌باشد.

$$Sc = W^T \times FV_{val} \qquad (۱۸-۳)$$

مرحله‌ی هشتم: روی مجموع داده‌ها روش تغییر مکان دایره‌ای[1] به تعداد داده‌های مجموعه‌ی اعتبارسنجی انجام می‌شود. (به این صورت در هر تکرار داده‌های اعتبارسنجی عوض شده و در پایان پنج تکرار هر داده‌ی مجموعه‌ی آموزش یکبار در مجموعه‌ی اعتبارسنجی ظاهر شده است)

مرحله‌ی نهم: به ابتدای مرحله‌ی سوم بازگشته و دوباره مراحل انجام می‌شود تا پنج تکرار تمام شود.

مرحله‌ی دهم: بعد از اتمام پنج تکرار، هر نمونه یکبار رتبه‌دهی شده است. برای اینکه یکبار دیگر مرحله‌ی آموزش تکرار شود به مرحله‌ی دوم بازگشته و مراحل دوباره تکرار می‌شود.

در پایان فرآیند آموزش برای مشخصه‌ی iاُم، بردار وزن نهایی برای این مشخصه با محاسبه‌ی مجموع بردارهای وزن بدست آمده از هر تکرار و تقسیم آن بر ۱۰ بدست می‌آید. مقدار بایاس نهایی هم به همین صورت میانگین‌گیری می‌شود. بردار سطری رتبه مشخصه‌ی نهایی با محاسبه‌ی مجموع بردارهای رتبه مشخصه‌ی بدست آمده از هر مرحله با تقسیم بر ۲ محاسبه می‌شود.

برای مشخصه‌های ۱ تا ۶۵۴ با همین روش بردار وزن بدست آمده است. به ازای هر مشخصه این بردار ستونی دارای ابعاد ۱×۲۴۵۷۶ می‌باشد. در نهایت با در نظر گرفتن همه‌ی مشخصه‌ها، ماتریس W

[1] Circular Shift

نهایی با ابعاد ۶۵۴×۲۴۵۷۶ ایجاد می‌شود. برای هر مشخصه، بردار سطری رتبه مشخصه‌ی آن، دارای ابعاد ۱۳۵۰۰×۱ است. که ماتریس رتبه مشخصه نهایی دارای ابعاد ۱۳۵۰۰×۶۵۴ می‌باشد.

برای هر مشخصه اگر تعداد نمونه‌های مثبت آن کمتر از ۲ بود، برای آن مشخصه، بردار وزن، بردار رتبه مشخصه و مقدار بایاس صفر در نظر گرفته شده و در واقع آموزشی صورت نمی‌گیرد.

۴-۳-۳ تعبیه رتبه مشخصه‌ها و برچسب‌ها در فضای برداری مشترک

در طی بخش‌های قبلی روش تعبیه رشته‌های متنی در فضای برچسب و تعبیه تصاویر در فضای مشخصه بیان شد. نمایش تصاویر کلمات بر پایه رتبه مشخصه و تعبیه آنها در فضای مشخصه، باعث ایجاد نمایشی می‌شود که در مقابل تغییرات ظاهری کلمات مقاوم است. فضای برچسب و مشخصه هر دو دارای ابعاد یکسانی هستند ولی به این دلیل که برچسب‌های متنی تعبیه شده باینری می‌باشند در حالیکه رتبه‌های مشخصه دارای محدوده‌ی متفاوتی هستند مقایسه آنها در این فضا مناسب نیست. بنابراین کالیبراسیون، رتبه‌های مشخصه و برچسب‌های PHOC مورد نیاز می‌باشد.

برای حل این مشکل از روش تحلیل هم‌بستگی کانونی(CCA) استفاده شده است. با استفاده از این روش رتبه مشخصه‌ها و برچسب‌های PHOC در یک فضای برداری مشترک تعبیه می‌شوند که در این فضا همبستگی بین آنها ماکزیمم است. در این فضا رتبه مشخصه و برچسب PHOC که مربوط به یک کلمه هستند، نزدیک به یگدیگر قرار دارند. به عبارت دیگر در این فضا تصاویر و رشته‌ی متنی که نشان‌دهنده‌ی کلمه یکسان هستند نزدیک به یکدیگرند و این اجازه را می‌دهد که عملیات بازشناسی به صورت مسئله نزدیک‌ترین همسایگی انجام شود.

تجزیه و تحلیل همبستگی کانونی، با تجزیه و تحلیل همزمان مجموعه‌ها و با شناسایی و مشخص کردن عناصری از یک مجموعه متغیر با بیشترین وابستگی و ارتباط با عناصر مجموعه متغیر دیگر، روابط آماری مستقل موجود بین دو مجموعه متغیر را مورد آزمون قرار می‌دهد[۵۳]. CCA ابزاری

است که توصیف‌گرهای تصویر و برچسب‌های منتاظرشان را به هم مرتبط می‌کند. این روش مزایای فراوانی برای عملیات بازیابی و خوشه‌بندی دارد. در این تحقیق هدف از این مرحله دستیابی به دو ماتریس نگاشت U_x و V_y است که توسط این دو ماتریس، داده‌ها از فضای مشخصه و برچسب به فضای برداری مشترک نگاشت می‌شوند.

در این روش ابتدا ماتریس رتبه مشخصه و برچسب PHOC بدست آمده از مراحل قبل با توجه به اینکه تعداد نمونه‌ها ۱۳۵۰۰ است به صورت تصادفی مرتب می‌شوند. سپس برای این مرحله ۳۰ درصد اول مجموعه برای اعتبارسنجی و ۷۰ درصد باقی‌مانده برای آموزش استفاده می‌شوند.

برای حل این مساله فرض می‌شود که $I=\{I_n, n=1,\ldots.N\}$ مجموعه‌ای از N تصویر موجود برای آموزش می‌باشد و $L=\{L_n, n=1,\ldots.N\}$ مجموعه‌ی برچسب‌های متناظر با آنها است. طبق رابطه‌ی (۳-۱۹)، X مجموعه‌ی N تصویری است که در فضای مشخصه‌ی d بُعدی (d=654) تعبیه شده‌اند و Y مجموعه‌ی N برچسب تعبیه شده در فضای برچسب d بعدی می‌باشد[۵۴].

$$X = \emptyset_I(I) \in R^{d\times N} \ , \ Y = \emptyset_L(L) \in \{1,0\}^{d\times N} \qquad (۳-۱۹)$$

میانگین مجموعه‌ی X وY در راستای سطرها با پارامترهای μ_x و μ_y مشخص می‌شود. ماتریس-های C_{xx}، C_{yy} ، C_{yx} و C_{xy} توسط روابط (۳-۲۰) تا (۳-۲۳) بدست می‌آیند.

$$C_{xx} = \frac{1}{N}(X-\mu_x)(X-\mu_x)^T + \rho I \qquad (۳-۲۰)$$

$$C_{yy} = \frac{1}{N}\left(Y-\mu_y\right)\left(Y-\mu_y\right)^T + \rho I \qquad (۳-۲۱)$$

$$C_{xy} = \frac{1}{N}(X-\mu_x)\left(Y-\mu_y\right)^T \qquad (۳-۲۲)$$

$$C_{yx} = \left(C_{xy}\right)^T \qquad (۳-۲۳)$$

در این روابط ρ ضریب تنظیم است(ρ=0.00001) و I ماتریس همانی است. هدف یافتن

ماتریس‌های U_x و V_y طبق رابطه‌ی (۳-۲۴) است:

$$argmax_{U_x,V_y} \frac{U_x^T C_{xy} V_y}{\sqrt{U_x^T C_{xx} U_x}\sqrt{V_y^T C_{yy} V_y}} \quad (۲۴-۳)$$

با در نظر گرفتن رابطه‌های (۳-۲۵) و (۳-۲۶) ماتریس U_x که معادل با بردارهای ویژه‌ی ماتریس Z است بدست می‌آید.

$$ZU_{xk} = \lambda_k^2 U_{xk} \ , \ Z = C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} \quad (۲۵-۳)$$

(۲۶-۳)

با در نظر گرفتن k=128 ماتریس U_x دارای ابعاد d×k است. با یک روش مشابه ماتریس V_y نیز بدست آمده است. توسط این دو ماتریس، رتبه مشخصه‌ها و برچسب‌ها از فضای d بُعدی در فضای k بُعدی تعبیه می‌شوند. در واقع این روش اجازه می‌دهد که ابعاد خروجی را کنترل نموده و کاهش ابعاد را نیز انجام داد.

حل معادلات بالا و بدست آوردن دو ماتریس U_x و V_y، یکبار بصورت برون‌خط در مرحله‌ی آموزش انجام می‌شود. در زمان آزمون یک نمونه ($S \in R^d$) با توجه به اینکه در فضای مشخصه قرار دارد یا فضای برچسب، توسط $U_x^T(S - \mu_x)$ یا $V_y^T(S - \mu_y)$ در فضای برداری مشترک تعبیه می‌شود.

۳-۳-۵ عملیات بازشناسی

داده‌های اعتبارسنجی با توجه به مرحله‌ی قبل در فضای برداری مشترک تعبیه می‌شوند. برای هر یک از نمونه‌ها، کلاس مورد نظر آن نمونه مشخص است. ماتریس رتبه مشخصه تعبیه شده در این فضا با φ_I و ماتریس برچسب تعبیه شده در این فضا با φ_L نمایش داده می‌شوند. عملیات بازشناسی طی دو مرحله انجام می‌شود:

همان طور که قبلا نیز بیان شد در این فضا تصاویری که نشان‌دهنده‌ی یک کلمه یکسان هستند

نزدیک به یگدیگر قرار دارند. در مرحله‌ی اول، پرسش توسط تصویر کلمات می‌باشد. در واقع تصویر هر نمونه به عنوان یک پرسش در نظر گرفته شده است. با محاسبه‌ی ضرب داخلی، نمونه‌ای در مجموعه-ی داده‌ها که به کلمه‌ی پرسشی مورد نظر نزدیکتر است مشخص می‌شود (ضرب داخلی بین ستون مربوط به کلمه‌ی پرسشی مورد نظر در ماتریس φ_I با ماتریس φ_I محاسبه می‌شود). در این حالت اگر کلاس این نمونه با کلاس مربوط به کلمه‌ی پرسشی مورد نظر یکسان بود عملیات بازشناسی در مورد این نمونه درست در غیر این صورت نادرست انجام شده است.

در مرحله دوم، با در نظر گرفتن این مساله که در این فضا رشته‌ی متنی و تصویری که نشان-دهنده‌ی یک کلمه یکسان هستند نزدیک به یکدیگرند عملیات بازشناسی انجام می‌شود. در این حالت پرسش توسط رشته‌ی متنی کلمات انجام می‌شود. در واقع هدف از بازشناسی یافتن تصویری در پایگاه داده است که متناسب با این رشته‌ی متنی باشد. برای هر رشته‌ی متنی ستون مربوط به آن در ماتریس φ_L در نظر گرفته شده و با محاسبه‌ی ضرب داخلی بین آن و ماتریس φ_I، نزدیکترین تصویر به این رشته‌ی متنی مشخص می‌شود. اگر کلاس مربوط به هر دوی آنها یکسان بود عملیات بازشناسی در مورد این رشته‌ی متنی درست در غیر این صورت نادرست انجام شده است.

۳-۳-۵-۱ نتایج مرحله‌ی بازشناسی

همان طور که قبلا نیز بیان شد عملیات بازشناسی بر روی داده‌های اعتبارسنجی پایگاه داده‌ی فارسا که شامل تصاویر کلمات بریده شده است انجام می‌شود. جدول ۳-۱ نتایج بازشناسی روی داده-های اعتبارسنجی برای دو حالت پرسش با تصویر و پرسش با متن را نشان می‌دهد. در این جدول P میانگین تعداد کلماتی است که به درستی بازشناسی شده‌اند و MAP میانگین متوسط دقت بازشناسی را نشان می‌دهد.

جدول ۳-۱: نتایج مرحله‌ی بازشناسی

	P	MAP
حالت QBE	۹۷/۴۱	۹۳/۵
حالتQBS	۱۰۰	۹۷/۹

فصل چهارم:

الگوریتم پیشنهادی و نتایج ارزیابی آن

۴-۱ مقدمه

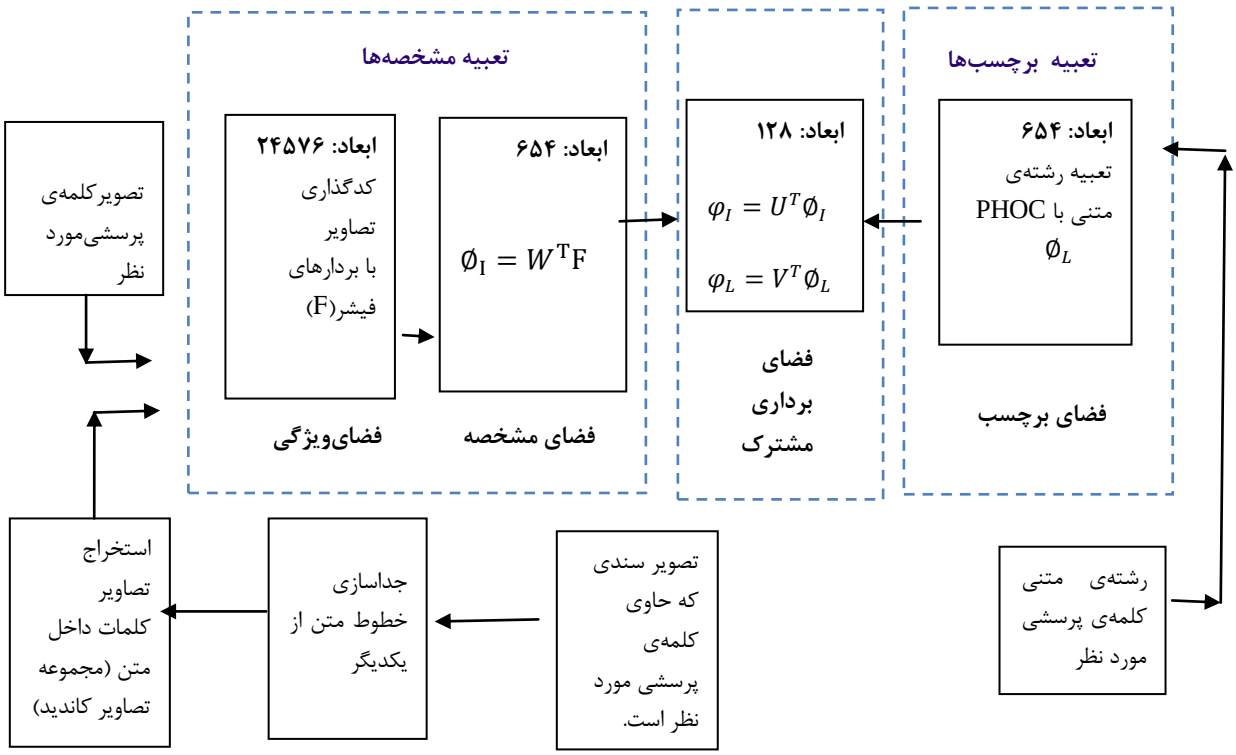
در این فصل روش پیشنهاد شده برای جستجوی کلمات کلیدی در تصویر دست‌نوشته‌ی فارسی بیان خواهد شد. به عبارت دیگر هدف پایان نامه این است که کلمه‌ی پرسشی مورد نظر در متن جستجو شده و مکان نمونه‌های آن کلمه در داخل متن مشخص شود. یکی از مهمترین تفاوت‌های این پایان نامه با تحقیق [۴۴] در این است که در این پایان نامه هدف، جستجوی کلمات در داخل اسناد تصویری می‌باشد در حالی که [۴۴] به بازشناسی تصاویر کلمات جدا شده از متن پرداخته است. در این روش کاربر می‌تواند کلمه‌ی مورد نظر خود را در داخل فایل Notepad تایپ کند و یا می‌تواند تصویر کلمه‌ی مورد نظر خود را (می‌توان تصویر کلمه‌ی مورد نظر را از داخل یک سند تصویری جدا کرد) به عنوان ورودی به سیستم بدهد، سپس سیستم مکان نمونه‌های کلمه‌ی پرسشی در داخل سند تصویری را مشخص می‌کند. مثلا با کشیدن کادر مستطیلی دور آن کلمه، آن را مشخص می‌کند. همانطور که در فصل سه بیان شد برای مرحله‌ی آزمون از پایگاه داده FHT و برای آموزش از پایگاه داده فارسا استفاده می‌شود. پایگاه داده FHT دارای ۱۰۰۰ تصویر دست‌نوشته‌ی فارسی از متون، در ۱۳ موضوع مختلف می‌باشد که شامل ۴۰ نوع متن متفاوت است و هر متن توسط ۲۵ فرد متفاوت نوشته شده است. در واقع از هر فرد تقاضا شده است ۴ متن متفاوت را بنویسد که در مجموع این پایگاه داده، ۲۵۰ نویسنده دارد. هر متن به طور میانگین دارای ۶/۴۵ خط، ۱۰۶/۶ کلمه، ۲۳۰/۱۷۵ زیر کلمه، ۱۴۰۶ حرف و ۱۳۲/۱ نقطه است. یکی از مزیت‌های این پایگاه داده وجود خطوط صحیح[1] برای هر خط می‌باشد. پایگاه داده‌ی فارسا شامل ۱۳۵۰۰تصویر دست‌نوشته از کلمات مختلف می‌باشد. برای هر کلمه ۴۵ نمونه از دست‌نوشته‌های افراد مختلف وجود دارد.

۴-۲ روش پیشنهادی اول

آموزش مشخصه‌ها، بدست آوردن ماتریس وزن W (برای تعبیه تصاویر از فضای ویژگی بـه فضـای

[1] Ground Truth

مشخصه) و ماتریس‌های نگاشت Ux و Vy (برای تعبیه رتبه مشخصه‌ها و برچسب‌ها به فضای برداری مشترک) با استفاده از پایگاه داده‌ی فارسا و روش بیان شده در فصل سوم انجام شد. برای مرحله‌ی آزمون از پایگاه داده‌ی FHT که شامل تصاویر اسناد است استفاده می‌شود. استفاده از دو پایگاه داده که دارای سبک نگارشی متفاوتی هستند نشان‌دهنده‌ی قدرت تعمیم بالای روش می‌باشد. در این حالت برای یافتن کلمه‌ی پرسشی مورد نظر در داخل متن، ابتدا خطوط متن از یگدیگر جدا شده سپس تصاویر کلمات موجود در هر خط استخراج می‌شود. با استخراج تصاویر کلمات داخل متن، یک مجموعه کلمات کاندید برای کلمه‌ی پرسشی مورد نظر ایجاد می‌شود. برای مجموعه کلمات کاندید مختصات مکانی آنها در داخل سند و شماره‌ای که مختص هر تصویر است از قبل ذخیره شده است. رشته‌ی متنی یا تصویر کلمه‌ی پرسشی به عنوان ورودی الگوریتم هستند. هدف این است که الگوریتم، نمونه‌های کلمه‌ی پرسشی را در بین مجموعه کلمات کاندید پیدا کند و مکان این نمونه‌ها در داخل سند را مشخص کند. مراحل روش در شکل ۴-۱ نمایش داده شده است.



شکل ۴-۱: مراحل روش

الگوریتم جداسازی خطوط و استخراج تصاویر کلمات متن در فصل بعد بیان خواهد شد. در پایگاه داده‌یFHT پیکسل‌های پس زمینه سیاه و پیکسل‌های قلم سفید هستند به همین دلیل بعد از استخراج تصاویر کلمات، آنها به تصاویری با پس زمینه سفید و قلم سیاه تبدیل می‌شوند.

مراحل روش در مرحله‌ی آزمون بصورت زیر می‌باشد:

۱) نرمال‌سازی ابعاد تصاویر مجموعه کلمات کاندید و کلمات پرسشی
۲) اعمال فیلتر گوسی و تبدیل تصاویر باینری به تصاویر سطح خاکستری
۳) استخراج ویژگی‌های فیشر تصاویر
۴) تعبیه تصاویر از فضای ویژگی به فضای مشخصه با ضرب در ماتریسW
۵) تعبیه تصاویر از فضای مشخصه به فضای برداری مشترک با ضرب در ماتریسU_x
۶) تعبیه رشته‌ی متنی کلمات پرسشی در فضای برچسب با استفاده از PHOC
۷) تعبیه برچسب‌هایPHOC از فضای برچسب به فضای برداری مشترک با ضرب در ماتریسV_x
۸) در فضای برداری مشترک، از بین مجموعه کلمات کاندید، نمونه‌هایی که به کلمه‌ی پرسشی موردنظر نزدیک‌تر هستند مشخص می‌شوند.
۹) انجام مرحله‌ی ازریابی

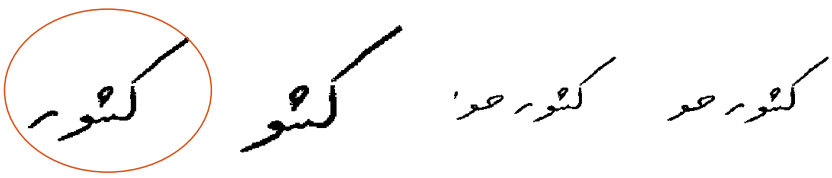
۴-۲-۱ روش ارزیابی

همان طور که در فصل سوم بیان شد در مرحله‌ی بازشناسی تمام تصاویری که در فضای برداری مشترک قرار دارند از قبل برچسب‌گذاری شده و کلاس مربوط به آنها مشخص است و هدف تعیین کلاس مربوط به کلمه‌ی پرسشی می‌باشد. ولی در این روش مجموعه کلمات کاندید دارای برچسب نیستند، به عبارتی کلاس مربوط به آنها مشخص نیست و فقط کلمات پرسشی برچسب‌گذاری می‌شوند. این مساله یکی از تفاوت‌های این روش با روش بازشناسی [۴۴] است. ارزیابی برای هر دو روش پرسش با تصویر و پرسش با رشته‌ی متنی انجام می‌شود. روش ارزیابی برای پرسش با تصویر برای یک

کلمه‌ی پرسشی شامل مراحل زیر است:

برای انجام ارزیابی لازم است ابتدا برای هر کلمه‌ی پرسشی، در مجموعه کلمات کاندید، تصاویری که نمونه‌های آن کلمه در متن هستند (تصاویری که فقط شامل کلمه‌ی مورد نظر باشند) مشخص شوند. مختصات مکانی این تصاویر در داخل متن که شامل مختصات گوشه‌ی سمت چپ بالا هر تصویر، طول و عرض می‌باشد از قبل ذخیره شده است. به این تصاویر، تصاویر صحیح متناسب با کلمه‌ی پرسشی گفته می‌شود.

در این روش فرض بر این است که اگر تصویر کلمه‌ای در مجموعه‌ی کاندید، از نظر مکانی ۸۰ درصد با این تصاویر صحیح تطبیق داشت و به عنوان خروجی سیستم آشکار شده بود به عنوان گزینه-ی درست پذیرفته شود. به این تصاویر که دارای ۸۰ درصد تطبیق هستند، مجموعه تصاویر تطبیقی گفته می‌شود. در واقع به ازای هر تصویر صحیح در متن پنجره‌های تطبیقی آن هم مشخص می‌شوند. با توجه به اینکه هر کدام از تصاویر استخراج شده از متن دارای شماره منحصربه‌فردی است، در یک بردار، شماره‌ی مربوط به این تصویر صحیح و تصاویر تطبیقی ذخیره می‌شوند به این بردار، بردار صحیح گفته می‌شود. نکته قابل توجه این است که به تعداد نمونه‌هایی که از یک کلمه‌ی پرسشی در متن وجود دارد، بردار صحیح بدست می‌آید. در شکل ۴-۲ تصویر صحیح منتاظر با کلمه‌ی پرسشی "کشور" و تصاویر تطبیقی متناظر با آن مشخص است (تصویر صحیح با بیضی مشخص شده است) .



شکل ۴-۲: تصویر صحیح و تطبیقی متناظر با کلمه‌ی کشور

در فضای برداری مشترک برای هر کلمه‌ی پرسشی، فاصله تصویر آن با هر یک از تصاویر مجموعه کلمات کاندید محاسبه می‌شود. (فاصله بین ستون مربوط به کلمه‌ی پرسشی مورد نظر در ماتریس φ_I با ماتریس φ_I محاسبه می‌شود). سپس این فاصله با رابطه‌ی (۴-۱) بین صفر تا یک نرمال شده بطوریکه هرچه تصویر موجود در مجموعه کاندید به کلمه‌ی پرسشی نزدیکتر باشد فاصله‌ی آن به یک نزدیکتر است و برعکس. در این رابطه بردار فاصله یک بردار سطری است که تعداد ستون‌های آن برابر با تعداد کلمات مجموعه‌ی کاندید است. هر ستون، فاصله با کلمه‌ی پرسشی را نشان می‌دهد. min و max به ترتیب برابر با کمترین و بیشترین مقدار در این بردار می‌باشند.

$$Distance = 1 - (\frac{distance - \min}{max}) \qquad (۴-۱)$$

در مرحله‌ی بعدی سه پارامتر زیر تعریف می‌شوند:

n : تعداد نمونه‌هایی که از کلمه‌ی پرسشی مورد نظر در متن وجود دارند.

r : تعداد کلماتی که به درستی به عنوان نمونه‌ی کلمه‌ی پرسشی آشکار شده‌اند.

N: تعداد کل کلماتی که به عنوان کلمه‌ی پرسشی مورد نظر آشکار شده‌اند.

با توجه به توضیحات بالا n بردار صحیح وجود دارد. با توجه به اینکه فاصله‌ی نزدیک‌ترین تصاویر مجموعه کلمات کاندید به کلمه‌ی پرسشی نزدیک به یک است، ماتریس فاصله (Distance) بصورت نزولی مرتب شده و متناسب با آن برداری شامل شماره تصاویر مجموعه کلمات کاندید بدست می‌آید. با مقایسه‌ی این بردار با هر یک از n بردار صحیح، رتبه[1] متناسب با تصویر صحیح و تصاویر تطبیقی برای هر نمونه مشخص می‌شود. این رتبه نشان‌دهنده‌ی این است که تصاویر صحیح و تطبیقی

[1] Rank

متناسب با نمونه‌ی کلمه‌ی پرسشی در متن به عنوان چندمین تصویر نزدیک به کلمه‌ی پرسشی تشخیص داده شده‌اند. پایین‌ترین رتبه به عنوان بهترین حالت در نظر گرفته می‌شود.

در مرحله‌ی بعدی ۱۳ آستانه در نظر گرفته شده، و هر بار با توجه به آستانه‌ی مورد نظر، تصاویری که فاصله آنها تا کلمه‌ی پرسشی از این آستانه بیشتر باشد به عنوان تصاویر آشکار شده‌ی سیستم در نظر گرفته می‌شوند. تعداد این کلمات همان N است. برای بدست آوردن r، پایین‌ترین رتبه برای هر نمونه از کلمه‌ی پرسشی در متن در نظر گرفته شده و تعداد رتبه‌هایی که از N کمتر هستند برابر با r می‌باشد.

اولین آستانه‌ی در نظر گرفته شده برابر با ۰/۹۹۲۵ است و ۱۳ آستانه‌ی دیگر به ترتیب با کم شدن از آستانه‌ی قبلی به میزان ۰/۰۱۲۵ بدست می‌آیند بطوریکه آخرین آستانه برابر با ۰/۸۴۲۵ است. برای ارزیابی کارایی روش معیارهای زیر تعریف شده‌اند که در هر آستانه محاسبه می‌شوند. در ادامه روش محاسبه‌ی هر یک از این معیارها بیان می‌شود:

۱) دقت[1]: دقت تشخیص یک کلمه‌ی پرسشی، در تمامی تحقیقات صورت گرفته در این زمینه طبق رابطه‌ی (۴-۲) محاسبه می‌شود.

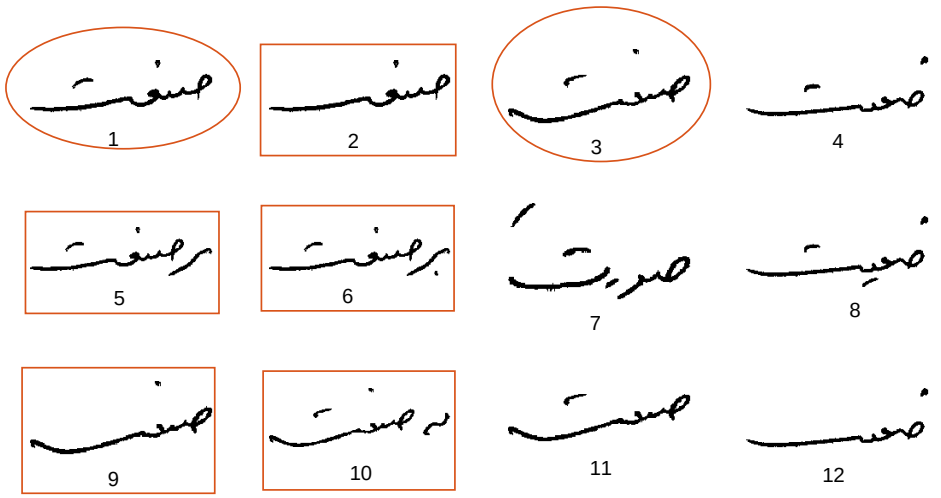
$$P = \frac{r}{N} \qquad (۴-۲)$$

در این تحقیق به دلیل وجود تصاویر تطبیقی این معیار نیاز به اصلاح دارد. این مساله در مثال زیر توضیح داده شده است:

در این مثال فرض شده است که کلمه‌ی پرسشی ورودی "صنعت" است و دو نمونه از این کلمه

[1] Precision(P)

در متن وجود دارد. سیستم N=12 کلمه‌ی موجود در شکل ۴-۳ را آشکار کرده است. اعداد ۱ تا ۱۲ رتبه مربوط به این کلمات آشکار شده را نشان می‌دهد. در این مثال دو کلمه‌ای که با بیضی مشخص هستند، دو نمونه‌ی صحیح متناظر با کلمه‌ی پرسشی می‌باشند و کلماتی که با مستطیل مشخص شده‌اند کلمات تطبیقی متناظر هستند. کلمات با رتبه ۱، ۲، ۵ و۶ مربوط به یک نمونه و کلمات ۳، ۹ و ۱۰ مربوط به نمونه‌ی دیگر "صنعت" در متن می‌باشند.



شکل ۴-۳: کلمات آشکار شده توسط سیستم برای کلمه‌ی "صنعت"

در این حالت برای محاسبه‌ی r، پایین‌ترین رتبه برای هر نمونه که ۱ و۳ می‌باشد در نظر گرفته شده و چون هر دو این اعداد از ۱۲ کوچکتر هستند پس در نتیجه r=2 است. در این سیستم کلمات تطبیقی آشکار شده به عنوان خطا در نظر گرفته نمی‌شوند. در این حالت فقط ۵ کلمه با رتبه ۴، ۷، ۸، ۱۱ و ۱۲ که مربوط به دو کلمه‌ی مشابه "وضعیت" و "صورت" می‌باشند به عنوان مواردی اشتباه آشکار شده‌اند به همین دلیل برای محاسبه‌ی دقت تعداد کلمات آشکار شده از ۱۲ به ۷ کاهش پیدا می‌کند. در واقع پارامتر جدیدی به صورت زیر تعریف شده که معیار دقت با توجه به آن با رابطه‌ی (۴-۳) اصلاح می‌شود:

N' : تعداد کل کلمات آشکار شده منهای تعداد تصاویر تطبیقی آشکار شده متناظر با کلمه‌ی پرسشی

$$P=\frac{r}{N'} \qquad (۳-۴)$$

۲)**فراخوانی**[1]: فراخوانی یک کلمه‌ی پرسشی از رابطه‌ی (۴-۴) محاسبه می‌شود.

$$R=\frac{r}{n} \qquad (۴-۴)$$

۳)**نرخ F_1**: این معیار توازن بین دقت و فراخوانی را منعکس می‌کند. زیرا اگر دقت بالا باشد فراخوانی کاهش پیدا می‌کند و اگر دقت پایین آید نرخ فراخوانی بالاتر می‌رود. روش‌هایی مناسب هستند که میزان دقت و فراخوانی آنها در یک حد باشد که در این حالت F_1 بیشترین مقدار را دارد. از این رو نرخ F_1 در هر آستانه‌ای که بیشترین مقدار را داشته باشد، همان آستانه به عنوان بهترین آستانه در نظر گرفته می‌شود. F_1 با رابطه‌ی (۴-۵) محاسبه می‌شود.

$$F_1=\frac{2\times P\times R}{P+R} \qquad (۵-۴)$$

برای هر کلمه‌ی پرسشی در یک متن دقت، فراخوانی و F_1 در هر آستانه محاسبه می‌شود. سپس این اطلاعات به همراه برچسب کلمه‌ی پرسشی ذخیره می‌شوند. برای تمام کلماتی که دارای برچسب مشابه هستند، در واقع کلماتی که مربوط به یک کلاس هستند، در هر آستانه، این معیارها میانگین‌گیری می‌شوند. تا در نهایت مجموعه‌ای حاصل شود که کلاس کلمات همه از هم متفاوت باشند و برای ارزیابی کلی، در هر آستانه در این مجموعه میانگین‌گیری می‌شود.

۴)**FP**[2]: به صورت تعداد کلماتی تعریف می‌شود که کلمه‌ی پرسشی مورد نظر نبوده ولی سیستم آن را آشکار کرده است. این تعداد برابر با N'-r می‌باشد. (در مثال بالا FP=5)

[1] Recall(R)
[2] False Positive

۵)**FN**[1]: به صورت تعداد کلماتی تعریف می‌شود که نمونه‌ای از کلمه‌ی پرسشی در متن هستند ولی توسط سیستم آشکار نشده‌اند. این تعداد برابر با n-r می‌باشد. (در مثال بالا FN=0)

۶)**TP**[2]: به صورت تعداد کلماتی تعریف می‌شود که نمونه‌ی از کلمه‌ی پرسشی در متن هستند و توسط سیستم آشکار شده‌اند. این تعداد برابر با r می‌باشد. (در مثال بالا TP=2)

۷)**TN**[3]: به صورت تعداد کلماتی تعریف می‌شود که در مجموعه‌ی کاندید، نمونه‌ای از کلمه‌ی پرسشی مورد نظر نبوده و توسط سیستم هم آشکار نشده است. با توجه به وجود پنجره‌های تطبیقی، ابتدا باید تعداد کل کلماتی که در مجموعه‌ی کاندید وجود دارد از تعداد کل تصاویر تطبیقی منتاظر با کلمه‌ی پرسشی کم شود که اگر این تعداد برابر با (W') در نظر گرفته شود، تعداد TN از رابطه‌ی (۴-۶) محاسبه می‌شود.

$$TN= W'-FP-TP-FN \qquad (۶-۴)$$

معیارهای دیگر توسط TN، TP، FN و FP محاسبه می‌شوند. این معیارها در جدول ۴-۱ معرفی شده‌اند.

جدول ۴-۱: تعریف معیارها

Nagative Predictive Value	NPV=TN/TN+FN
Specificity or true negative rate	SPC=TN/TN+FP
Accuracy	ACC=TP+TN/TP+FP+FN+TN

تنها تفاوت روش پرسش با تصویر و روش پرسش با رشته‌ی متنی در این است که در روش پرسش با رشته‌ی متنی، در فضای برداری مشترک برای هر کلمه‌ی پرسشی، فاصله بین برچسب آن، با

[1] False Negative
[2] True Positive
[3] True Negative

تصویر هر یک از نمونه‌های مجموعه کلمات کاندید محاسبه می‌شود.

۴-۲-۲ روش پیشنهادی برای کاهش کلمات در مجموعه‌ی کاندید

همان‌طور که در بخش قبلی بیان شد هر کلمه‌ی پرسشی با تمام کلمات موجود در مجموعه‌ی کاندید مقایسه می‌شود. ولی روش مناسب‌تر کاهش مجموعه کلمات کاندید متناسب با تعداد حروف هر کلمه‌ی پرسشی است. به طور مثال برای کلمه‌های پرسشی مانند "سال" یا "سه" یا حروف اضافه-ای مانند "در" یا "با" که دارای دو حرف می‌باشند، مقایسه با کلماتی مثل "سالیانه" یا "مسئولیت" که دارای هفت حرف هستند منطقی نیست.

روش استخراج کلمات تصویری که در فصل ۵ بیان خواهد شد بر پایه‌ی استخراج مولفه‌های متصل است. در این روش با در نظر گرفتن این قانون که اکثر حروف اضافه و کلمات فارسی از ترکیب یک تا هفت مولفه‌ی متصل تشکیل شده‌اند استخراج کلمات از داخل متن انجام می‌شود. در این روش متناسب با هر تصویر کلمه‌ای که استخراج می‌شود تعداد مولفه‌های متصل آن نیز مشخص است.

با توجه به این مساله، برای تمام کلمات پرسشی و کلمات مجموعه‌ی کاندید، تعداد مولفه‌های متصل آنها نیز به عنوان یک پارامتر در نظر گرفته می‌شود. با فرض اینکه تعداد مولفه‌های متصل کلمه‌ی پرسشی ورودی برابر با NCC باشد مرحله‌ی کاهش مجموعه کلمات کاندید به این صورت انجام می‌شود:

- اگر NCC= 1 باشد کلماتی از مجموعه‌ی کاندید برای مقایسه انتخاب می‌شوند که تعداد مولفه‌های متصل آنها بین ۱ تا ۳ باشد.
- اگر NCC= 2 باشد کلماتی از مجموعه‌ی کاندید برای مقایسه انتخاب می‌شوند که تعداد مولفه‌های متصل آنها بین ۱ تا ۴ باشد.

- اگر NCC= 6 باشد کلماتی از مجموعه‌ی کاندید برای مقایسه انتخاب می‌شوند که تعداد مولفه‌های متصل آنها بین ۴ تا ۷ باشد.
- اگر NCC= 7 باشد کلماتی از مجموعه‌ی کاندید برای مقایسه انتخاب می‌شوند که تعداد مولفه‌های متصل آنها بین ۵ تا ۷ باشد.
- اگر 3<=NCC<=5 باشد کلماتی از مجموعه‌ی کاندید برای مقایسه انتخاب می‌شوند که تعداد مولفه‌های متصل آنها در بازه‌ی [NCC-2,NCC+2] قرار گیرد.

استفاده از این روش باعث کاهش حجم محاسبات می‌شود.

۴-۲-۳ نتایج مرحله‌ی جستجوی کلمات

برای انجام مرحله‌ی ارزیابی از پایگاه داده‌ی FHT، ۱۰ نوع متن متفاوت انتخاب شده است که از هر نوع متن،۱۰دست‌خط متفاوت وجود دارد. به طور کلی برای ارزیابی این روش از ۱۰۰ سند تصویری استفاده شده است. جزئیات مربوط به تعداد کلمات پرسشی به صورت زیر می‌باشد:

مجموعه‌ای شامل ۵۵ کلمه‌ی متمایز به عنوان کلمات پرسشی انتخاب شده‌اند. ۱۰ متن انتخابی دارای موضوعات تقریبا مشابهی هستند تا فراوانی کلمات پرسشی داخل آنها زیاد باشد به همین دلیل در مجموعه‌ی کلمات پرسشی، ۲۴ کلمه در بیشتر از یک متن ظاهر شده‌اند. در مجموع تعداد این ۲۴ کلمه در داخل ۱۰ متن، ۷۸ می‌باشد. با در نظر گرفتن این ۷۸ کلمه و تعداد کلمات متمایز(۳۱ کلمه)، تعداد کل کلمات پرسشی که برای ۱۰ متن در نظر گرفته شده، ۱۰۹ می‌باشد. برای هر کدام از این ۱۰۹ کلمه ، حداقل ۱۰ سند با دست‌خط متفاوت وجود دارد. بیشترین فراوانی یک کلمه در داخل یک متن ۷ و کمترین آن یک است. جدول ۴-۲ تعداد کلمات با فراوانی بین ۱ تا ۷ را نشان می‌دهد. با در نظر گرفتن تعداد فراوانی کلمات، به طور کلی ۱۸۱۰ نمونه کلمه در این ۱۰۰ سند وجود دارد که این روش به جستجوی آنها پرداخته است. متوسط تعداد نمونه‌های کلمات پرسشی در متن برابر با ۱/۵۶۵۱می‌باشد. این الگوریتم در روش پرسش با تصویر، ۱۲۵۰ نمونه از کلمات پرسشی موجود در

اسناد تصویری را رتبه اول آشکار کرده است که معادل با ۶۹ ٪ می‌باشد و در روش پرسش بـا رشـته‌ی متنی موفق به آشکارسازی ۱۳۱۲ نمونه از کلمات پرسشی در رتبه اول شده است کـه معـادل بـا ۷۲ ٪ است. مجموعه‌ی کلمات پرسشی از جهت تعداد حروف تشکیل دهنـده‌ی کلمـه متنـوع مـی‌باشـند بـه طوری که شامل کلماتی با تعداد حروف کم مثل "ما"، "در"، "شد" و "دو"، کلمـات بـا تعـداد حـروف متوسط مثل"دارو"،"کشور"، "مردم" و "درآمد" و کلمـاتی بـا تعـداد حـروف زیـاد مثـل "مـدیریت"، "اولویت"،" اقدامات" و "مستقیم" است.

جدول ۴-۲: تعداد کلمات با فراوانی بین ۱ تا ۷

تعداد نمونه‌های کلمات در متن	۱	۲	۳	۴	۵	۶	۷
تعداد کلمات	۶۷	۲۷	۷	۴	۲	۱	۱

در جدول ۴-۳ متوسط معیارهای تعریف شده در بخش ۴-۲-۱ بر روی ۱۰۰ سند دسـت‌نوشـته‌ی فارسی برای پرسش با تصویر و متن در آستانه‌های متفاوت نشان داده شده است. در ایـن جـداول پـنج آستانه‌ی اول و آستانه‌ی یازدهم و سیزدهم در نظر گرفته شده‌اند.

جدول ۴-۳ : متوسط معیارهای کارایی روش بر روی ۱۰۰ سند متنی در آستانه‌های متفاوت برای حالت پرسش با تصویر و متن

	Performance Criterion	Threshold						
		۰/۹۹۲۵	۰/۹۸	۰/۹۶۷۵	۰/۹۵۵	۰/۹۴۲۵	۰/۸۶۷۵	۰/۸۴۲۵
QBE	Precision	**۸۱/۱۹**	۷۹/۹۶	۷۸/۵۱	۷۶/۹۷	**۷۵/۴۵**	۵۶/۷۸	۴۹/۶۱
	Recall	۶۷/۰۹	۷۰/۱۲	۷۳/۵۹	۷۷	**۸۰/۴۲**	۹۱/۱۷	**۹۳/۴۷**
	F_1	۷۰/۴۱	۷۱/۲۷	۷۱/۹۴	۷۲/۵۱	**۷۲/۵۴**	۶۰/۸۴	۵۴/۸۴
QBS	Precision	**۸۳**	۸۲/۸۰	۸۲/۸۲	۸۱/۳۹	**۸۰/۲۸**	۶۶/۰۳	۵۸/۶۲
	Recall	۶۷/۳۴	۷۰/۴۹	۷۳/۵۶	۷۶/۱۹	**۷۸/۹۲**	۹۰/۷۱	**۹۳/۱۴**
	F_1	۷۱/۵۲	۷۳/۴۰	۷۴/۶۷	۷۵/۵۴	**۷۵/۹۲**	۶۹/۰۲	۶۳/۵۷

با کاهش آستانه تعداد کل کلماتی که توسط الگوریتم آشکار شده‌اند افزایش می‌یابد به گونه‌ای که برای حالت پرسش با تصویر، در بیشترین آستانه متوسط این تعداد برابر با ۱/۱۱ و در کمترین آستانه برابر با ۱۸/۶۴ است. بنابراین با کاهش آستانه، متوسط دقت کاهش می‌یابد و بیشترین مقدار آن در آستانه اول حاصل شده که برابر با ۸۱/۱۹٪ است. همچنین با کاهش آستانه، تعداد کلماتی که به درستی به عنوان نمونه‌ای از کلمه‌ی پرسشی آشکار شده‌اند (TP) افزایش داشته و از تعداد کلماتی که نمونه‌ای از کلمه‌ی پرسشی هستند ولی الگوریتم قادر به شناسایی آنها نبوده (FN) کاسته می‌شود به گونه‌ای که بیشترین مقدار فراخوانی در آستانه‌ی آخر حاصل شده و برابر با ۹۳/۴۷٪ است. بیشترین مقدار F_1 در آستانه‌ی ۰/۹۴۲۵ بدست آمده است که برابر با ۷۲/۵۴٪ می‌باشد که در این آستانه میزان دقت ۷۵/۴۵٪ و میزان فراخوانی ۸۰/۴۲٪ است. این آستانه به عنوان بهترین آستانه انتخاب می‌شود. برای حالت پرسش با متن بیشترین مقدار F_1 در آستانه‌ی ۰/۹۴۲۵ بدست آمده است که برابر با ۷۵/۹۲٪ می‌باشد که در این آستانه میزان دقت ۸۰/۲۸٪ و میزان فراخوانی ۷۸/۹۲٪ است. در مقایسه

با حالت پرسش با تصویر، F_1 به میزان ۳/۳۸ و دقت به میزان ۴/۸۳ افزایش و فراخوانی به میزان ۱/۵ کاهش یافته است.

سه معیار SPC، ACC و NPV در جدول ۴-۴ گزارش شده‌اند.

جدول ۴-۴: معیارهای کارایی روش بر روی ۱۰۰ سند متنی در آستانه‌های متفاوت برای حالت پرسش با تصویر و متن

	Performance Criterion	Threshold						
		۰/۹۹۲۵	۰/۹۸	۰/۹۶۷۵	۰/۹۵۵	**۰/۹۴۲۵**	۰/۸۶۷۵	۰/۸۴۲۵
QBE	ACC	۹۹/۹۳	۹۹/۹۳	۹۹/۹۲	۹۹/۹۱	**۹۹/۸۹**	۹۹/۴۳	۹۸/۹۶
	SPC	۹۹/۹۸	۹۹/۹۷	۹۹/۹۶	۹۹/۹۴	**۹۹/۹۲**	۹۹/۴۴	۹۸/۹۳
	NPV	۹۹/۹۵	۹۹/۹۶	۹۹/۹۶	۹۹/۹۷	**۹۹/۹۷**	۹۹/۹۹	۹۹/۹۹
QBS	ACC	۹۹/۹۴	۹۹/۹۴	۹۹/۹۴	۹۹/۹۳	**۹۹/۹۲**	۹۹/۴۳	۹۸/۹۳
	SPC	۹۹/۹۹	۹۹/۹۸	۹۹/۹۷	۹۹/۹۶	**۹۹/۹۴**	۹۹/۴۳	۹۸/۹۴
	NPV	۹۹/۹۶	۹۹/۹۶	۹۹/۹۷	۹۹/۹۷	**۹۹/۹۷**	۹۹/۹۹	۹۹/۹۹

با توجه به جدول ۴-۴، یکی از مزیت‌های الگوریتم، بالا بودن TN است. به عبارت دیگر از بین مجموعه کلمات کاندید الگوریتم قادر است تعداد زیادی از کلمات را به عنوان مواردی که کلمه‌ی پرسشی مورد نظر کاربر نبوده‌اند به درستی رد کند. به همین دلیل مقادیر SPC و NPV بزرگ می‌باشند. با توجه به مقادیر گزارش شده در جدول ۴-۴، بزرگی SPC نشان‌دهنده‌ی این است که تعداد کلماتی که توسط الگوریتم به درستی پذیرفته نشده‌اند (TN) نسبت به تعداد کلماتی که اشتباها آشکار شده‌اند (FP) بسیار بزرگتر می‌باشد. این معیار با کاهش آستانه تغییرات جزئی دارد. بزرگی NPV نشان‌دهنده‌ی این است که تعداد کلماتی که کلمه‌ی پرسشی مورد نظر بوده‌اند ولی توسط سیستم آشکار نشده‌اند (FN) نسبت به تعداد کلماتی که توسط الگوریتم به درستی پذیرفته نشده‌اند

بسیار اندک می‌باشد. ACC نسبت مجموع تعداد نتایج مثبت (TP و TN) به کل مجموعه‌ی کاندید می‌باشد. در این آزمایشات این معیار تقریبا ۹۹ ٪ است. که بزرگی آن نشان‌دهنده‌ی این است که مجموع تعداد نتایج منفی (FN وFT) کم می‌باشد.

جدول ۴-۵ تا جدول ۴-۷ تعداد TP، TN، FP، FN و تعداد کلمات آشکار شده (N') را برای سه کلمه‌ی پرسشی در آستانه‌های متفاوت نشان می‌دهد. برای هر کلمه‌ی پرسشی تعداد مجموعه کلمات کاندید(W') هم گزارش شده است.

جدول ۴-۵: تعداد معیارها برای کلمه‌ی پرسشی ملی (W'=1142)

Performance Criterion	**Threshold**						
	۰/۹۹۲۵	۰/۹۸	۰/۹۶۷۵	۰/۹۵۵	۰/۹۴۲۵	۰/۸۶۷۵	۰/۸۴۲۵
N'	۱	۱	۲	۳	۳	۴	۴
TP	۱	۱	۲	۳	۳	۴	۴
TN	۱۱۳۸	۱۱۳۸	۱۱۳۸	۱۱۳۸	۱۱۳۸	۱۱۳۸	۱۱۳۸
FP	۰	۰	۰	۰	۰	۰	۰
FN	۳	۳	۲	۱	۱	۰	۰

جدول ۴-۶: تعداد معیارها برای کلمه‌ی پرسشی برای (W'=1868)

Performance Criterion	Threshold						
	۰/۹۹۲۵	۰/۹۸	۰/۹۶۷۵	۰/۹۵۵	۰/۹۴۲۵	۰/۸۶۷۵	۰/۸۴۲۵
N'	۱	۲	۲	۲	۲	۴	۴
TP	۱	۲	۲	۲	۲	۲	۲
TN	۱۸۶۶	۱۸۶۶	۱۸۶۶	۱۸۶۶	۱۸۶۶	۱۸۶۴	۱۸۶۴
FP	۰	۰	۰	۰	۰	۲	۲
FN	۱	۰	۰	۰	۰	۰	۰

جدول ۴-۷: تعداد معیارها برای کلمه‌ی پرسشی تولید (W'=1652)

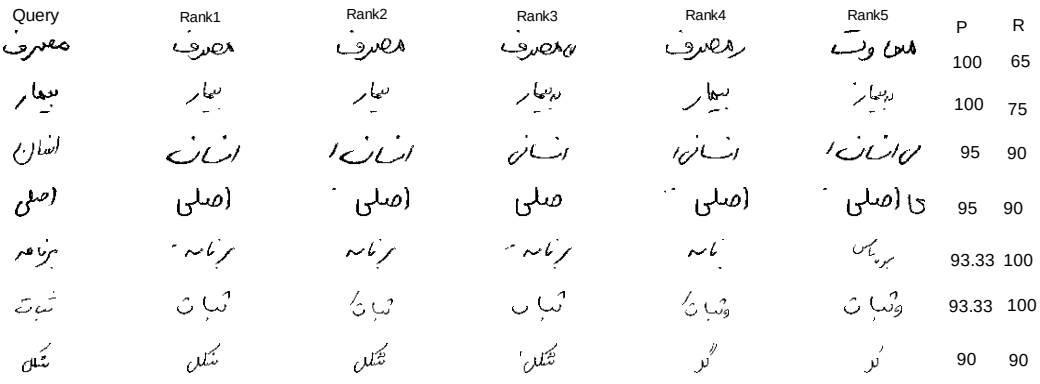
Performance Criterion	Threshold						
	۰/۹۹۲۵	۰/۹۸	۰/۹۶۷۵	۰/۹۵۵	۰/۹۴۲۵	۰/۸۶۷۵	۰/۸۴۲۵
N'	۱	۱	۱	۲	۳	۳	۳
TP	۱	۱	۱	۲	۳	۳	۳
TN	۱۶۴۹	۱۶۴۹	۱۶۴۹	۱۶۴۹	۱۶۴۹	۱۶۴۹	۱۶۴۹
FP	۰	۰	۰	۰	۰	۰	۰
FN	۲	۲	۲	۱	۰	۰	۰

همان طور که در این جداول مشخص است، تعداد TN زیاد و تعـداد FP وFN تعـداد کمـی مـی-باشند که این مسئله کارایی الگوریتم را نشان می‌دهد.

در شکل ۴-۴ و شکل ۴-۵ ، سیزده کلمه‌ی پرسشی که دارای بیشترین متوسط دقت در مجموعـه

کلمات پرسشی هستند نشان داده شده‌اند. در این شکل‌ها، ستون اول مربوط به تصویر کلمه‌ی پرسشی است. در ستون دوم تا ششم، پنج تصویر اولی که توسط الگوریتم آشکار شده‌اند نمایش داده شده است. همان طور که مشخص است الگوریتم موفق به یافتن نمونه‌های کلمات پرسشی از بین مجموعه کلمات کاندید شده است. ستون هفتم و هشتم به ترتیب دقت متوسط و فراخوانی را روی کلاس کلمه‌ی پرسشی مورد نظر نشان می‌دهد. متوسط تعداد کلمات آشکار شده در آستانه‌ی مورد نظر برابر با ۲/۸ است. به همین دلیل تشخیص تصاویر صحیح و تطبیقی متناظر با کلمه‌ی پرسشی بین سه تصویر اول آشکار شده توسط الگوریتم موجب افزایش دقت می‌شود. در مورد کلمات پرسشی شکل ۴-۴ و شکل ۴-۵ این مساله برقرار است.

شکل ۴-۴: کلمات پرسشی که دارای بالاترین دقت متوسط می‌باشند.



شکل ۴-۵: کلمات پرسشی که دارای بالاترین دقت متوسط می‌باشند.

در شکل ۴-۶، کلماتی که چند بار در متن ظاهر شده و دارای تعداد نمونه‌ی بین ۳ تا ۷ در یک متن می‌باشند نمایش داده شده‌اند. ستون یک تا هفت مربوط به هفت تصویر اولی است که توسط الگوریتم آشکار شده است. ستون هشتم تعداد نمونه‌های یک کلمه در متن را نشان می‌دهد. همان طور که مشخص است الگوریتم قادر به آشکارسازی تمام نمونه‌های کلمه‌ی پرسشی موجود در متن شده است. این نمونه‌ها با کادر مستطیلی مشخص شده‌اند.

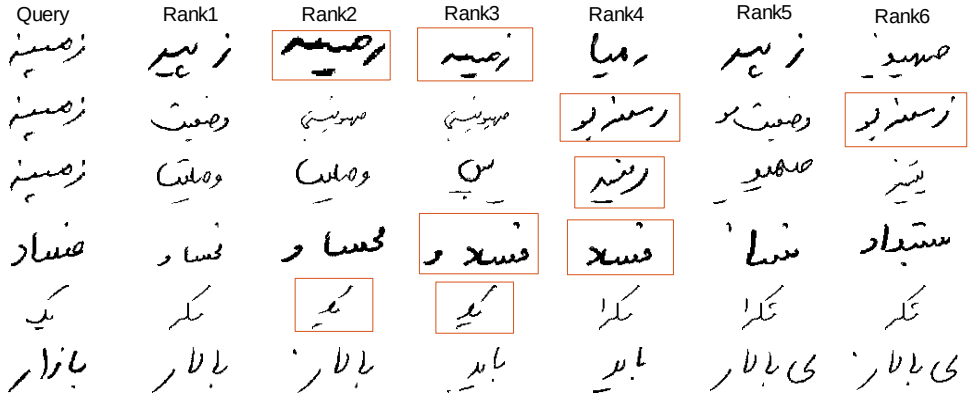
شکل ۴-۶: آشکارسازی تمام نمونه‌های متناظر با کلمه‌ی پرسشی موجود در متن

در شکل ۴-۷ و شکل ۴-۸، کلماتی که دارای پایین‌ترین دقت متوسط در مجموعه کلمات پرسشی هستند نمایش داده شده‌اند. دلیل کاهش دقت متوسط، وجود کلمات مشابه با این کلمات پرسشی در متن می‌باشد. در شکل ۴-۷، نمونه‌ی کلمه‌پرسشی مورد نظر در متن به عنوان اولین تصویر آشکار شده است ولی به دلیل وجود کلمات مشابه با کلمه‌ی پرسشی و تشخیص آنها در رتبه‌های بعدی، دقت متوسط در آستانه‌ی ۰/۹۴۲۵ کاهش یافته است. به طور مثال در شکل ۴-۷، در مورد کلمه‌ی پرسشی "فساد" وجود کلمه‌ای مانند "فحشا"، در مورد "منابع" وجود "صلاح"، در مورد "ما" وجود "ها"، در مورد "لاتین" وجود "این"، در مورد "سال" وجود "شمال" و ..... باعث کاهش دقت در این کلمات شده است.

دلیل دیگر برای کاهش متوسط دقت در بعضی از کلمات، ناخوانا نوشته شدن آن کلمه در متن

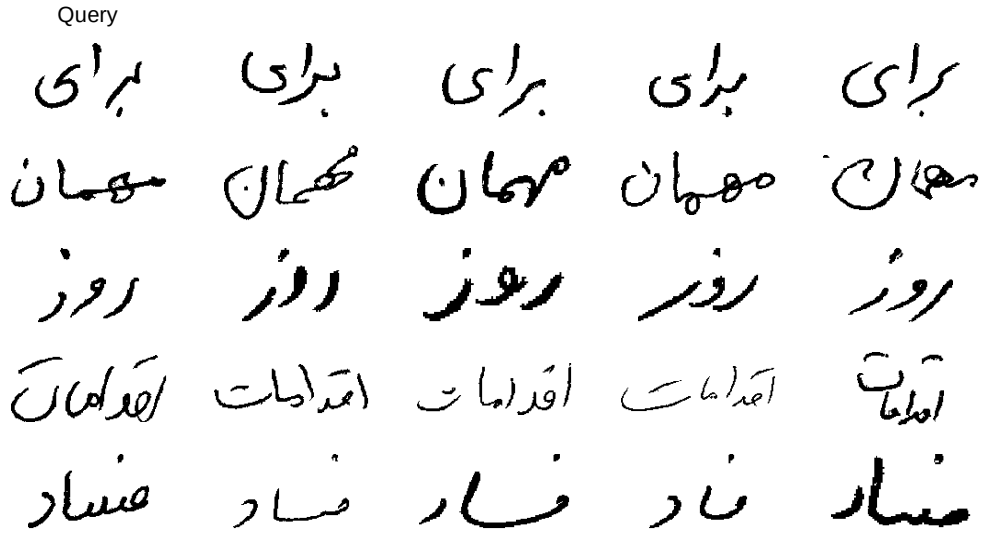
می‌باشد. بعضی از این موارد در شکل ۴-۸ نشان داده شده است. به طور مثال ناخوانا بودن کلمه‌ی "زمینه" یا "فساد" موجب عدم تشخیص آن در رتبه یک شده است. این کلمات با کادر مستطیلی مشخص شده‌اند.

شکل ۴-۷: کلمات پرسشی که دارای پایین‌ترین دقت متوسط هستند.

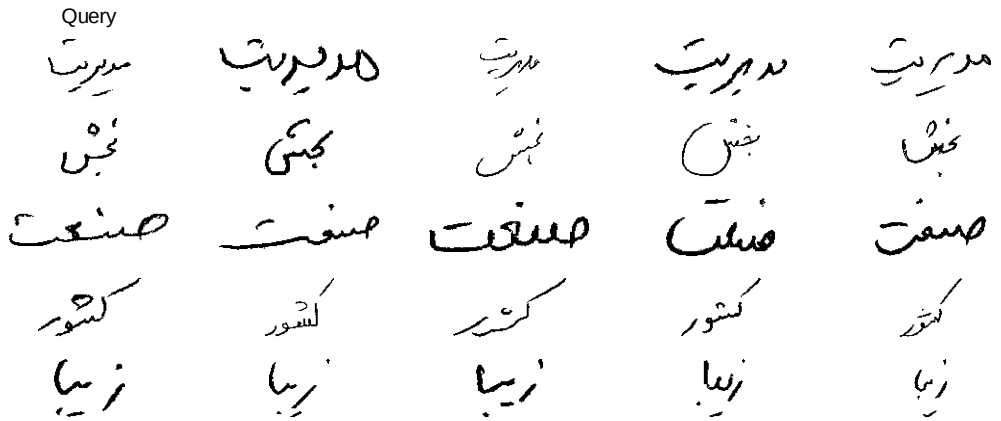


شکل ۴-۸: ناخوانا نوشته شدن بعضی از کلمات موجب عدم تشخیص آنها در رتبه یک شده است.

در شکل ۴-۹ و شکل ۴-۱۰، تصویر کلمه‌ی پرسشی و نمونه‌های آن که مربوط به پنج دست‌خط متفاوت هستند نمایش داده شده است. این روش موفق به تشخیص این نمونه‌ها به عنوان اولین تصویر متناظر با کلمه‌ی پرسشی در متن شده است. این تصاویر تنوع درون کلاسی کلمات را نشان می‌دهند. الگوریتم پیشنهادی با در نظر گرفتن این تنوع دورن کلاسی باز هم نتایج قابل قبولی ارائه می‌دهد.



شکل ۴-۹: نمونه‌های متناظر با کلمه‌ی پرسشی که الگوریتم موفق به تشخیص آنها شده است.

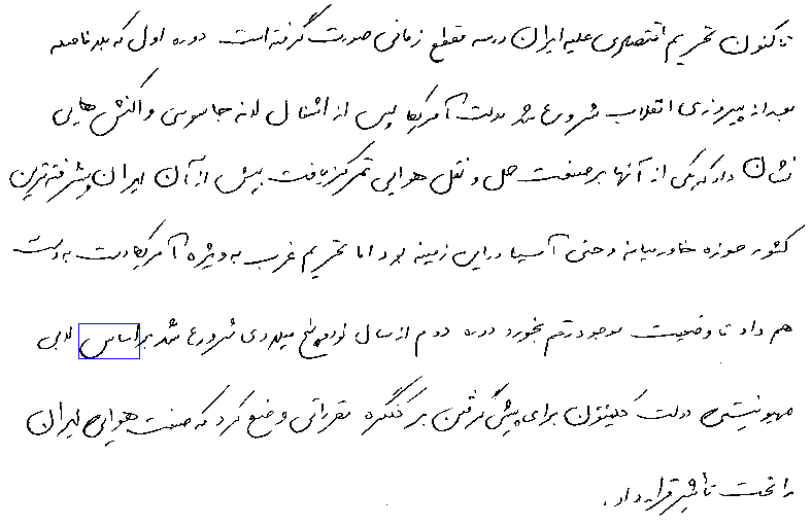


شکل ۴-۱۰: نمونه‌های متناظر با کلمه‌ی پرسشی که الگوریتم موفق به تشخیص آنها شده است.

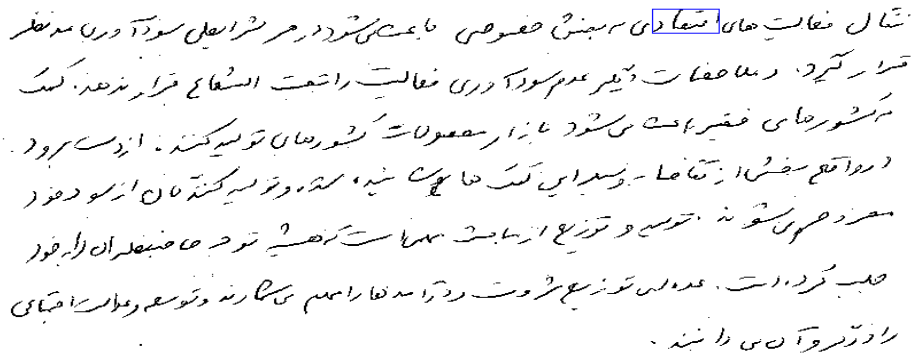
همان طور که قبلا نیز به آن اشاره شد استفاده از مشخصه‌ای مانند PHOC، موجب می‌شود الگوریتم بتواند کلمات دیده نشده در فاز آموزش را نیز تشخیص دهد. در مجموعه‌ی کلمات پرسشی برای کلماتی مانند "لاتین" و "فوتبال" آموزش صورت نگرفته ولی در مرحله‌ی آزمون با دقت متوسط ۳۳/۵۴ ٪ و ۸۵/۳۳ ٪ عملیات تشخیص برای این کلمات انجام شده است.

برای تمام تصاویر موجود در مجموعه‌ی کاندید، مختصات مکانی آنها در متن از قبل ذخیره شده است. با توجه به این موضوع، بعد از تشخیص نمونه‌ی کلمه‌ی پرسشی توسط الگوریتم، میتوان آن را

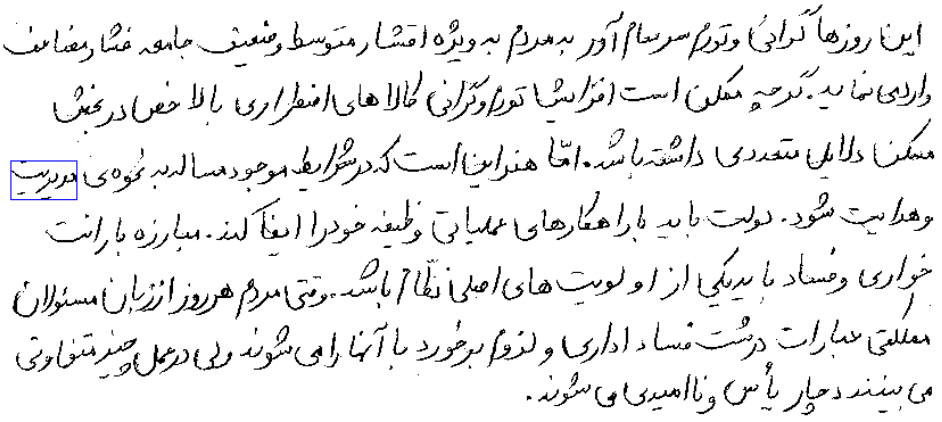
در متن مشخص کرد. این روش در شکل ۴-۱۱ تا شکل ۴-۱۳ نمایش داده شده است.



شکل ۴-۱۱:جستجوی کلمه‌ی اساس در متن



شکل ۴-۱۲:جستجوی کلمه‌ی اقتصاد در متن



شکل ۴-۱۳:جستجوی کلمه‌ی مدیریت در متن

۴-۲-۳-۱ زمان اجرای الگوریتم

مرحله‌ی آموزش روی پایگاه داده فارسا، مرحله‌ی جداسازی خطوط و استخراج دنباله‌ای از مولفه-های متصل روی پایگاه داده FHT بصورت برون‌خط انجام شده و نتایج آن ذخیره می‌شود. در مرحله‌ی آزمون با وارد شدن تصویر و رشته‌ی متنی کلمه‌ی پرسشی بصورت برخط عملیات جستجو انجام می-شود. زمان اجرای هر یک از مراحل برای کلمه‌ی پرسشی "کالا" بصورت زیر است:

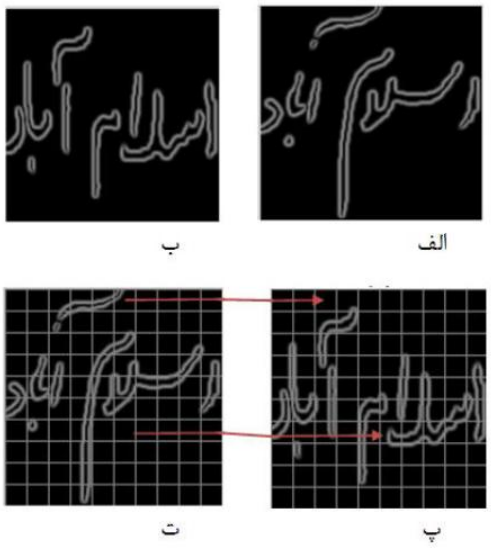
- زمان تعبیه رشته‌ی متنی در فضای برچسب با PHOC = ۰/۰۰۶۵ثانیه
- زمان استخراج ویژگی DSIFT از تصویر =۰/۰۳۱۰ثانیه
- زمان استخراج بردار فیشر= ۰/۰۲۲۲۱ثانیه
- زمان تعبیه بردار فیشر در فضای مشخصه= ۰/۰۱۹۳ ثانیه
- زمان تعبیه برچسب PHOC در فضای برداری مشترک=۰/۰۰۲۱ ثانیه
- زمان تعبیه رتبه مشخصه در فضای برداری مشترک=۰/۰۱۰۹ ثانیه
- زمان جستجوی کلمه‌ی پرسشی مورد نظر در بین مجموعه‌ی کاندید در زیر فضای عمومی=۴/۷ ثانیه

مرحله‌ی آخر شامل کاهش مجموعه تصاویر کاندید براساس روش بخش ۴-۲-۲ از تعداد ۲۸۶۳ به ۱۶۷۴ می‌باشد. سپس کلمه‌ی پرسشی مورد نظر با این ۱۶۷۴ تصویر مقایسه شده و نزدیک‌ترین آن به عنوان نمونه‌ی کلمه‌ی در نظر گرفته شده و مکان آن روی سند تصویری مشخص می‌شود.

۴-۳ روش پیشنهادی دوم (آموزش *GMM* با ناحیه‌ی‌بندی تطبیقی)

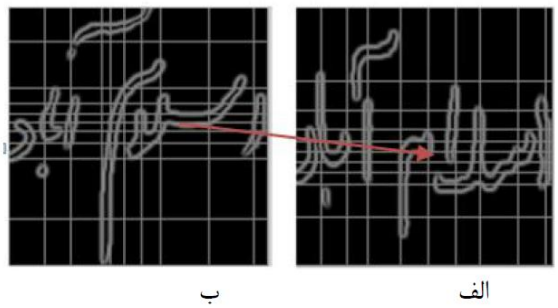
همان‌طور که در بخش ۳-۳-۲ بیان شد برای آموزش GMM از ویژگی‌ها کل تصویر بصورت یکجا استفاده نمی‌شود بلکه ابتدا تصویر به ۱۲ ناحیه تقسیم شده و ویژگی‌های هر ناحیه به صورت مستقل از هم در نظر گرفته می‌شوند و GMM روی هر ناحیه با توجه به ویژگی‌های همان ناحیه آموزش داده

می‌شود. در روش قبلی با توجه به اینکه مختصات مکانی ویژگی‌ها برای هر تصویر کلمه، بین ۰/۵و ۰/۵- نرمال شده است، شش نوار عمودی و دو نوار افقی روی تصویر ایجاد شد. مکان این نوارها برای تمامی تصاویر ثابت است. در این بخش برای آموزش GMM، ناحیه‌بندی تطبیقی پیشنهاد داده می-شود. برای ناحیه‌بندی تطبیقی از روش پیشنهادی در [۵۵] استفاده شده است. امکان وجود جابه‌جایی و تغییر مکان در کلمات دست‌نوشته، در ناحیه‌بندی با اندازه‌ی ثابت، باعث می‌شود پیکسل‌های قلم موجود در ناحیه‌های متناظر در تصاویر از یک جنس نباشند. برای حل این مساله از ناحیه‌بندی تطبیقی که تناسب بیشتری با دست‌نوشته فارسی دارد استفاده شده است. (شکل ۴-۱۴و شکل ۴-۱۵)



شکل ۴-۱۴: (الف و ب) تصاویر مربوط به یک کلاس

و (پ و ت) ناحیه‌بندی با اندازه‌ی ثابت [۵۵]



شکل ۴-۱۵: ناحیه‌بندی تطبیقی برای نمونه‌های یک کلاس [۵۵]

در این روش متناسب با تصویر هر کلمه مکان نوارهای عمودی و افقی تغییر می‌کند. برای ناحیه-بندی تطبیقی، ابتدا تصویر به دو نوار افقی تقسیم می‌شود به طوری که تعداد پیکسل‌های سیاه در این دو نوار افقی با هم برابر باشند سپس تصویر به شش نوار عمودی تقسیم می‌شود به طوری که تعداد پیکسل‌های سیاه در همه‌ی نوارهای عمودی با هم برابر باشند. این روش باعث می‌شود اندازه‌ی ناحیه-ها در هر تصویر کلمه با توجه به توزیع پیکسل‌های سیاه تغییر کند. مزیت این روش این است که اندازه‌ی ناحیه‌ها طبق مکان و نحوه‌ی نوشتن کلمه دست نوشته تغییر می‌کند و ناحیه‌های کوچکتر در اطراف خط اصلی کلمه دست‌نوشته ایجاد شده است.

۴-۳-۱نتایج روش آموزش *GMM* با ناحیه‌بندی تطبیقی

نتایج این روش در جدول ۴-۸ برای پنج آستانه‌ی اول نشان داده شده است. بیشترین مقدار F_1 در آستانه‌ی ۰/۹۵۵ بدست آمده است که برابر با ۷۴/۵۸٪ می‌باشد که در این آستانه میزان دقت ۷۹/۴۶٪ و میزان فراخوانی ۷۷/۴۱٪ است. مقایسه‌ی بین ناحیه‌بندی تطبیقی و ناحیه‌بندی ثابت در روش پرسش با تصویر برای دو حالت بازشناسی و جستجوی کلمه‌ی پرسشی در جدول ۴-۹ و جدول ۴-۱۰ آورده شده است. همان طور که جدول نشان می‌دهد دقت روش ناحیه‌بندی تطبیقی بهتر است.

جدول ۴-۸: متوسط معیارهای کارایی روش بر روی ۱۰۰ سند متنی در پنج آستانه‌های اول

Performance Criterion	**Threshold**				
	۰/۹۹۲۵	۰/۹۸	۰/۹۶۷۵	**۰/۹۵۵**	**۰/۹۴۲۵**
Precision	۸۳/۰۲	۸۱/۸۵	۸۱	**۷۹/۴۶**	۷۷/۰۹
Recall	۶۷/۷۲	۷۱/۰۹	۷۴/۶۱	**۷۷/۴۱**	۸۰/۳
F_1	۷۱/۶۶	۷۲/۹۸	۷۴/۲	۷۴/۵۸	۷۴/۳۹

جدول ۴-۹:مقایسه‌ی بین دو روش برای حالت جستجوی کلمه‌ی پرسشی در متن

روش	دقت	فراخوانی	F_1
ناحیه‌ی‌بندی تطبیقی	**۷۹/۴۶**	۷۷/۴۱	۷۴/۵۸
ناحیه‌ی بندی ثابت	**۷۶/۹۷**	۷۷	۷۲/۵۱

جدول ۴-۱۰: مقایسه‌ی بین دو روش برای حالت بازشناسی

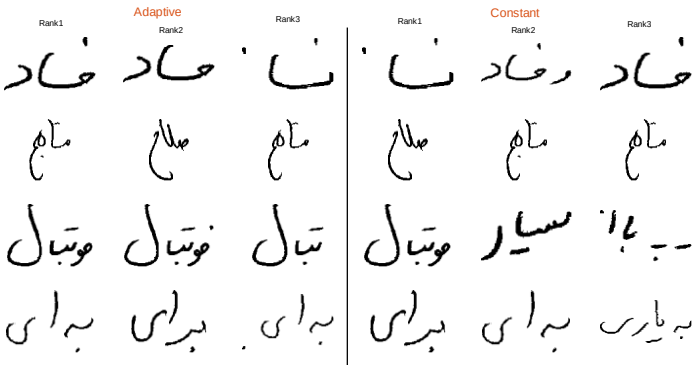
روش	P	MAP
ناحیه‌ی بندی تطبیقی	۹۸/۰۵	**۹۴/۶۶**
ناحیه‌ی بندی ثابت	۹۷/۴۱	**۹۳/۵**

این روش در مورد کلمات پرسشی که در روش ناحیه‌بندی با انـدازه‌ی ثابـت دارای دقـت متوسـط پایین بوده‌اند عملکرد بهتری داشته و دقت متوسط برای این کلمات را به میزان قابل توجه‌ای افـزایش داده است ولی در مورد کلماتی که در روش ناحیه‌بندی با اندازه‌ی ثابت نیـز دارای دقـت متوسـط بـالا هستند تفاوت قابل ملاحضه‌ای ایجاد نکرده است. برخـی از ایـن کلمـات در جـدول ۴-۱۱ نشـان داده شده‌اند.

جدول ۴-۱۱ : مقایسه بعضی از کلمات در دو روش ناحیه‌بندی تطبیقی و ثابت

کلمه‌ی پرسشی	زمینه	فساد	منابع	لاتین	شد	یک	جمهوری	فوتبال
متوسط دقت (تطبیقی)	۶۸/۳۳	۷۶/۶۷	۶۸/۶۱	۵۷/۳۳	۴۲/۲۳	۵۵/۸۶	۹۰	۱۰۰
متوسط دقت (ثابت)	۳۰/۶	۵۸/۳۳	۵۰/۷۱	۴۱/۷۰	۳۳/۰۳	۴۷/۷۲	۷۵	۸۵/۳۳

در شکل ۴-۱۶ برای هر کلمه‌ی پرسشی، سه تصویر اولی که توسط الگوریتم ناحیه‌بندی تطبیقی و ناحیه‌بندی ثابت آشکار شده‌اند نمایش داده شده است. همان طور که تصاویر نشان می‌دهند روش تطبیقی عملکرد بهتری دارد.



شکل ۴-۱۶:مقایسه‌ی دو روش

فصل پنجم:

الگوریتم پیشنهادی برای جداسازی خطوط و استخراج دنباله‌ای از مولفه‌های متصل

۵-۱ مقدمه

جداسازی خطوط متن یکی از اصلی‌ترین مراحل در بازشناسی نوری حروف است. عملکرد این مرحله بر تمام مراحل بعدی مانند قطعه‌بندی کلمات و بازشناسی متون تاثیرگذار است. به طور کلی بازشناسی متون در دو حوزه‌ی متفاوت صورت می‌گیرد:

۱) متون چاپی

۲) متون دست‌نویس

از آنجایی که حروف و کلمات در متون چاپی به صورت استاندارد هستند، جداسازی خطوط این گونه متون نیز ساده‌تر است. در متون چاپی معمولا فاصله‌ی بین کلمات، فاصله‌ی بین خطوط و شکل حروف به صورت منظم و یکسان است. کجی خطوط متن نیز در متون چاپی یا وجود ندارد یا بسیار کم است. اما در متون دست نوشته وضعیت کاملا متفاوت است. بر این اساس روش‌های ارائه شده برای این دو دسته نیز متفاوت است. از جمله مواردی که جداسازی متون دست‌نویس را مشکل می‌کند می‌توان به کج نوشته شدن خطوط و هم پوشانی بین حروف و خطوط اشاره کرد.

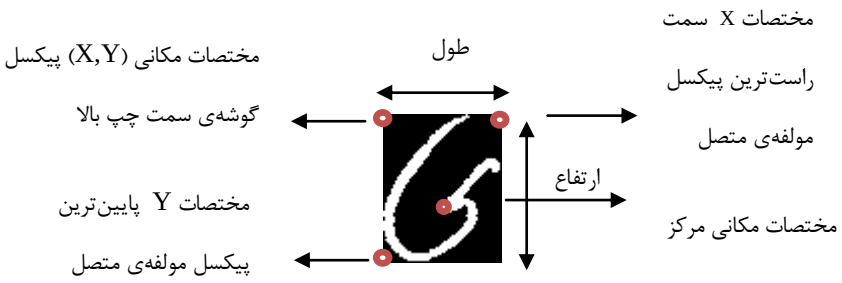
در این پایان نامه برای جستجوی کلمه‌ی کلیدی، یکی از اصلی‌ترین مراحل جداسازی خطوط متن می‌باشد. در مرحله‌ی بعدی می‌بایست کلمات و یا زیرکلمات موجود در سند تصویری و اطلاعات مکانی آنها استخراج شوند. استخراج کلمات می‌تواند بصورت برون‌خط یا برخط انجام شود؛ در روش برون‌خط فقط اطلاعات مکانی از تصاویر دست‌نوشته در دسترس است اما در روش برخط علاوه بر اطلاعات مکانی، اطلاعات زمانی، جهت و نحوه‌ی حرکت قلم مغناطیسی ورودی های سیستم می‌باشند. روش‌های برون‌خط هم شامل بازشناسی نوشتار چاپی و هم نوشتار دست‌نویس می‌شوند ولی روش‌های برخط فقط در بازشناسی دست‌نویس کاربرد دارند. این پایان نامه بر استخراج برون‌خط دنباله‌ای از مولفه‌های متصل متمرکز شده است.

روش پیشنهادی برای انجام این مرحله مستقل از نحوه‌ی نگارش می‌باشد که بر پایه‌ی استخراج

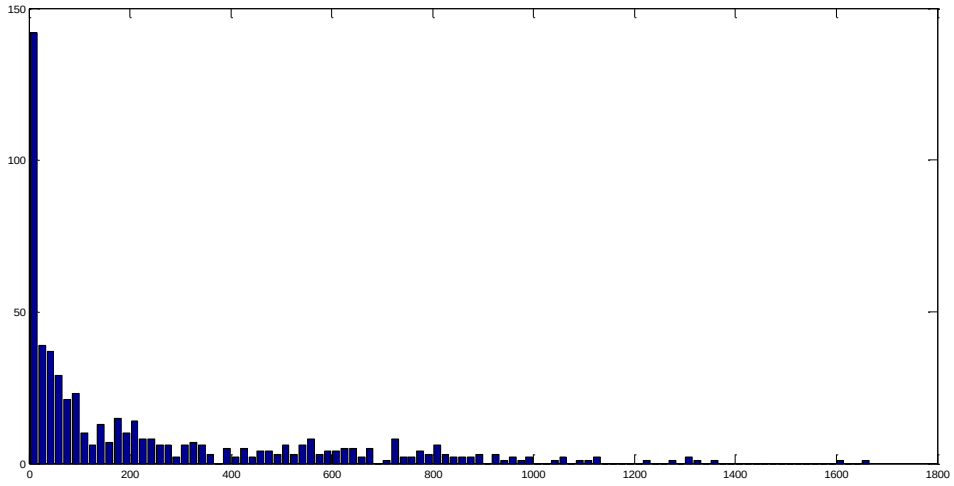
مولفه‌های متصل یک متن و اختصاص دادن هر مولفه‌ی متصل به خط مربوط به خودش است. سپس مولفه‌های متصلی که مربوط به یک کلمه مستقل در متن هستند به همان ترتیبی که در کلمه قرار دارند با این روش شناسایی می‌شوند تا در نهایت بتوان کلمات متن را از یکدیگر تفکیک کرد. در واقع این روش به استخراج دنباله‌ای از مولفه‌های متصل یک کلمه و قطعه‌بندی برون‌خط متون دست‌نویس فارسی در سطح کلمه می‌پردازد. در ادامه نحوه‌ی انجام مراحل بیان خواهد شد.

۵-۲ استخراج مولفه‌های متصل اصلی هر خط

تصاویر پایگاه داده‌ی مورد استفاده برای این مرحله بصورت باینری می‌باشند به طوری که پیکسل‌های پس زمینه سیاه و پیکسل‌های قلم سفید هستند. در این مرحله ابتدا تمام مولفه‌های متصل مربوط به متن استخراج می‌شوند. مختصات مرکز، بالاترین، پایین‌ترین، سمت راست‌ترین، سمت چپ‌ترین و تمام پیکسل‌های مربوط به هر مولفه متصل به همراه طول، عرض و تعداد پیکسل‌های آن به صورت جداگانه در یک لیست اولیه‌ای ذخیره می‌شود. این نقاط در شکل ۵-۱نشان داده شده‌اند. سپس هیستوگرام تعداد پیکسل‌های تمام مولفه‌های متصل موجود در متن با در نظر گرفتن ۱۰۰ بازه محاسبه می‌شود. مرکز اولین بازه در نظر گرفته شده و مولفه‌های متصلی که تعداد پیکسل آن از این حد کمتر است حذف می‌شوند. انجام این مرحله باعث حذف نویزهایی که به صورت نقاط کوچک در تصاویر وجود دارند می‌شود. تعداد مولفه‌ی متصل حذف شده در این مرحله قابل توجه می‌باشد که باعث کاهش محاسبات و افزایش دقت در مراحل بعدی است. در شکل ۵-۲ این هیستوگرام برای یکی از تصاویر نمایش داده شده است.



شکل ۵-۱: نقاطی که برای هر مولفه‌ی متصل ذخیره می‌شود.



شکل ۵-۲:هیستوگرام تعداد پیکسل‌های تمام مولفه‌های متصل موجود در متن

در ادامه با در نظر گرفتن اینکه در فارسی از بالا سمت راست نوشتن آغاز می‌شود؛ مولفه‌ی متصلی که مختصات مرکز آن دارای بیشترین مقدار X و کمترین مقدار Y است و تعداد پیکسل‌های آن از یک آستانه‌ی در نظر گرفته شده بیشتر می‌باشد، به عنوان مولفه‌ی متصل اولیه انتخاب می‌گردد.(در نظر گرفتن این آستانه به این علت است که مولفه‌ی متصلی کوچکی مانند نقطه به عنوان اولین انتخاب نشود. این آستانه برای تمام تصاویر ۱۵۰ در نظر گرفته شده است.)

برای انتخاب این مولفه‌ی متصل از رابطه‌ی (۵-۱) استفاده می‌شود. در این رابطه X و Y مختصات مرکز هستند و مقدار FCC به ازای تمام مولفه‌های متصل متن محاسبه شده است. مولفه‌ی

متصلی که دارای بیشترین مقدار FCC است شروع کننده‌ی اولین خط هر متن است.

$$FCC = 0.3X - Y \quad (۱-۵)$$

در مرحله‌ی بعدی نزدیک‌ترین مولفه‌ی متصل در متن به این مولفه‌ی متصل اولیه با در نظر گرفتن دو شرط زیر انتخاب می‌شود:

۱) مختصات X مرکز نزدیک‌ترین مولفه‌ی متصل باید از مختصات X مرکز مولفه‌ی متصل اولیه کمتر باشد.(تا حرکت همیشه از شروع یک خط به سمت انتهای آن باشد.)

۲) قدر مطلق اختلاف بین مختصاتY مراکز آنها از یک آستانه که برای همه متون ۸۵ در نظر گرفته شده است کمتر باشد.(این باعث می‌شود احتمال اینکه هر دو مربوط به یک خط باشند را افزایش دهد)

انتخاب نزدیک‌ترین مولفه‌ی متصل با محاسبه‌ی فاصله‌ی اقلیدسی است. در رابطه‌ی (۵-۲) X_{FCC} و Y_{FCC} مختصات مرکز مولفه‌ی متصل اولیه، X و Y مختصات مراکز سایر مولفه‌های متصل متن و D میزان فاصله بین آنها می‌باشد. کمترین فاصله مربوط به نزدیک‌ترین مولفه‌ی متصل است. با در نظر گرفتن شرط ۱ و ۲، نزدیک‌ترین مولفه‌ی متصل، در واقع مولفه‌ی متصلی است که در همین خط بعد از مولفه‌ی متصل اولیه قرار گرفته است. (با در نظر گرفتن حرکت از سمت راست به چپ خط)

$$D = \sqrt{(X_{FCC} - X)^2 + 10(Y_{FCC} - Y)^2} \quad (۲-۵)$$

با توجه به اینکه خطوط متن در راستای محور Y قرار دارند، در رابطه‌ی بالا ضریب ۱۰ برای این راستا

در نظر گرفته شده است تا مولفه‌های متصلی که مختصات Y آنها در یک محدوده است، در واقع مربوط به یک خط هستند، به عنوان نزدیک‌ترین انتخاب شوند. به عبارت دیگر این ضریب احتمال اینکه مولفه‌ی متصلی از خط بعدی به عنوان نزدیک‌ترین انتخاب شود را تا حدی کاهش می‌دهد.

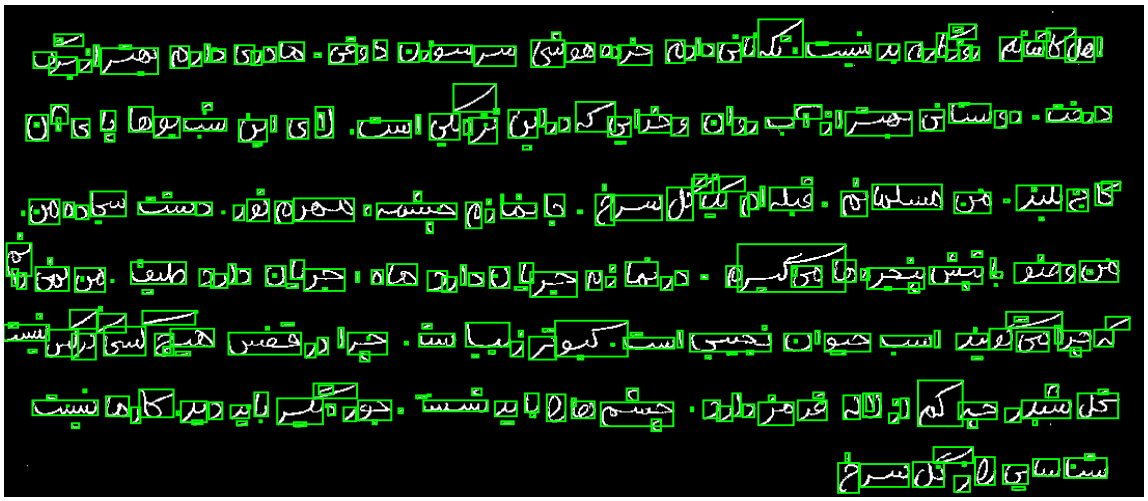
در ادامه نزدیکترین مولفه متصل بدست آمده در بالا به عنوان مولفه‌ی متصل اولیه قرار داده شده و با رابطه‌ی (۵-۲) و دو شرط بیان شده، مولفه‌ی متصل نزدیک به آن محاسبه می‌شود. در نهایت با تکرار این الگوریتم تمام مولفه‌های متصل خط اول بصورت یک زنجیره و به همان ترتیبی که در متن قرار دارند بدست می‌آیند. این الگوریتم تا جایی ادامه پیدا می‌کند که دیگر هیچ مولفه‌ی متصلی که در دو شرط بالا صدق کند وجود نداشته باشد.

خروجی این قسمت به عنوان لیستی از مراکز مولفه‌های متصل اصلی اولین خط ذخیره می‌شود. به دلیل قرار دادن شرط اول این مولفه‌های متصل با همان ترتیبی که در متن از سمت راست به چپ پشت سر هم قرار گرفته‌اند، در این لیست هم بدست آمده‌اند. مختصات بالاترین، پایین‌ترین، سمت راست‌ترین، سمت چپ‌ترین و تمام پیکسل‌های مربوط به این مولفه‌های متصل به همراه طول، عرض و تعداد پیکسل‌های آن به صورت جداگانه ذخیره می‌شوند. شکل ۵-۳ تا شکل ۵-۵ این مراحل را نشان می‌دهند.

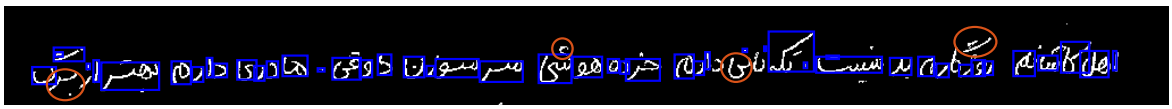
در شکل ۵-۳ محدوده‌ی هر مولفه‌ی متصل با رنگ سبز مشخص شده است. شکل ۵-۴ را در نظر بگیرید همان طور که مشخص است بعضی از مولفه‌های متصلی که مربوط به خط اول بوده‌اند به دلیل اینکه در شرط اول صدق نکرده‌اند جز مولفه‌های متصل اصلی قرار نگرفته‌اند. این مولفه‌های متصل با رنگ قرمز مشخص شده‌اند. در مرحله‌ی آخر به جایگذاری آنها در مکان مربوط به خودشان پرداخته می‌شود. به طور مثال در شکل ۵-۶ مولفه‌های متصل "هو"، "سی" و "سر" به ترتیب در لیست اصلی قرار دارند. مولفه‌ی متصل نقطه با توجه به رابطه‌ی (۵-۲) نزدیک‌ترین به بدنه‌ی خود یعنی مولفه‌ی متصل "سی" است اما به دلیل اینکه مختصات X مرکز آن بیشتر از مختصات X مرکز "سی" است در

لیست اصلی قرار نگرفته است.

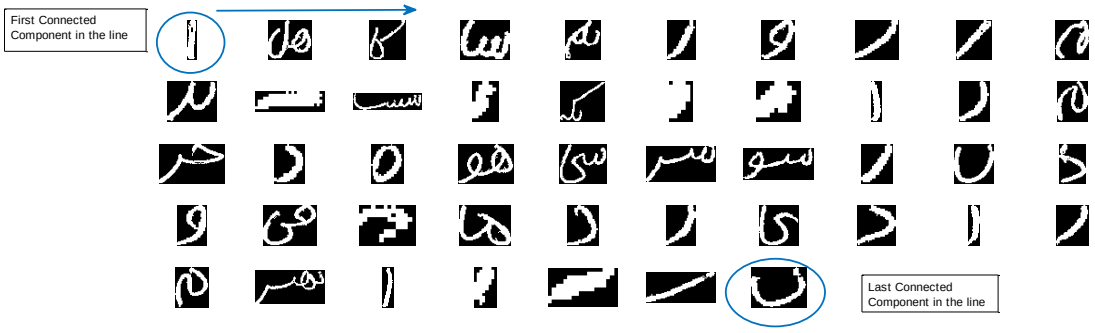
با انجام پس‌پردازش‌هایی بر روی این لیست اصلی، لیست نهایی مولفه‌های متصل خط اول بدست می‌آید. با حذف مولفه‌های متصل نهایی خط اول از لیست کل مولفه‌های متصل موجود در متن و تکرار الگوریتم از ابتدا، این بار اولین مولفه متصل مربوط به خط دوم پیدا شده و این خط مورد پردازش قرار می‌گیرد و مولفه‌های متصل مربوط به آن ذخیره می‌شوند. در واقع این الگوریتم به تعداد خطوط متن انجام می‌شود تا هر بار یک خط با حذف خط‌های قبل از خودش مورد پردازش قرار گیرد.



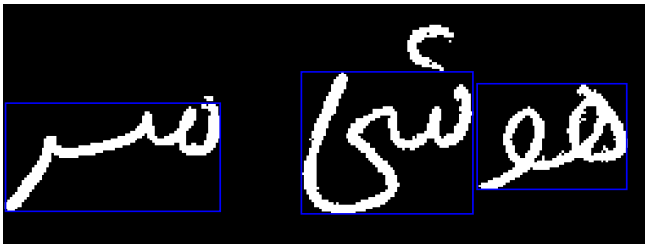
شکل ۵-۳:استخراج تمام مولفه‌های متصل داخل متن



شکل ۵-۴:ذخیره مولفه‌های متصل اصلی خط اول(رنگ آبی)



شکل ۵-۵:نمایش مولفه‌های متصل خط اول به ترتیب شناسایی آنها



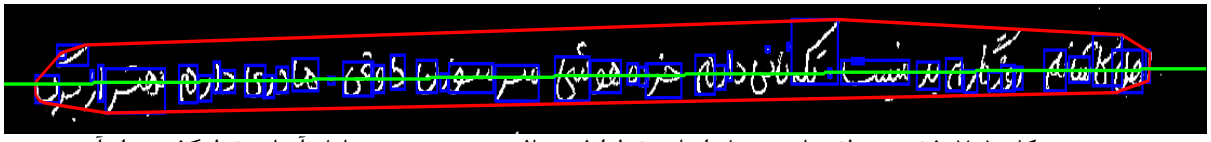
شکل ۵-۶: قرار نگرفتن نقطه‌ی هوشی در لیست مولفه‌های متصل خط اول به دلیل صدق نکردن در شرط یک

۵-۳ پس‌پردازش روی مولفه‌های متصل اصلی هر خط

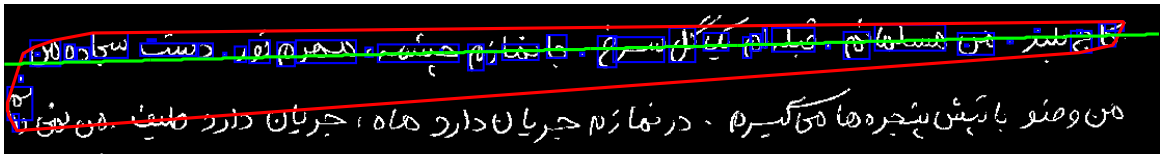
قبل از مرحله‌ی جایگذاری پس‌پردازشی انجام شده تا اگر مولفه متصلی اشتباها در مرحله قبل به یکی از خطوط اختصاص داده شده از لیست حذف گردد. این پس‌پردازش درصد خطا را کاهش می‌دهد. در واقع این پس‌پردازش بعد از بدست آمدن لیست مولفه‌های متصل اصلی هر خط انجام می‌شود و بعد از انجام آن، خط حذف می‌شود.

عملکرد این روش به این صورت است که ابتدا مختصات مکانی (X,Y) پیکسل‌های همه‌ی مولفه‌های متصل درون لیست اصلی، مربوط به هر خط، بصورت یک مجموعه در نظر گرفته شده و بدنه محدب شامل آنها رسم می‌شود. سپس خط گذرنده از این بدنه محدب محاسبه شده که تقریبا محدوده‌ی خط پایه را نشان می‌دهد. در مرحله‌ی بعدی برای هر مولفه‌ی متصل اصلی خط، مختصات مکانی پایین‌ترین پیکسل(گوشه‌ی پایین سمت چپ هر مولفه‌ی متصل) در نظر گرفته شده و فاصله‌ی

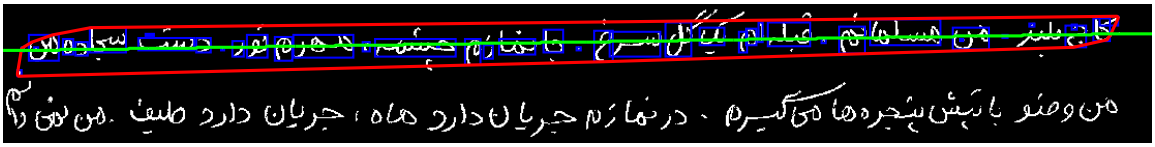
آن از خط پایه در جهت راستای Y محاسبه می‌شود. اگر این فاصله از یک حد آستانه بیشتر بود این مولفه‌ی متصل مربوط به این خط نبوده و از لیست حذف می‌شود. (این آستانه برای تمام تصاویر ۹۰ در نظر گرفته شده است.) به این صورت لیست نهایی مولفه‌های متصل مربوط به هر خط بدست می‌آید. با بدست آوردن این لیست نهایی با همان روش قبلی، دوباره معادله خط نهایی به دست آمده و ذخیره می‌شود. مراحل این قسمت در شکل ۵-۷ تا شکل ۵-۹ نمایش داده شده است.



شکل ۵-۷: ذخیره مولفه‌های متصل اصلی خط اول و یافتن بدنه محدب شامل آنها و خط گذرنده از آن



شکل ۵-۸: دو مولفه متصلی که مربوط به خط چهارم هستند و به اشتباه به خط سوم اختصاص یافته‌اند.



شکل ۵-۹: تصحیح خطای بالا با در نظر گرفتن فاصله‌ی مولفه‌های متصل با معادله خط سوم

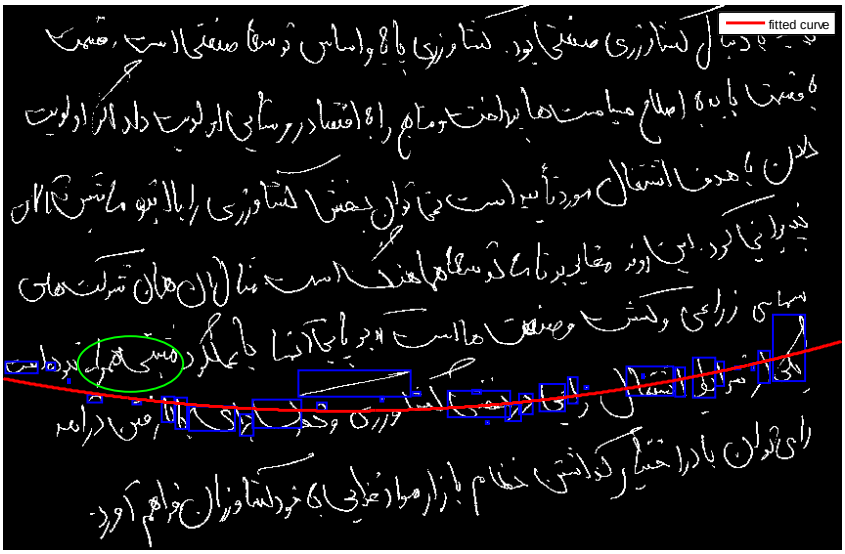
به طور کلی برای محاسبه‌ی خط، ابتدا از یک تابع درجه دوم بصورت ($aX^2 + bX + c$) استفاده شده است. در مورد خطوط کج متن ، این خط به صورت یک سهمی است. در بعضی از موارد اگر شیب این سهمی از حدی بیشتر باشد در مرحله‌ی پس‌پردازش خطا ایجاد می‌کند. این خطا در دو مورد بوجود می‌آید.

به طور مثال در شکل ۵-۱۰مولفه‌های متصل نقطه، "ه" و "ست" در آخر خط پنجم جزء مولفه-

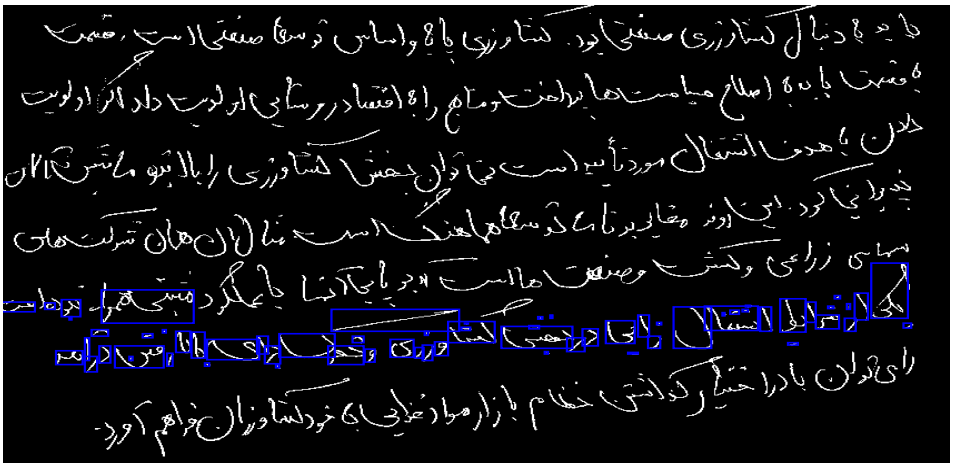
های متصل اصلی این خط قرار نگرفته‌اند و در پردازش خط بعدی جزء مولفه‌های متصل خط ششم قرار می‌گیرند. در مرحله‌ی پس‌پردازش به دلیل اینکه خط ششم به صورت یک منحنی است باز هم فاصله‌ی این مولفه‌های متصل با این خط از مقدار آستانه‌ی در نظر گرفته شده کمتر است و قادر به تصحیح این اشتباه نیست. (شکل ۵-۱۱) ولی اگر همین خط به صورت تابع درجه اول aX+b تخمین زده شود این خطا تصحیح می‌شود.(شکل ۵-۱۲ تا شکل ۵-۱۴).

در حالت بعدی مولفه‌ی متصلی که در شکل ۵-۱۰ با رنگ سبز مشخص است جزء مولفه‌ی متصل جا افتاده از خط پنجم است. در مرحله‌ی جایگذاری به دلیل اینکه این مولفه‌ی متصل به منحنی خط ششم نزدیکتر از خط پنجم می‌باشد اشتباها به این خط اختصاص پیدا می‌کند.(شکل ۵-۱۱) ولی با تخمین خط ششم به صورت تابع درجه اول این خطا تصحیح می‌شود.(شکل ۵-۱۳)

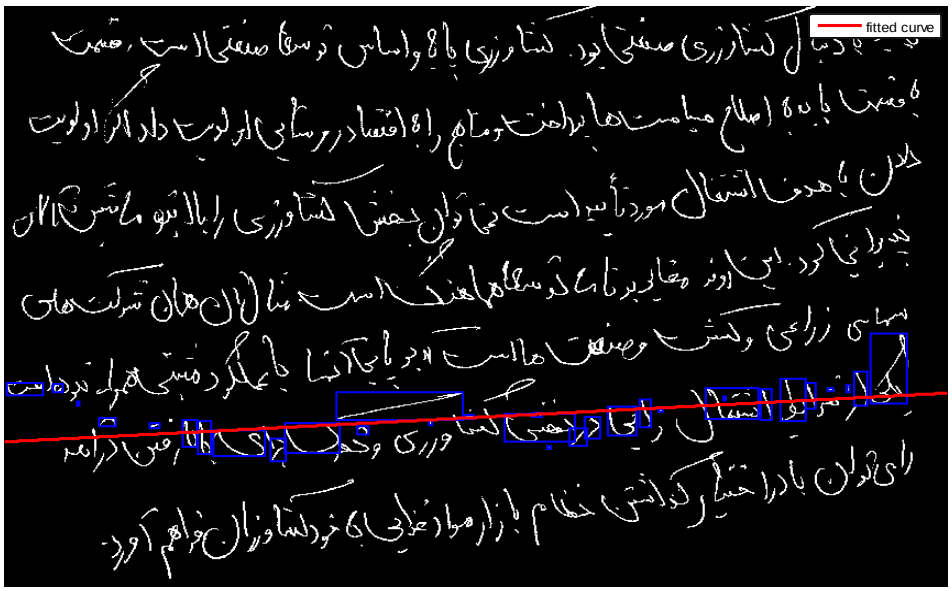
با توجه به توضیحات بالا به طور کلی اگر ضریب X^2 (یعنی a) از $۰/۵\times۱۰^{(-۵)}$ بیشتر بود خط با تابع درجه اول تخمین زده می‌شود و در غیر این صورت بصورت تابع درجه دوم است. اعمال این شرط باعث کاهش خطا و بهبود نتایج می‌شود.



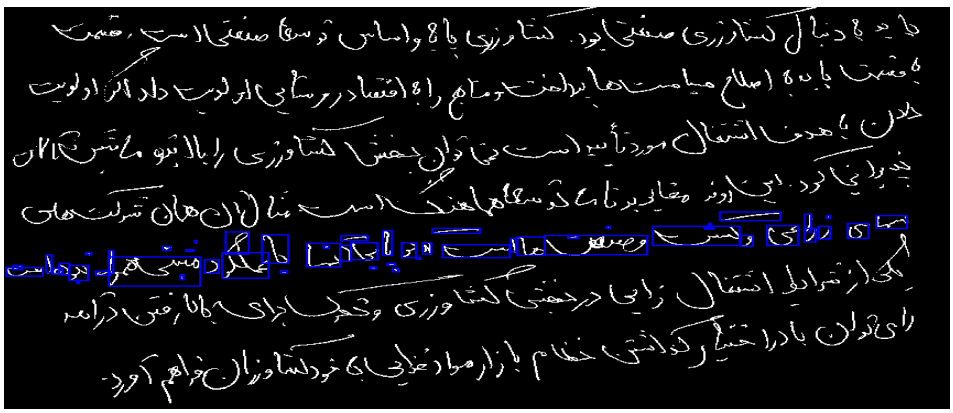
شکل ۵-۱۰: در نظر گرفتن خط ششم بصورت تابع درجه دو (خط قرمز)



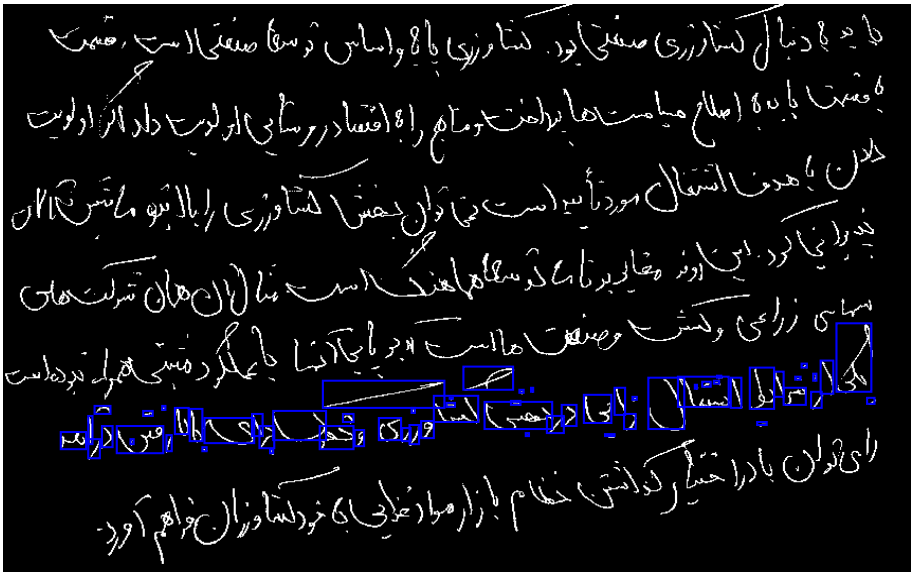
شکل ۵-۱۱:ایجاد خطا در جداسازی خط ششم



شکل ۵-۱۲:در نظر گرفتن خط ششم بصورت تابع درجه اول(خط قرمز)



شکل ۵-۱۳:تصحیح خطا در مرحله جداسازی خط پنجم



شکل ۵-۱۴:تصحیح خطا در مرحله جداسازی خط ششم

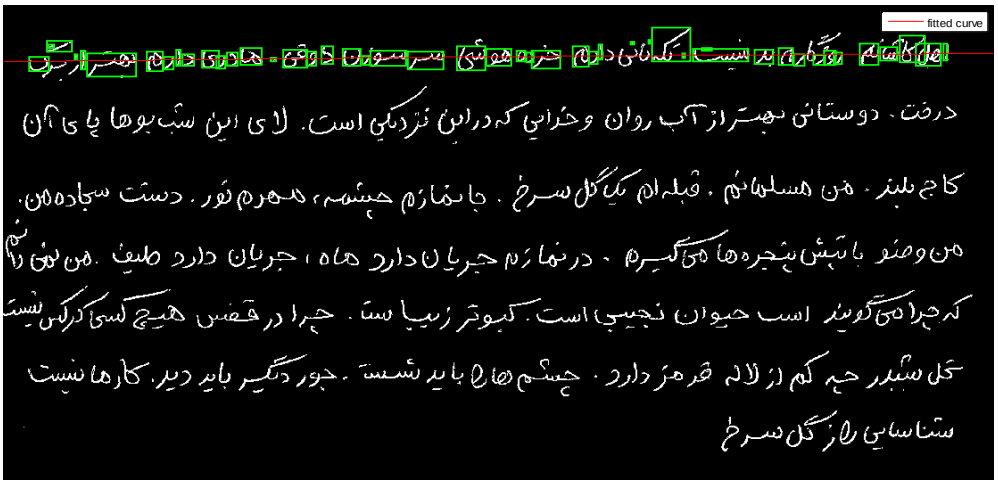
بعد از اتمام این مرحله ابتدا مشخصات مولفه‌های متصل لیست نهایی برای هر خط به صورت جداگانه ذخیره و سپس از داخل لیست اولیه که شامل کل مولفه‌های متصل متن بود حذف می‌گردنند.

همان‌طور که در شکل ۵-۴ به آن اشاره شد مولفه‌های متصلی وجود دارند که به دلیل صدق نکردن در شرط ۱ جز مولفه‌های نهایی خط اول قرار نگرفتند حال اگر این مولفه‌های متصل نزدیک به ابتدای خط باشند در مرحله تکرار الگوریتم دوباره به عنوان یک خط جدا در نظر گرفته شده و مورد پردازش قرار می‌گیرند. برای اجتناب از این حالت بعد از بدست آوردن مولفه‌های متصل هر خط و قبل از پردازش خط بعدی مرحله‌ی زیر انجام می شود:

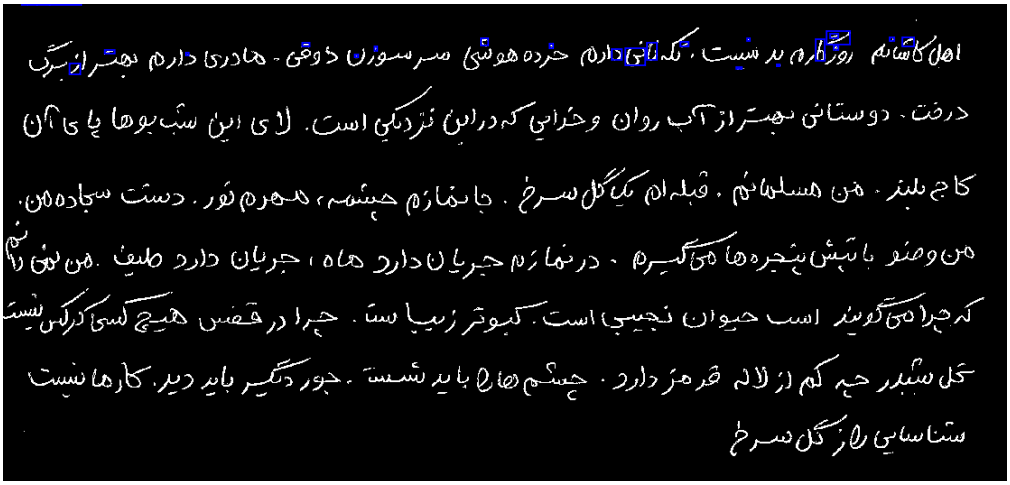
مختصات y پایین‌ترین پیکسل همه‌ی مولفه‌های متصل نهایی خط را در نظر گرفته و بیشترین آنها بدست می‌آید. سپس مابقی مولفه‌های متصل داخل متن از نظر این مختصات با این بیشترین مقایسه می‌شوند و آنهایی که کمتر از این بیشترین هستند از لیست کل مولفه‌های متصل متن حذف شده و در لیست جدیدی ذخیره می‌شوند. در واقع این لیست جدید بیشتر شامل مولفه‌های متصلی است که از خط مربوط به خودشان جا مانده‌اند و بعدا جایگذاری می‌شوند. شکل ۵-۱۶ این مولفه‌های

متصل که مربوط به خط اول هستند را نشان می‌دهد. در پایان این الگوریتم برای هر خط لیست مربوط به مولفه‌های متصل نهایی هر خط، معادله هر خط و لیست مولفه‌های متصل جا افتاده از خطوط بدست می‌آید.

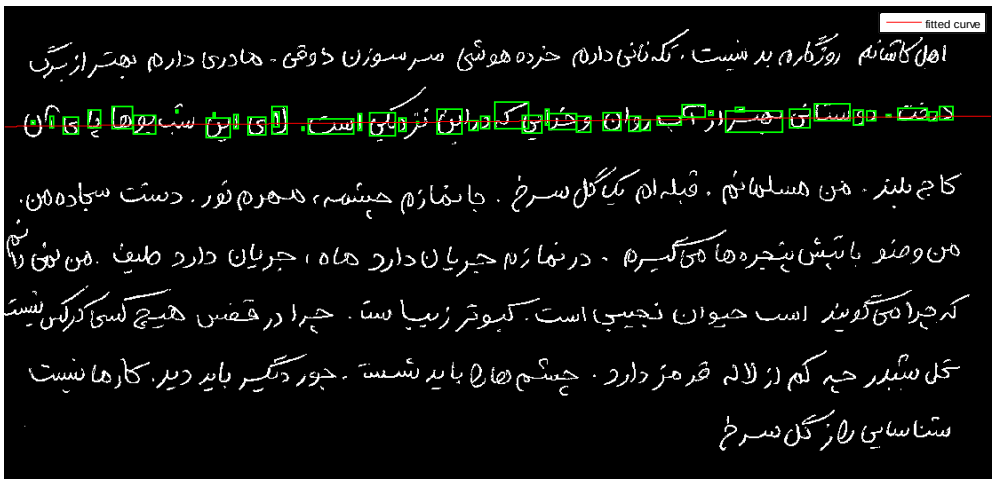
بعد از پردازش تمام خطوط متن، اگر مولفه‌ی متصلی در لیست کل مولفه‌های متصل موجود در متن باقی مانده و به خطی اختصاص پیدا نکرده بود، در مرحله‌ی بعدی در جای مناسب جایگذاری می‌شود. شکل ۵-۱۵ تا شکل ۵-۲۲ این الگوریتم را برای چهار خط اول به ترتیب اجرای آن نشان می‌دهد. در شکل ۵-۲۳ تصویری هم سایز با تصویر اصلی که در آن معادلات خطوط رسم شده‌اند نمایش داده شده است.



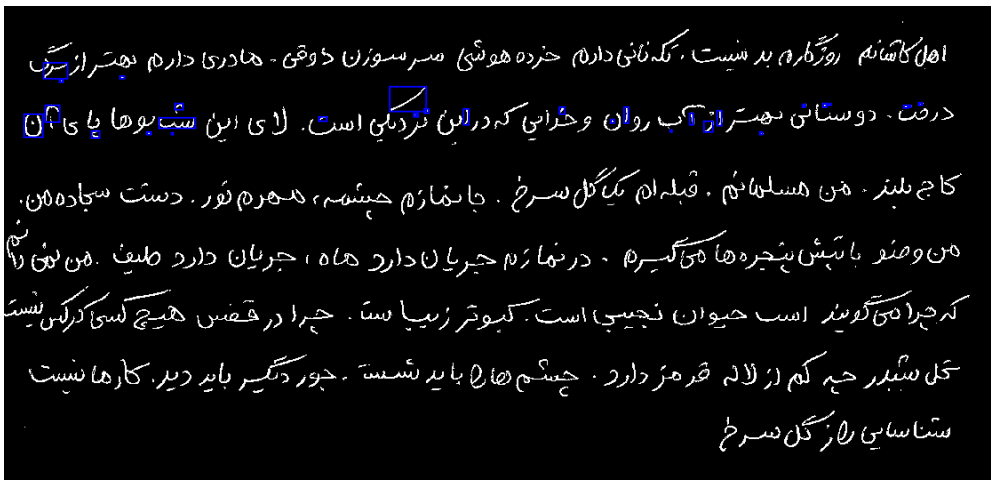
شکل ۵-۱۵:مولفه‌های متصل نهایی خط اول بعد از پس‌پردازش



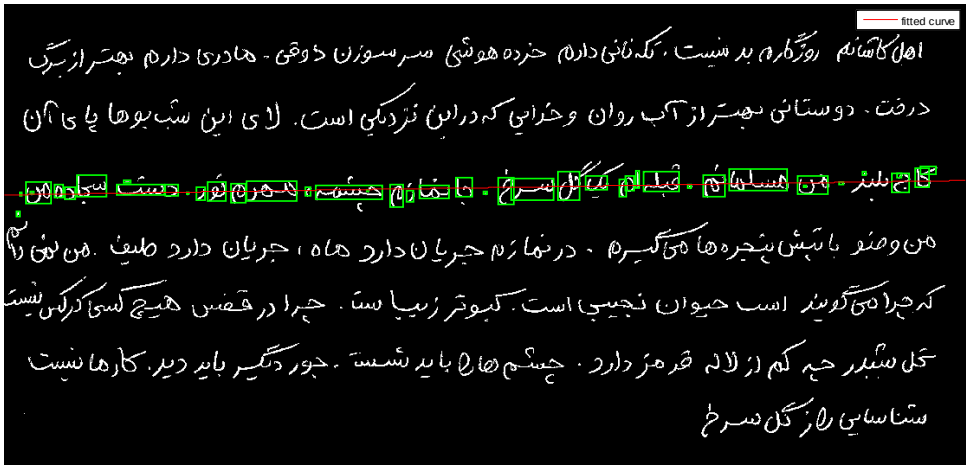
شکل ۵-۱۶:حذف بعضی از مولفه‌های متصل از کل متن و ذخیره جداگانه‌ی آنها(مرحله‌ی پس‌پردازش)



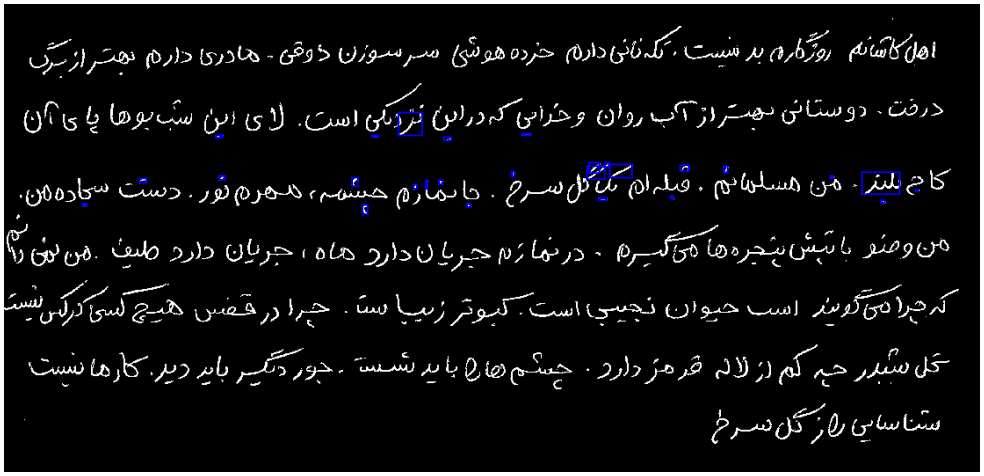
شکل ۵-۱۷: مولفه‌های متصل نهایی خط دوم بعد از پس‌پردازش



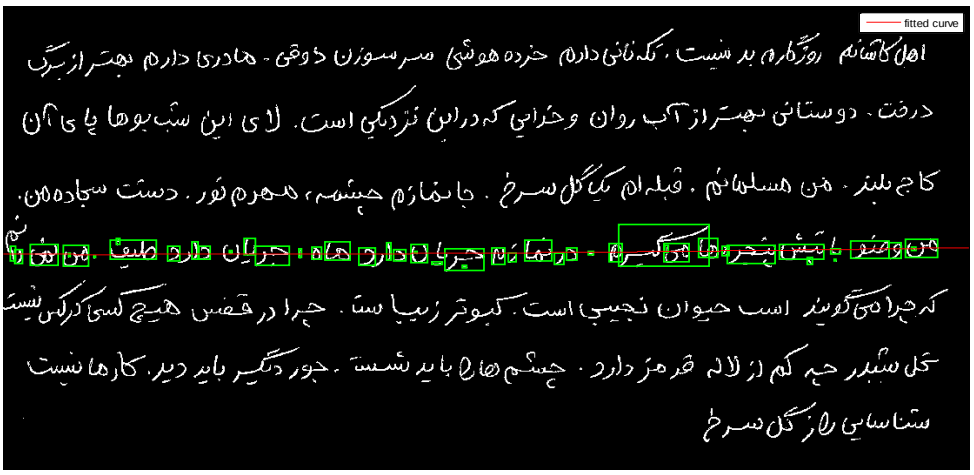
شکل ۵-۱۸:حذف بعضی از مولفه‌های متصل از کل متن و ذخیره جداگانه‌ی آنها(مرحله‌ی پس‌پردازش)



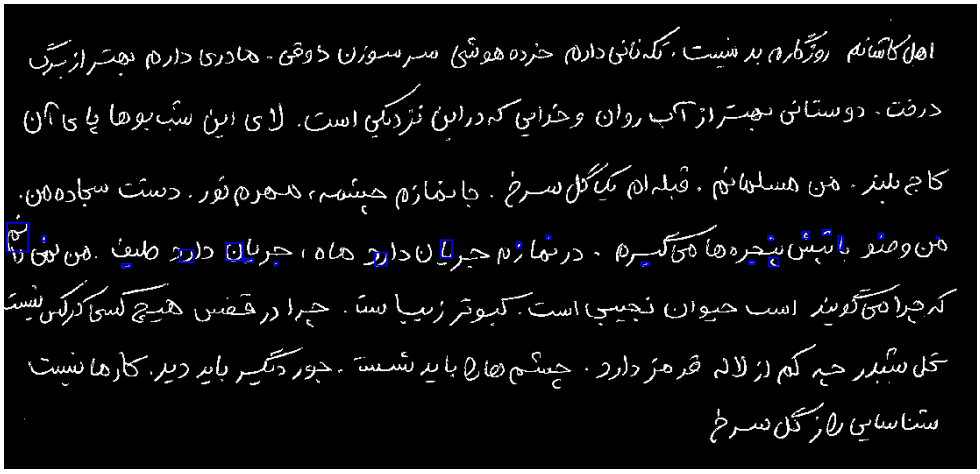
شکل ۵-۱۹:مولفه‌های متصل نهایی خط سوم بعد از پس پردازش



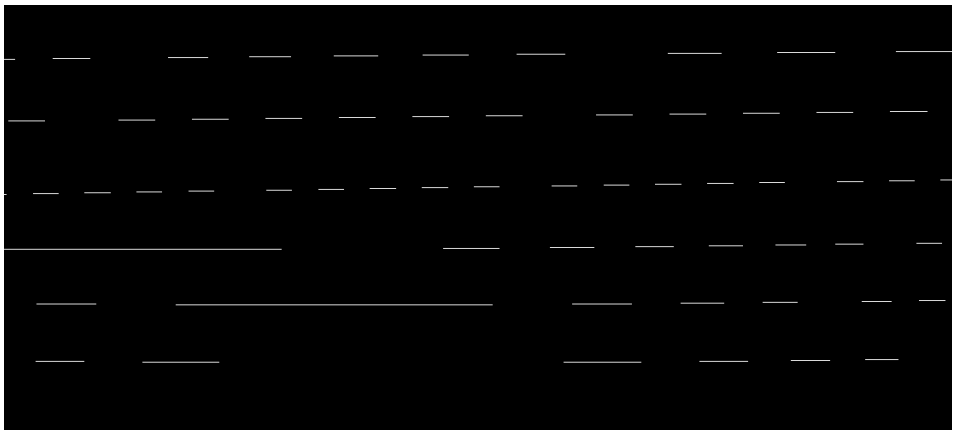
شکل ۵-۲۰:حذف بعضی از مولفه‌های متصل از کل متن و ذخیره جداگانه‌ی آنها(مرحله‌ی پس‌پردازش)



شکل ۵-۲۱:مولفه‌های متصل نهایی خط چهارم بعد از پس‌پردازش



شکل ۵-۲۲:حذف بعضی از مولفه‌های متصل از کل متن و ذخیره جداگانه‌ی آنها(مرحله‌ی پس‌پردازش)



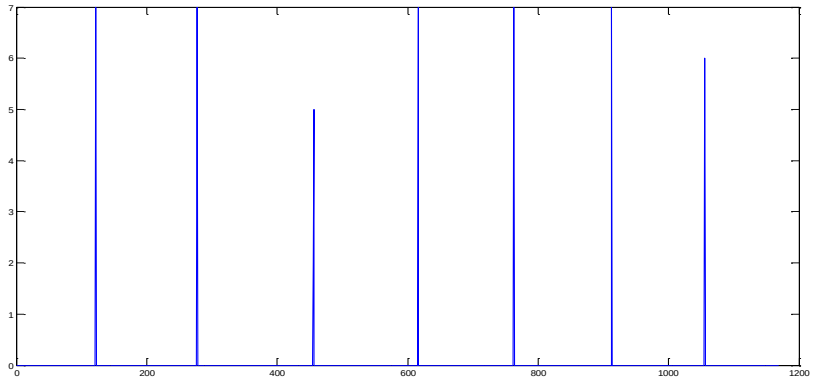
شکل ۵-۲۳: نمایش معادلات خطوط

۵-۴ جایگذاری

در مرحله‌ی جایگذاری در مورد تمام مولفه‌های متصلی که به هیچ خطی اختصاص نیافته‌اند تصمیم‌گیری می‌شود. بیشتر این مولفه‌های متصل را سرکش‌ها و نقاط تشکیل داده‌اند. ابتدا همه‌ی این مولفه‌های متصل در یک لیست ذخیره می‌شوند. سپس مختصات x و y پایین‌ترین پیکسل هر مولفه‌های متصل در نظر گرفته شده و همه آنها با توجه به مختصات x از سمت راست به چپ تصویر (به صورت نزولی) مرتب شده و به همین ترتیب جایگذاری می‌شوند. برای هر مولفه متصل، مختصات x را در نظر گرفته و یک نوار ستونی روی تصویر شکل ۵-۲۳ با مرکزیت این مولفه متصل و با در نظر گرفتن سه پیکسل در سمت راست و سه پیکسل در سمت چپ آن تشکیل داده و سپس تصویر در

راستای افقی جاروب می‌گردد. همان‌طور که در شکل ۵-۲۴ مشخص است به تعداد خطوط متن قله تشکیل شده است. این قله‌ها به ترتیب از سمت چپ به راست مربوط به خط اول تا هفتم می‌باشند.

مختصات هر کدام از این قله‌ها با مختصات y مولفه‌ی متصل مقایسه شده و به هر کدام که نزدیکتر بود این مولفه‌ی متصل مربوط به همان خط است. مزیت این روش در این است که چون بصورت محلی برای هر مولفه‌ی متصل اجرا می‌شود سرعت محاسبات بالاتری دارد. این مرحله برای تمام مولفه‌های متصل داخل لیست اجرا شده تا مشخص شود هر مولفه‌ی متصل جا افتاده مربوط به کدام خط است.



شکل ۵-۲۴:نمودار افقی افکنش معادلات خطوط

بعد از مشخص شدن خط مربوط به هر مولفه‌ی متصل، با در نظر گرفتن فاصله‌ی هر مولفه‌ی متصل با خط مربوط به خودش (در راستای y)، این مولفه متصل در گروه بالارونده، پایین‌رونده یا روی خط قرار می‌گیرد:

۱) اگر این فاصله بیشتر از ۱۸ بود این مولفه‌ی متصل جزء گروه بالارونده است. سرکش‌ها و نقاط بالایی کلمات در این گروه قرار دارند.

۲) اگر این فاصله کمتر از ۱۵- بود این مولفه‌ی متصل جزء گروه پایین رونده است. نقاط پایینی کلمات در این گروه قرار دارند.

۳) در غیر دو شرط بالا جزء حروف روی خط است. معمولا مولفه‌های متصل غیر از نقاط و سرکش در این گروه قرار دارند.

سپس از خط اول شروع کرده و برای مولفه‌های متصل مختص به این خط، مکان مناسب برای جایگذاری با روش زیر مشخص می‌شود. این روش به ترتیب برای تمام خطوط اجرا می‌شود.

همان طور که قبلا بیان شد مولفه‌های متصل نهایی هر خط که از مرحله‌ی قبل بدست آمده‌اند با در نظر گرفتن ویژگی زبان فارسی، در متن از راست به چپ شناسایی شده‌اند. به عبارت دیگر از نظر مولفه‌ی x به صورت نزولی مرتب شده‌اند. به همین دلیل در این مرحله برای مشخص کردن محدوده‌ی مناسب برای جایگذاری مولفه‌ی متصل، ابتدا مختصات x مرکز آن با مختصات x مراکز مولفه‌های متصل نهایی خط مربوطه مقایسه شده و به هر کدام از مولفه‌های متصل اصلی خط که نزدیکتر بود محدوده‌ی مناسب برای جایگذاری است. این محدوده‌ی به دست آمده به سه صورت زیر است:

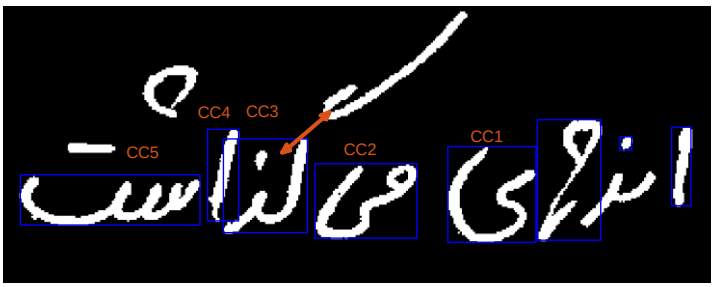
۱) مختصات x مرکز مولفه‌ی متصل از مختصات x تمام مراکز مولفه‌های متصل خط اصلی بزرگتر است که در این صورت در ابتدای هر خط جایگذاری می‌شود.
۲) مختصات x مرکز مولفه‌ی متصل از مختصات x تمام مراکز مولفه‌های متصل خط اصلی کوچکتر است که در این صورت در انتهای هر خط جایگذاری می‌شود.
۳) در غیر دو مورد بالا با توجه به اینکه مولفه‌ی متصل در گروه بالارونده، پایین‌رونده یا روی خط است، در بین مولفه‌های متصل خط اصلی جایگذاری می‌شود.

برای قسمت سوم با توجه به ساختار زبان فارسی سه قاعده‌ی کلی در نظر گرفته شده است.

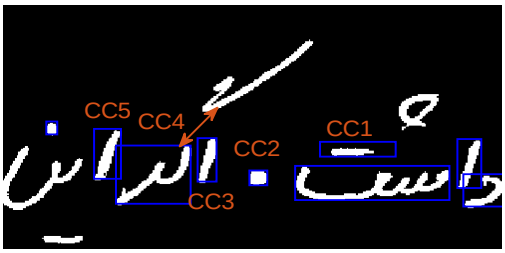
۵-۴-۱روش‌های جایگذاری مولفه‌ی متصل بالارونده

در این حالت با توجه به محدوده‌ی بدست آمده در بالا، پنج مولفه‌ی متصل خط اصلی که نزدیک‌ترین مرکز را به مرکز مولفه‌ی متصل دارد در نظر گرفته می‌شود. اکثر مولفه‌های متصلی که در

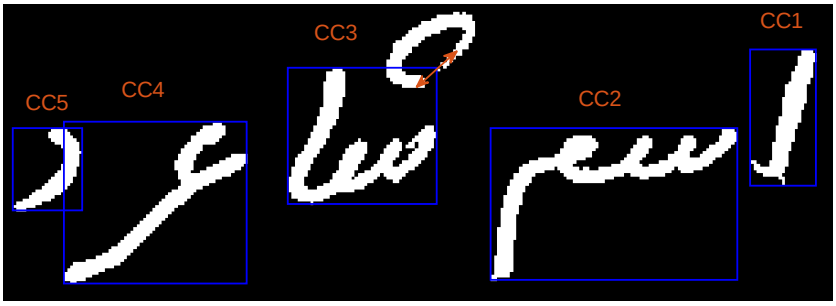
گروه بالارونده قرار دارند نقطه یا سرکش هستند. برای آنها از این قاعده استفاده شده که سمت چپ‌ترین پیکسل آنها به کدام سمت راست‌ترین پیکسل، پنج مولفه‌ی متصل خط اصلی نزدیک است. با یافتن این مولفه‌ی متصل اصلی خط، مولفه متصل بالارونده قبل از آن قرار می‌گیرد. این روش جایگذاری در شکل ۵-۲۵ و شکل ۵-۲۷ نمایش داده شده است. در این شکل‌ها پنج مولفه‌ی متصل خط اصلی با رنگ قرمز مشخص شده‌اند. مولفه‌ی متصل بالارونده و مولفه‌ی متصل خط اصلی که پیکسل سمت راست آن به پیکسل سمت چپ مولفه‌ی بالارونده از همه نزدیکتر است با فلش مشخص شده‌اند. مولفه‌ی بالارونده قبل از این مولفه‌ی متصل قرار می‌گیرد.



شکل ۵-۲۵:جایگذاری سرکش"گ" قبل از بدنه‌ی اصلی" گذا"



شکل ۵-۲۶:جایگذاری سرکش"گ" قبل از بدنه‌ی اصلی" گر"



شکل ۵-۲۷:جایگذاری نقطه قبل از بدنه‌ی اصلی"سا"

۵-۴-۲ روش‌های جایگذاری مولفه‌ی متصل پایین‌رونده

اکثر مولفه متصلی که در گروه پایین‌رونده هستند نقاط می‌باشند. با توجه به این اصل که هر نقطه با مولفه متصلی که بدنه‌ی اصلی نقطه مورد نظر است بیشترین اشتراک از نظر طولی (مختصات x) را دارد قبل از آن جای داده می‌شود. در این حالت دو مولفه‌ی متصل اصلی خط که بیشترین احتمال را دارد که مولفه‌ی متصل پایین‌رونده مربوط به آنها باشد توسط رابطه‌ی (۵-۳) مشخص می‌شوند:

$$C(X)_{B-1} < C(X)_I < C(X)_B \qquad (۵-۳)$$

در رابطه‌ی (۵-۳)، $C(X)_{B-1}$ و $C(X)_B$ مختصات x مرکز مولفه‌ی متصل خط اصلی هستند و $C(X)_I$ مختصات x مرکز مولفه‌ی متصل پایین‌رونده است. با توجه به اینکه حرکت از سمت راست به چپ خط است در لیست مولفه‌های متصل خط اصلی ابتدا مولفه‌ی متصل با اندیس B و بعد از آن B-1 قرار دارد.

نحوه‌ی بدست آوردن بازه‌ی اشتراک بین مولفه‌ی متصل اصلی B-1 و B با مولفه‌ی متصل پایین‌رونده‌ی I در زیر توضیح داده شده است.

- L(X) = مختصات x سمت چپ‌ترین پیکسل هر مولفه‌ی متصل
- R(X) = مختصات x سمت راست‌ترین پیکسل هر مولفه‌ی متصل

- Start = مختصات شروع بازه
- End = مختصات پایان بازه

- اگر $R(X)_{B-1} < L(X)_I$ برقرار باشد دو مولفه‌ی متصل هیچ اشتراکی با هم ندارند.
- اگر $R(X)_I < L(X)_B$ برقرار باشد دو مولفه‌ی متصل هیچ اشتراکی با هم ندارند.

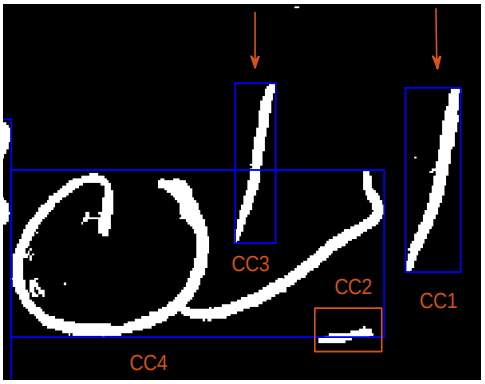
در صورتی که شرط بالا برقرار نباشد بازه‌ی اشتراکی به صورت زیر است:

- اگر $L(X)_{B-1} < L(X)_I$ در نتیجه $Start = L(X)_I$ در غیر این صورت $Start = L(X)_{B-1}$.
- اگر $R(X)_I < R(X)_{B-1}$ در نتیجه $End = R(X)_I$ در غیر این صورت $End = R(X)_{B-1}$.

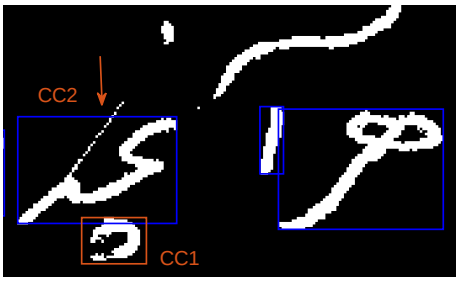
برای مولفه‌ی متصل B نیز با همین روش بازه‌ی اشتراکی بدست می‌آید. با بدست آوردن ابتدا و انتهای بازه، میزان اشتراک از رابطه‌ی (۵-۴) بدست می‌آید:

$$U \qquad (۴-۵)$$

در صورتی که بازه‌ی اشتراکی برای هر دو مولفه‌ی متصل اصلی با مولفه‌ی متصل پایین‌رونده برابر بود، مولفه‌ی متصل پایین‌رونده بین آنها جایگذاری می‌شود. این روش جایگذاری در شکل ۵-۲۸ و شکل ۵-۲۹ نمایش داده شده است. در این شکل‌ها مولفه‌ی متصل پایین‌رونده با مستطیل قرمز و دو مولفه‌ی متصل خط اصلی با فلش مشخص هستند.



شکل ۵-۲۸: نقطه با دو مولفه‌ی متصل "الف" اشتراک نداشته و بین آنها جایگذاری می‌شود.



شکل ۵-۲۹: جایگذاری نقطه قبل از "نچه"

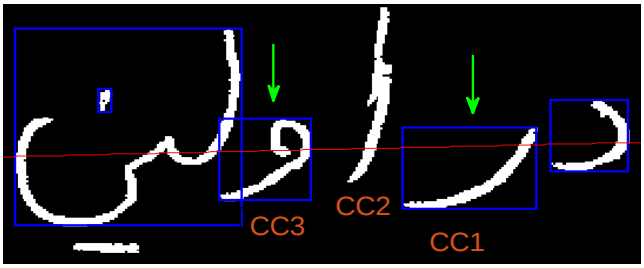
در شکل ۵-۲۹، نقطه با مولفه‌ی متصل بدنه خود، "نچه" بیشترین اشتراک را داشته و قبل آن جایگذاری می‌شود.

۵-۴-۳ روش‌های جایگذاری مولفه‌ی متصل روی خط

در این حالت با توجه به رابطه‌ی (۵-۳) دو مولفه‌ی متصل اصلی که بیشترین احتمال را دارد که مولفه‌ی متصل مربوط به آنها باشد مشخص می‌شوند. این مولفه متصل یا باید بین، قبل یا بعد از دو مولفه‌ی متصل اصلی جایگذاری شود. منظور از مولفه‌های متصل روی خط در واقع مولفه‌های متصلی است که بعضی از پیکسل‌های آنها روی خط هستند یا فاصله‌ی بسیار کمی با خط دارند. در این قسمت از شش قاعده که مربوط به ساختار زبان فارسی است، استفاده شده است. این قواعد به ترتیب در شکل ۵-۳۰ تا شکل ۵-۳۴ مشخص هستند. بعد از انجام مرحله‌ی جایگذاری، ترتیب قرارگرفتن مولفه‌های متصل با اعداد ۱ تا ۳ مشخص شده است.

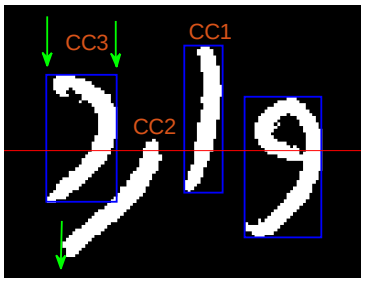
حالت اول: اگر مولفه‌ی متصل با دو مولفه‌ی متصل اصلی B-1 (CC3) و B(CC1) اشتراکی

نداشته باشد بین آنها جا داده می‌شود. دو مولفه‌ی متصل اصلی با فلش سبز رنگ مشخص هستند. (شکل ۵-۳۰)



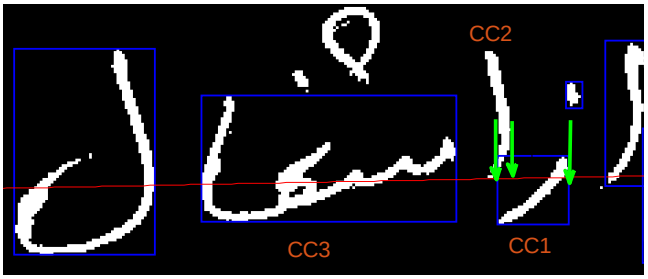
شکل ۵-۳۰:مولفه‌ی متصل"الف" بین دو مولفه‌ی متصل "ر" و "و" جایگذاری می‌شود.

حالت دوم: انتهای مولفه‌ی متصل بین ابتدا و انتهای مولفه‌ی متصل B-1 قرار دارد(این حالت با فلش سبز رنگ مشخص شده است) و با مولفه‌ی متصل B اشتراک ندارد؛ در این حالت بین آنها جا داده می‌شود. (شکل ۵-۳۱)



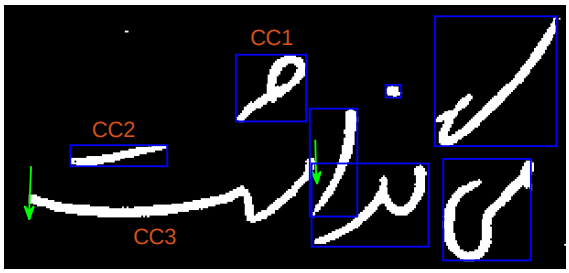
شکل ۵-۳۱:مولفه‌ی متصل"ر" بین "الف" و "د" جایگذاری می‌شود.

حالت سوم: برعکس حالت بالا ابتدای مولفه‌ی متصل بین ابتدا و انتهای مولفه‌ی متصل B قرار دارد(این حالت با فلش سبز رنگ مشخص شده است) و با مولفه‌ی متصل B-1 اشتراک ندارد؛ در این حالت بین آنها جا داده می‌شود. (شکل ۵-۳۲)



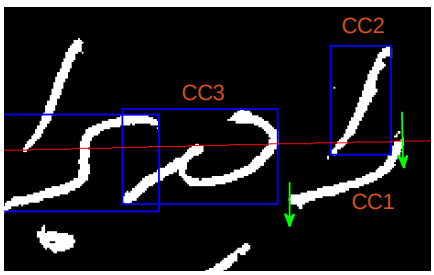
شکل ۵-۳۲:مولفه‌ی متصل"الف" بین "ر" و "سعا" جایگذاری می‌شود.

حالت چهارم: مولفه‌ی متصل جا افتاده جزء بدنه‌ی اصلی کلمه مورد نظر است. در این حالت بعد از دو مولفه‌ی متصل اصلی که معمولا نقاط هستند جایگذاری می‌شود.(شکل ۵-۳۳)



شکل ۵-۳۳:بدنه‌ی اصلی "شت" بعد از نقاط آن جایگذاری می‌شود.

حالت پنجم: در این حالت طول مولفه‌ی متصل از طول مولفه‌ی متصل B (CC2) بزرگتر است به طوریکه آن را دربرمی‌گیرد و با مولفه‌ی متصل B-1 (CC3) اشتراکی ندارد. مولفه‌ی متصل قبل از هر دو مولفه‌ی اصلی جایگذاری می‌شود.(شکل ۵-۳۴)



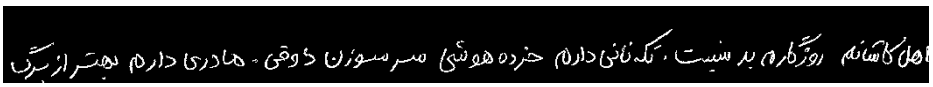
شکل ۵-۳۴:مولفه‌ی متصل "ر" قبل از "الف" و "در" قرار می‌گیرد.

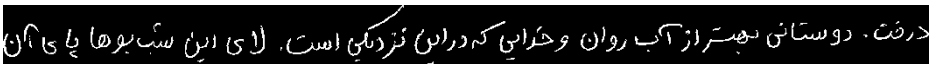
حالت ششم: در غیر از موارد بالا، مولفه‌ی متصل بین مولفه‌ی متصل B و B-1 قرار می‌گیرد.

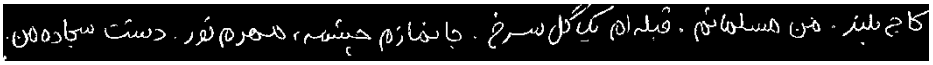
۵-۵ جداسازی خطوط

بعد از انجام مرحله‌ی جایگذاری برای تمام خطوط، هر مولفه متصل موجود در متن به یکی از خطوط اختصاص پیدا کرده است. سپس می‌توان هر خط را با مولفه‌های متصلش به طور مستقل از مابقی خطوط نمایش داد. نتیجه این مرحله جداسازی خطوط از یکدیگر می‌باشد. شکل ۵-۳۵ تمام مولفه‌های متصل خط اول را بعد از مرحله جایگذاری نشان می‌دهد. در این شکل تمام مولفه‌های متصل جا مانده از خط اول جایگذاری شده و کل مولفه‌های متصل این خط به ترتیبی که در متن از سمت راست به چپ قرار گرفته‌اند ذخیره می‌شوند. شکل ۵-۳۶ تا شکل ۵-۳۷ نتیجه اجرای روش و جداسازی خطوط متن از یگدیگر را نشان می‌دهد.

شکل ۵-۳۵:جایگذاری مولفه‌های متصل جا مانده از خط اول







شکل ۵-۳۶:نتیجه جداسازی خطوط متن

شکل ۵-۳۷: نتیجه جداسازی خطوط متن

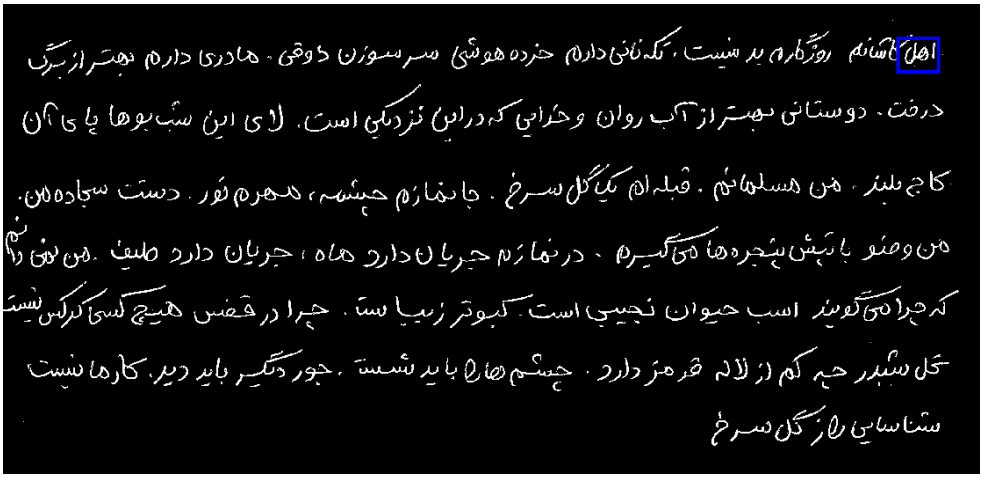
۵-۶ استخراج برون خط تصویر کلمات متن

در این مرحله با در نظر گرفتن این قانون که اکثر حروف اضافه و کلمات فارسی از ترکیب یک تا هفت مولفه‌ی متصل تشکیل شده‌اند استخراج کلمات از داخل متن انجام می‌شود. مولفه‌های متصل هر خط از نظر مکانی به ترتیب از سمت راست متن به سمت چپ ذخیره شده‌اند. از اولین مولفه‌ی متصل هر خط شروع کرده و به ترتیب برای ساختن ترکیب‌های دوتایی، دو تا مولفه‌ی متصل اول انتخاب شده‌اند. سپس تصویری با زمینه‌ی سیاه هم ابعاد با تصویر اصلی در نظر گرفته شده است. داخل این تصویر در مکان پیکسل‌های مربوط به دو مولفه‌ی متصل در متن اصلی، یک قرار داده می‌شود. در واقع در این تصویر مولفه‌های متصل سفید هستند.

در مرحله‌ی بعدی با توجه به ابعاد و مختصات دو مولفه‌ی متصل، فقط تصویری که شامل این دو مولفه‌ی متصل است از این تصویر استخراج می‌شود. این روش تا آخر برای کل مولفه‌ها متصل این خط با در نظر گرفتن گام یک ادامه پیدا می‌کند. سپس همین روش برای ترکیب سه تایی تا هفت تایی و روی تمام خطوط اجرا می‌شود. در نهایت تصویر تمام این ترکیبات یک تا هفت‌تایی که در متن وجود دارند ذخیره می‌شود. موقعیت مکانی هر تصویر کلمه در داخل متن اصلی در مرحله جستجو برای

مشخص کردن آن لازم می‌باشد. از این رو با ذخیره هر تصویر، موقعیت مکانی آن روی تصویر اصلی نیز ذخیره می‌شود. مزیت روش ارائه شده این است که به دلیل اینکه گام یک در نظر گرفته شده هر مولفه متصلی که آغاز کننده یک کلمه است قطعا در ابتدای یک ترکیب قرار می‌گیرد. سپس اگر در ترتیب قرار گرفتن مولفه‌های متصلی که در ترکیب وجود دارند خطا رخ نداده باشد استخراج آن به درستی انجام می‌شود.

به طور مثال در شکل ۵-۳۸ برای استخراج ترکیب‌های دوتایی، مکان پیکسل‌های دو مولفه‌ی متصل "ا" و "هل" در نظر گرفته شده است و تصویر شکل ۵-۳۹ بدست می‌آید. سپس با توجه به الگوریتم زیر، تصویر شامل کلمه‌ی "اهل" از تصویر شکل ۵-۳۹ استخراج می‌شود.



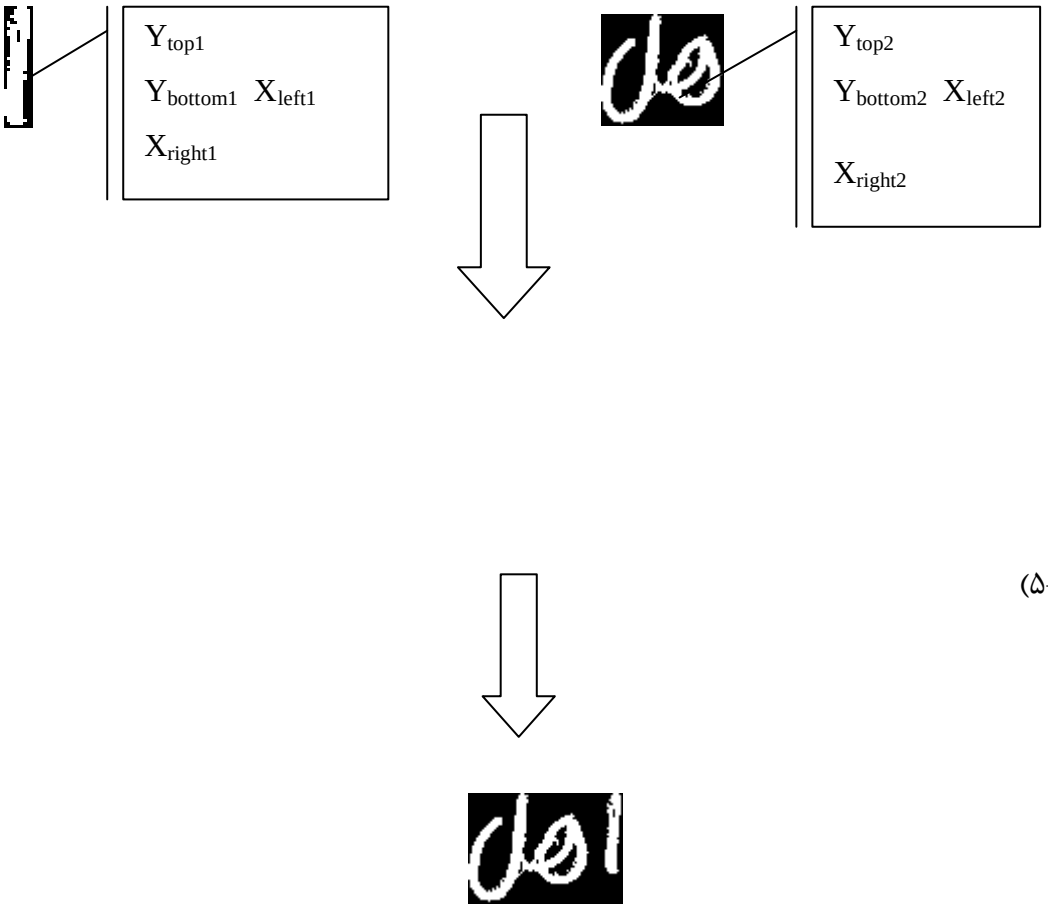
شکل ۵-۳۸:در نظر گرفتن ترکیب دوتایی مولفه‌های متصل

شکل ۵-۳۹:آشکار کردن ترکیب مولفه‌های متصل در تصویری هم اندازه با تصویر اصلی

برای بدست آوردن موقعیت مکانی هر تصویر الگوریتم زیر انجام می‌شود:

برای مولفه‌های متصل داخل ترکیب، چهار مختصات معرفی شده در زیر در نظر گرفته می‌شود. سپس با توجه به رابطه‌ی (۵-۵) مختصات مناسب برای استخراج ترکیب مورد نظر از تصویر شکل ۵-۳۹ بدست می‌آید. همان‌طور که اشاره شد موقعیت مکانی به صورت [$|Y_{top}-Y_{bottom}|$, $|X_{left} - X_{right}|$, Y_{top} ,X_{left}] برای هر تصویر ذخیره می‌شود. این روش در شکل ۵-۴۰ نمایش داده شده است.

- Y_{top} = مختصات Y گوشه‌ی سمت چپ بالا مولفه‌ی متصل
- Y_{bottom}= مختصات Y گوشه‌ی سمت چپ پایین مولفه‌ی متصل
- X_{left}= مختصات X گوشه‌ی سمت چپ بالا مولفه‌ی متصل
- X_{right}= مختصات X گوشه‌ی سمت راست بالا مولفه‌ی متصل



Y

Y

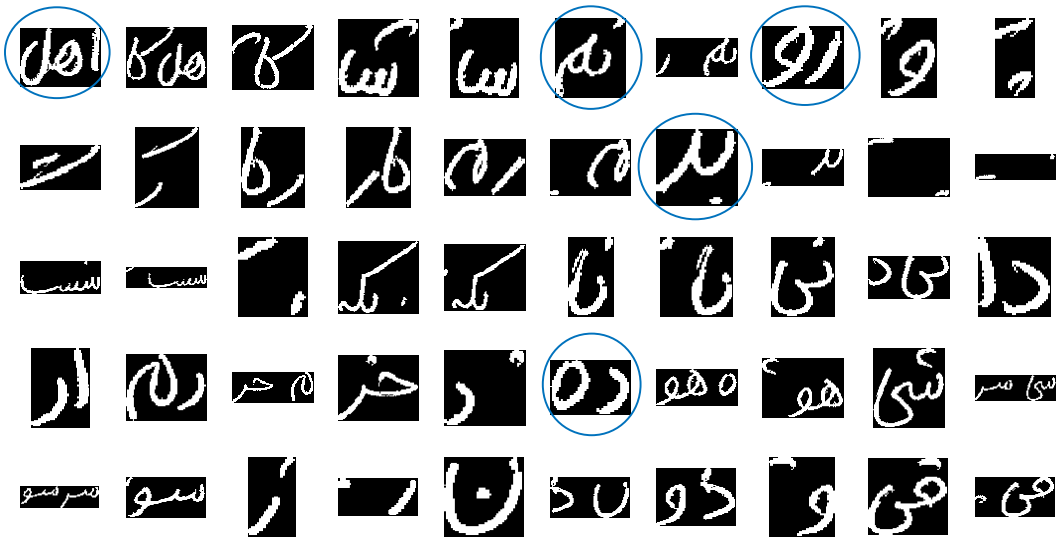
X

X

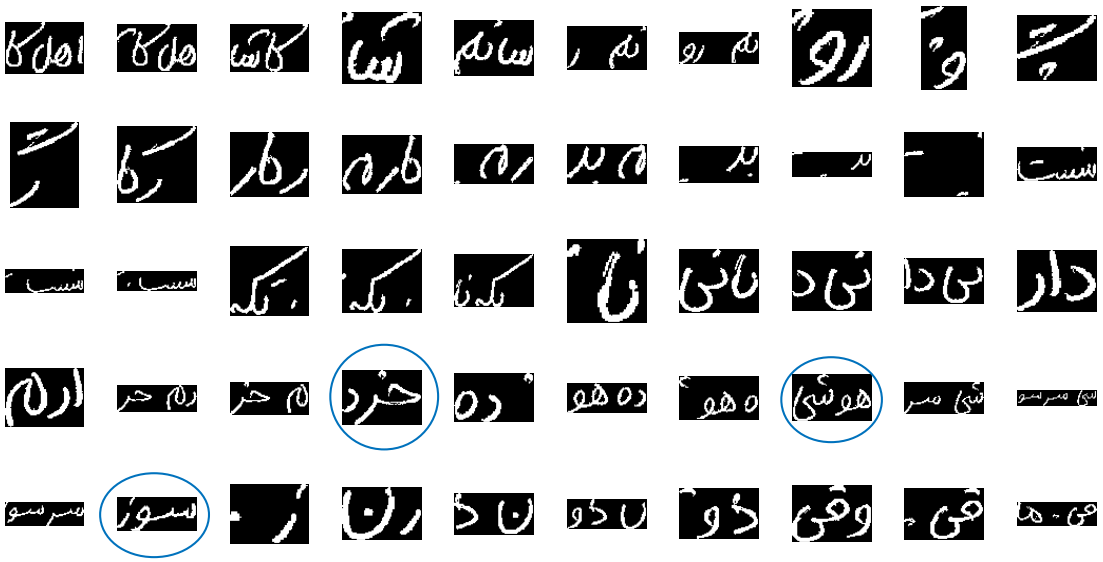
(۵-۵)

شکل ۵-۴۰:روش بدست آوردن موقعیت مکانی تصویر هر کلمه

در شکل ۵-۴۱ تا شکل ۵-۴۴ نتایج اجرای این روش برای ترکیب‌های ۲ تا ۵ تـایی بـر روی خـط اول نمایش داده شده است. کلمات با دایره مشخص شده‌اند.

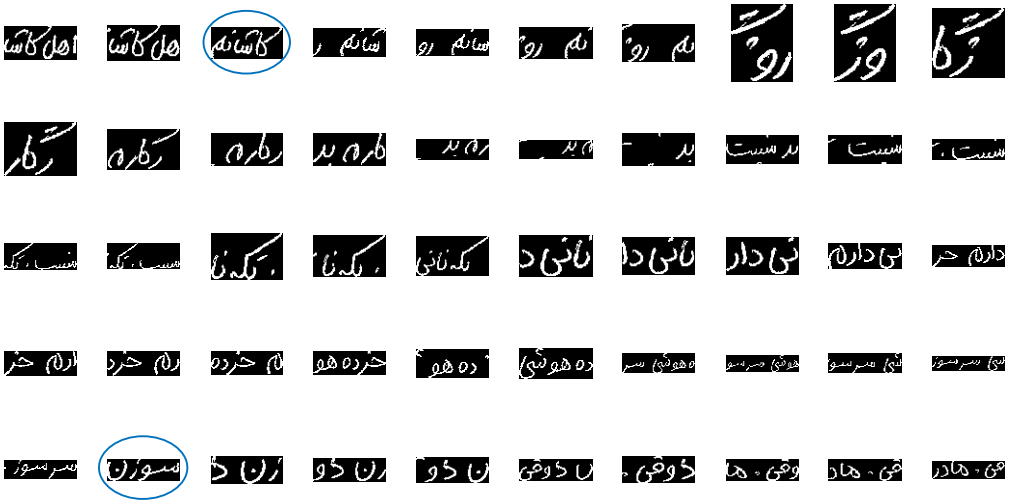


شکل ۵-۴۱:استخراج ترکیب‌های دو تایی از مولفه‌های متصل شکل



شکل ۵-۴۲:استخراج ترکیب‌های سه تایی از مولفه‌های متصل

شکل ۵-۴۳: استخراج ترکیب‌های چهار تایی از مولفه‌های متصل



شکل ۵-۴۴: استخراج ترکیب‌های پنج تایی از مولفه‌های متصل

۵-۷ نتایج روش پیشنهادی

برای ارزیابی روش پیشنهادی از پایگاه داده FHT که در فصل چهارم معرفی شد استفاده شده است. همان طور که بیان شد این پایگاه داده شامل ۱۰۰۰ سند می‌باشد که دارای ۴۰ دسته متن متفاوت است که توسط ۲۵ نویسنده متفاوت نوشته شده‌اند. برای ارزیابی این قسمت از هر دسته متن، ۴ تا با در نظر گرفتن تنوع از جهت کج و افقی بودن خطوط و فاصله‌ی نوشته شدن خطوط از هم که متوسط یا کم است انتخاب شده‌اند که در مجموع ۱۶۰ سند می‌باشد. در این اسناد از نظر اندازه، انواع دست‌خط‌های کوچک و بزرگ و متوسط وجود دارد. این اسناد شامل هم دست‌خط خوانا و خوش‌خط و هم غیرخوانا و بد خط است. ولی روش پیشنهادی مستقل از نوع دست‌خط و اندازه است.

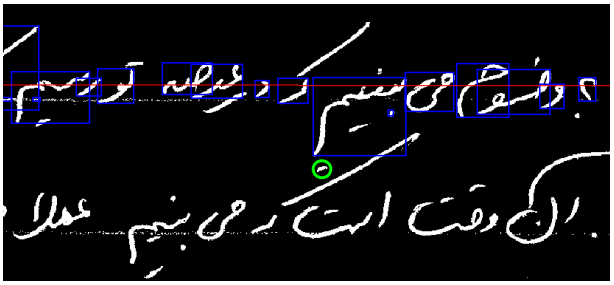
روش پیشنهادی در دو قسمت جداسازی خطوط و استخراج برون خط دنباله‌ای از مولفه‌های متصل مورد ارزیابی قرار گرفته است. برای جداسازی خطوط از نرخ آشکارسازی که در[۵۶، ۵۷] تعریف شده استفاده می‌شود. طبق این معیار اگر اشتراک بین مجموعه پیکسل‌های خط صحیح و خط بدست آمده‌ی متناظر با آن، تقسیم بر اجتماع بین پیکسل‌های دو مجموعه از ۹۰ درصد بیشتر باشد خط درست استخراج شده است. در رابطه‌ی (۵-۶)، G نشان‌دهنده‌ی خط صحیح و R خط بدست آمده‌ی متناظر با آن است. P مجموعه پیکسل‌های پیش زمینه و M میزان انطباق را نشان می‌دهد.

$$M \qquad (۵-۶)$$

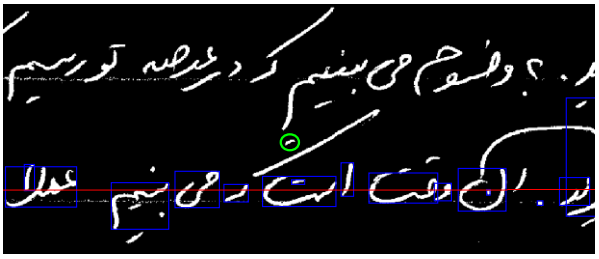
تعداد کل خط‌های مورد آزمایش ۱۰۵۱ است که طبق معیار بالا ۱۰۴۶ خط درست استخراج شده‌اند که معادل با نرخ آشکارسازی ۹۹/۵٪ درصد است. مزیت روش معرفی شده در این پایان نامه را می‌توان به این صورت بیان کرد که برای متونی که داری خطوط تا حدی کج می‌باشند و یا خطوط با فاصله‌ی کمی از هم نوشته شده‌اند نیز نتایج قابل قبولی ارائه می‌کند. ایجاد خطا در مرحله‌ی

استخراج برون‌خط تصاویر کلمات به دو عامل مربوط می‌شود؛ جداسازی خطوط و جایگذاری. بیشتر خطایی که در مرحله جداسازی خطوط رخ می‌دهد مربوط به نقاط است. که در واقع اشتباه در اختصاص یک نقطه به یک خط، در این قسمت هم ایجاد خطا می‌کند. در قسمت جایگذاری هم با توجه به این نحوه نگارش فارسی بخصوص در متون دست‌نویس بسیار متنوع است، امکان این که با توجه به قواعد در نظر گرفته شده در بخش ۵-۴، مولفه متصلی در مکان اشتباه جایگذاری شده باشد وجود دارد. خطا ایجاد شده در مرحله‌ی جایگذاری زمانی رخ می‌دهد که مولفه‌ی متصل جا افتاده مربوط به شروع و یا انتهای کلمه‌ی مورد نظر باشد و ترتیب جایگذاری آن به شکلی باشد که مولفه‌ی متصلی که مربوط به کلمه‌ی بعدی یا قبلی است زودتر از آن شناسایی شوند. در ادامه چند مورد از خطاهای بوجود آمده نشان داده شده است.

در شکل ۵-۴۵ نقطه‌ی کلمه‌ی "می‌بینیم" که با سبز مشخص است در مرحله‌ی جایگذاری به دلیل اینکه با فاصله از بدنه‌ی اصلی نوشته شده است به معادله‌ی خط (خط قرمز رنگ) زیرین نزدیکتر است. به همین دلیل این نقطه به اشتباه، به این خط اختصاص پیدا می‌کند.(شکل ۵-۴۶).

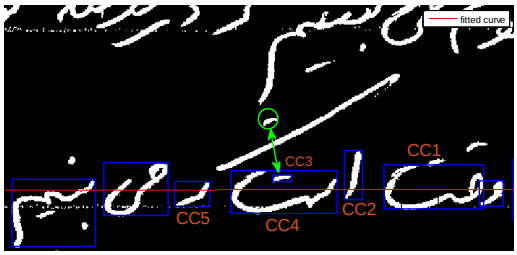


شکل ۵-۴۵: نمایش فاصله‌ی نقطه‌ی "می‌بینیم" با معادله‌ی خط پایه

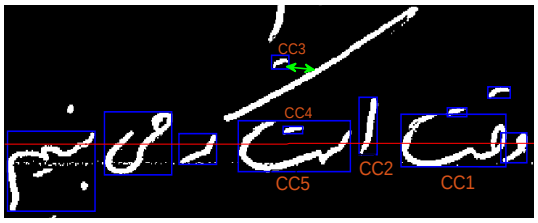


شکل ۵-۴۶: نمایش فاصله‌ی نقطه‌ی "می‌بینیم" با معادله‌ی خط زیرین

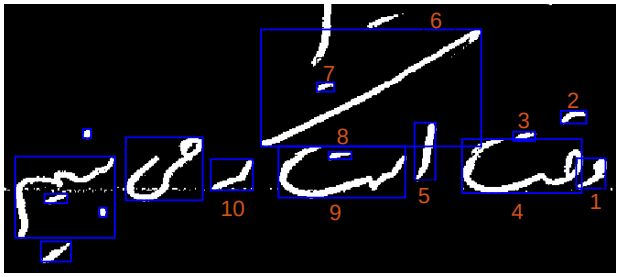
اشتباه در اختصاص این نقطه به این خط و اینکه سرکش حرف "که" بسیار کشیده و با فاصله از بدنه‌ی اصلی خود نوشته شده است باعث می‌شود جایگذاری حرف اضافه‌ی "که" نیز اشتباه انجام شود. با توجه به اینکه نقطه‌ی"می‌بینیم" برای خط زیرین خود مولفه‌ی متصل بالارونده محسوب می-شود با توجه به قواعد بخش۵-۴-۱ جایگذاری می‌شود. ابتدا مولفه‌ی متصل نقطه و بعد از آن مولفه‌ی متصل سرکش جایگذاری می‌شوند. در شکل ۵-۴۷ و شکل ۵-۴۸ پنج مولفه‌ی متصل خط اصلی که به مولفه‌های متصل بالارونده نقطه و سرکش نزدیک هستند مشخص شده‌اند. مولفه‌ی متصلی که با فلش سبز رنگ مشخص است نزدیکترین مولفه‌ی متصل اصلی به مولفه‌ی متصل جا افتاده است. مولفه‌ی متصل بالارونده قبل از آن قرار می‌گیرد. در نهایت در شکل ۵-۴۹ ترتیب کلی قرار گرفتن مولفه‌های متصل مشخص شده است.



شکل ۵-۴۷: جایگذاری مولفه‌ی متصل نقطه



شکل ۵-۴۸: جایگذاری مولفه‌ی متصل سرکش



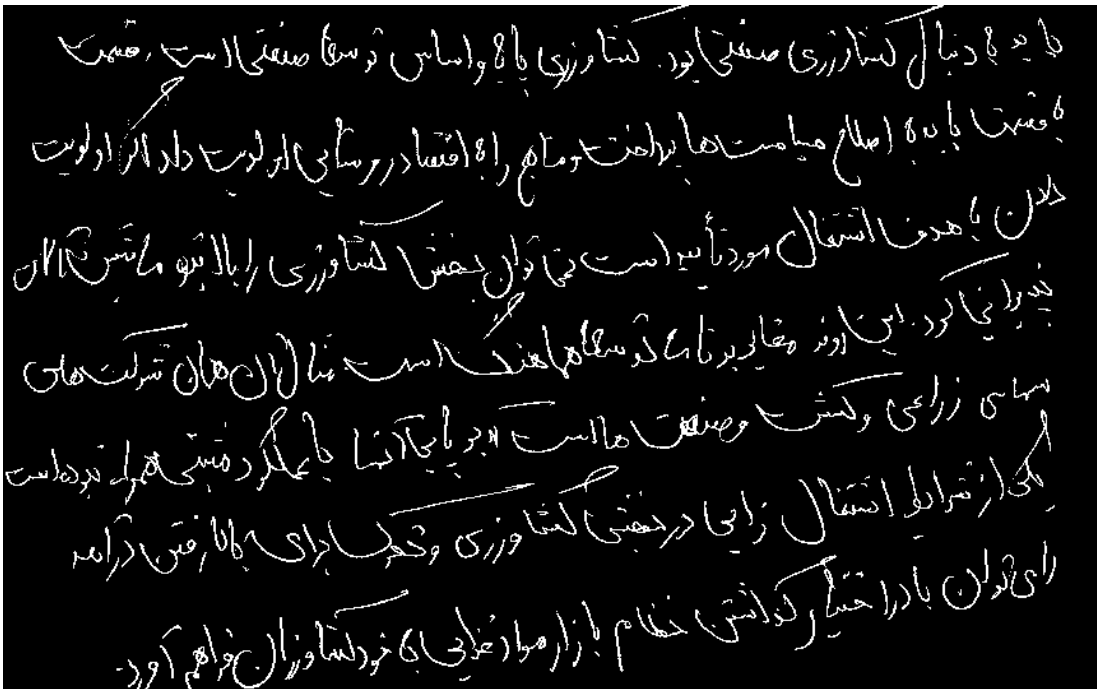
شکل ۵-۴۹: ترتیب قرار گرفتن مولفه‌های متصل پشت سر هم بعد از جایگذاری نقطه و سرکش

تعداد کل کلمات مورد آزمایش ۱۳۱۰۸ و تعداد کل حروف اضافه ۳۸۰۰ می‌باشد که از این تعداد ۷۳۹ کلمه و ۸۸ حرف اضافه به درستی آشکار نشده‌اند که معادل با درصد دقت ۹۴/۳۶٪ برای کلمات و ۹۷/۶٪ برای حروف اضافه است. جدول ۵-۱ میزان ایجاد خطا توسط دو عامل بالا را به تفکیک نشان می‌دهد. در این جدول در سطر اول تعداد کلماتی قرار دارند که بر اثر جایگذاری اشتباه مولفه متصلی که جز بدنه اصلی کلمه است دچار خطا شده‌اند. در سطر دوم کلماتی که بر اثر جایگذاری اشتباه نقاط دچار خطا شده‌اند و به همین ترتیب در سطر سوم کلماتی که بر اثر اشتباه در اختصاص یک نقطه به یک خط دچار خطا شده‌اند را نشان می‌دهد.

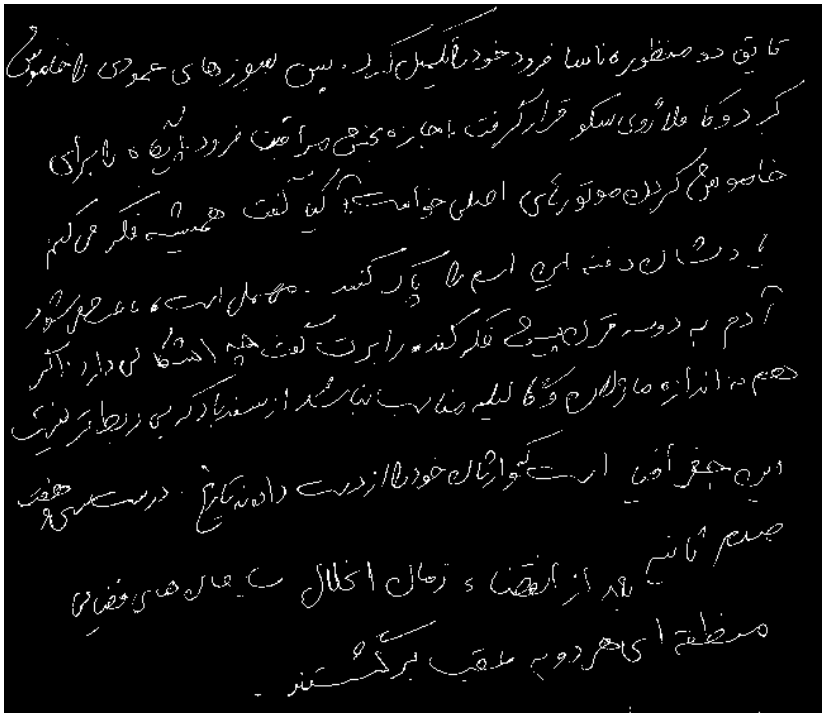
جدول ۵-۱: میزان ایجاد خطا توسط دو عامل بالا

عامل ایجاد خطا	تعداد کلمات	درصد ایجاد خطا
جایگذاری اشتباه بدنه اصلی	۴۱۱	۵۵/۶
جایگذاری اشتباه نقاط	۵۸	۷/۸
جداسازی اشتباه خطوط در نقاط	۲۷۰	۳۶/۵

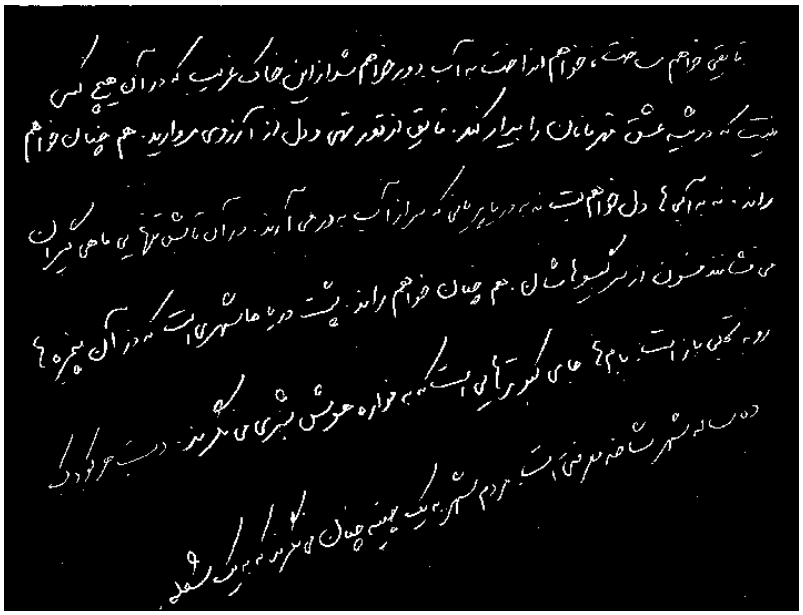
در ادامه چند سند تصویری که با روش پیشنهادی خطوط متن آنها بـه درسـتی شناسـایی شـده، نشان داده شده است. (شکل ۵-۵۰ تا شکل ۵-۵۳)



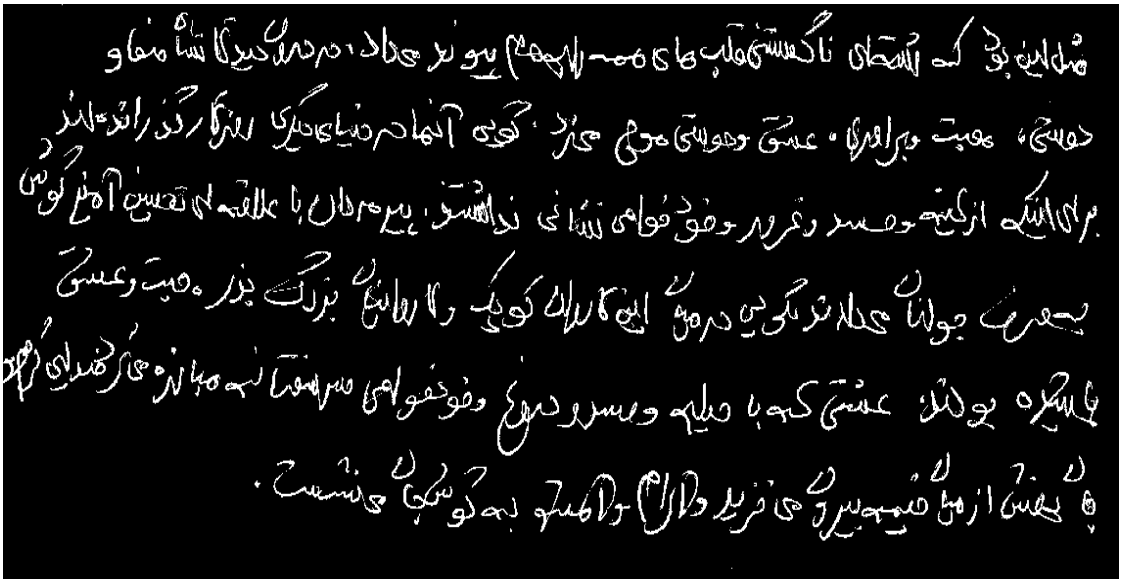
شکل ۵-۵۰:جداسازی خطوط متن با بکارگیری روش پیشنهادی



شکل ۵-۵۱: جداسازی خطوط متن با بکارگیری روش پیشنهادی

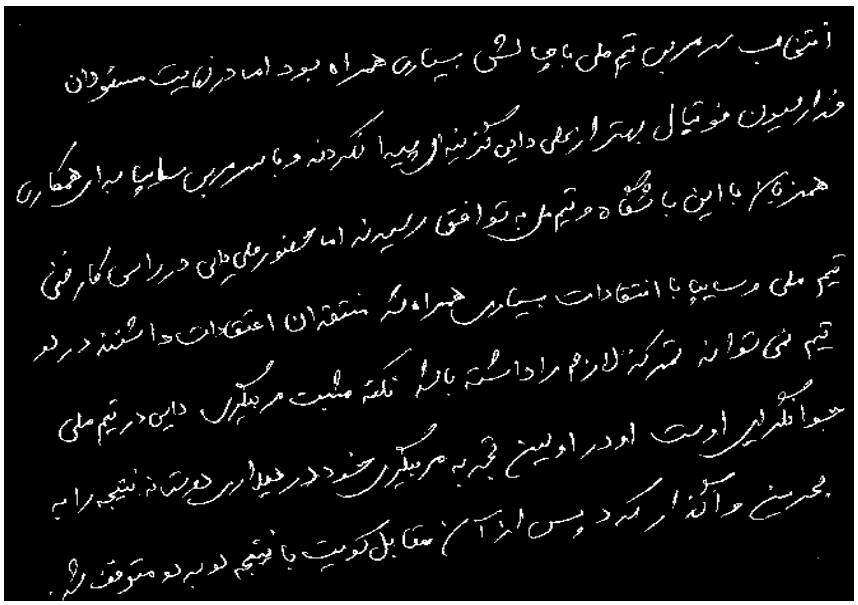


شکل ۵-۵۲: جداسازی خطوط متن با بکارگیری روش پیشنهادی



شکل ۵-۵۳: جداسازی خطوط متن با بکارگیری روش پیشنهادی

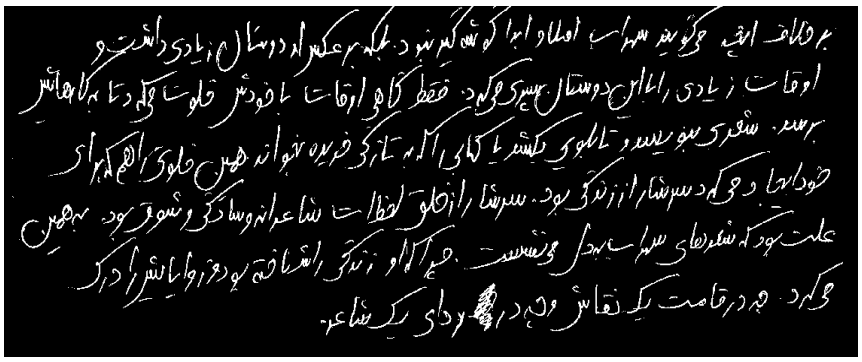
در شکل ۵-۵۴ تا شکل ۵-۵۹ عملکرد روش پیشنهادی در مورد متونی که دارای خطوط کج هستند یا خطوط با فاصله‌ی کم از هم نوشته شده‌اند نمایش داده شده است.



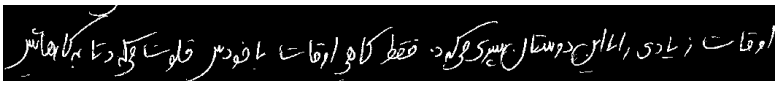
شکل ۵-۵۴: عملکرد روش پیشنهادی در مورد خطوط متن که تا حدی کج نوشته شده‌اند

شکل ۵-۵۵: سه خط اول متن بالا که به درستی جدا شده‌اند

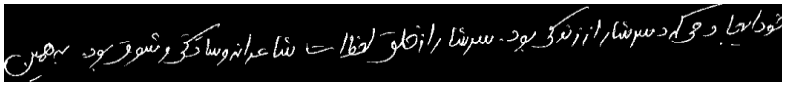
شکل ۵-۵۶:چهار خط آخر متن بالا که به درستی جدا شده‌اند

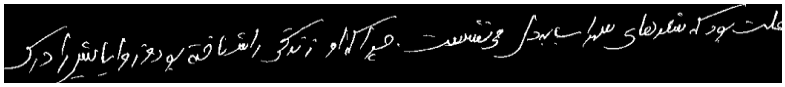


شکل ۵-۵۷: عملکرد روش پیشنهادی در مورد خطوط متنی که با فاصله کم از هم نوشته شده‌اند.



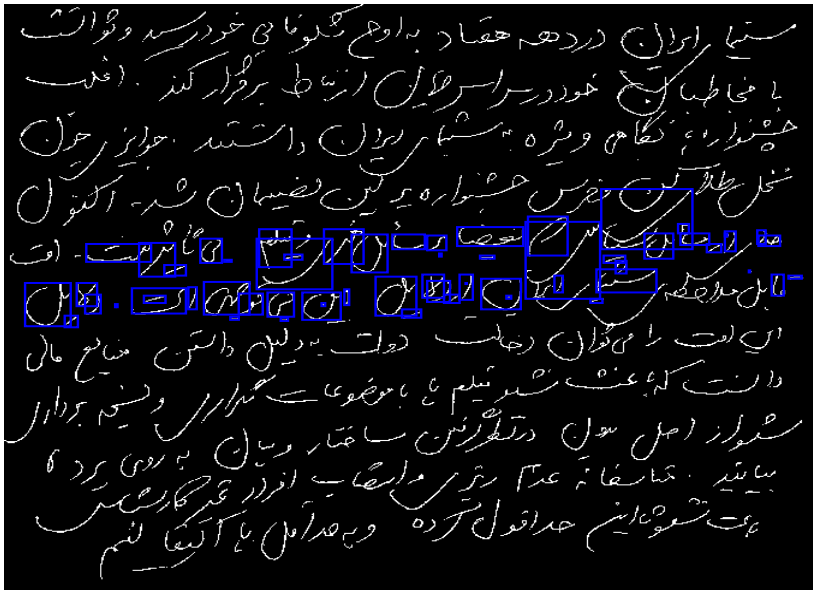
شکل ۵-۵۸: سه خط اول متن بالا که به درستی جدا شده‌اند.



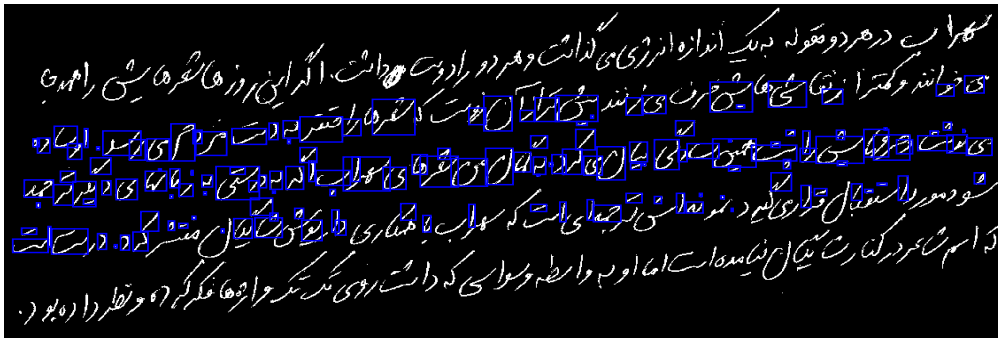


شکل ۵-۵۹: سه خط آخر متن بالا که به درستی جدا شده‌اند.

دو نمونه از اسناد تصویری که روش پیشنهادی قادر به جداسازی خطوط در مورد آنها نبوده است در شکل ۵-۶۰ و شکل ۵-۶۱ نمایش داده شده‌اند. در شکل ۵-۶۰ مولفه‌های متصل خط ۵ و ۶ به دلیل تو هم رفتگی شدید خطوط و اتصال بین مولفه‌های متصل این دو خط به یکدیگر، همه به عنوان مولفه‌های متصل خط ۵ تشخیص داده شده‌اند که باعث عدم تشخیص جداگانه این دو خط از هم شده است. در شکل ۵-۶۱ به دو دلیل، کجی خطوط و فاصله‌ی نزدیک خطوط به یکدیگر، روش قادر جداسازی خطوط نشده است.



شکل ۵-۶۰: عدم توانایی تشخیص مولفه‌های متصل مربوط به هر خط به دلیل اتصال بین مولفه‌های متصل خطوط



شکل ۵-۶۱: عدم توانایی تشخیص مولفه‌های متصل مربوط به هر خط به دلیل کجی زیاد و فاصله‌ی کم خطوط از هم

فصل ششم:

نتیجه‌گیری و پیشنهادات

روش پیشنهادی این پایان نامه یک رهیافت جستجوی کلمه برای دست‌نوشته‌های زبان فارسی است که هدف آن یافتن یک کلمه پرسشی در یک مجموعه داده شامل اسناد تصویری است. در این روش کلمه‌ی پرسشی مورد نظر کاربر در اسناد تصویری جستجو شده و مکان نمونه‌های آن کلمه در داخل اسناد مشخص می‌شود. پرسش می‌تواند بصورت تصویر یا رشته‌ی متنی کلمه باشد.

در این سیستم در اولین گام لازم است کلمات موجود در اسناد تصویری استخراج شوند. برای این مرحله ابتدا کل مولفه‌های متصل موجود در متن استخراج شده‌اند. سپس با در نظر گرفتن اینکه در فارسی از بالا سمت راست نوشتن آغاز می‌شود؛ مولفه‌ی متصلی که مختصات مرکز آن دارای بیشترین مقدار x و کمترین مقدار y است و تعداد پیکسل های آن از یک آستانه‌ی در نظر گرفته شده بیشتر می‌باشد، به عنوان مولفه‌ی متصلی که شروع کننده خط است انتخاب می‌گردد. نزدیک‌ترین مولفه متصل در متن به این مولفه متصل اولیه با در نظر گرفتن دو شرط انتخاب می‌شود (۱. مختصات x مرکز نزدیک‌ترین مولفه متصل باید از مختصات x مرکز مولفه متصل اولیه کمتر باشد۲. قدر مطلق اختلاف بین مختصاتy مراکز آنها از یک آستانه در نظر گرفته شده کمتر باشد). با اعمال روش نزدیک‌ترین همسایه و در نظر گرفتن دو شرط بالا تمام مولفه‌های متصل خط بصورت یک زنجیزه و به همان ترتیبی که در متن قرار دارند بدست می‌آیند. بعد از بدست آمدن مولفه‌های متصل مربوط به هر خط عملیات پس‌پردازشی روی آنها انجام می‌شود.

این پس‌پردازش شامل یافتن معادله‌ی خط و حذف مولفه‌های متصلی است که در مرحله‌ی قبل اشتباها به خط اختصاص داده شده‌اند. در این مرحله اگر فاصله‌ی هر مولفه متصل با معادله‌ی خط از یک حد آستانه بیشتر بود این مولفه متصل مربوط به این خط نبوده و از لیست آن حذف می‌شود. بعد از انجام این پس‌پردازش، مولفه‌های متصل نهایی این خط از لیست کل مولفه‌های متصل متن حذف می‌شوند. این الگوریتم به تعداد خطوط متن انجام می‌شود تا هر بار یک خط با حذف خط‌های قبل از خودش مورد پردازش قرار گیرد و مولفه‌های متصل مربوط به آن مشخص شوند.

بعد از اجرای این الگوریتم بر روی تمام خطوط متن، مولفه‌های متصلی باقی خواهند ماند که به هیچ خطی اختصاص نیافته‌اند که با پیدا کردن مکان آنها در متن، جایگذاری می‌شوند. به عبارت دیگر در مرحله جایگذاری در مورد تمام مولفه‌های متصلی که در مرحله‌ی قبلی به هیچ خطی اختصاص نیافته‌اند تصمیم‌گیری می‌شود. بیشتر این مولفه‌های متصل را سرکش‌ها و نقاط تشکیل داده‌اند. در این مرحله، با در نظر گرفتن یک نوار ستونی با مرکزیت مختصات x هر مولفه‌ی متصل و جاروب افقی تصویر شامل معادلات خطوط، به تعداد خطوط متن قله تشکیل خواهد شد که مختصات y این مولفه-ی متصل به هر قله که نزدیکتر بود به آن خط اختصاص دارد. بعد از مشخص شدن خط، با یافتن نزدیکترین مولفه‌های متصل خط اصلی، این مولفه‌ی متصل بین، قبل یا بعد از آنها جایگذاری می‌شود.

با در نظر گرفتن این قانون که اکثر حروف اضافه و کلمات فارسی از ترکیب یک تا هفت مولفه متصل تشکیل شده‌اند استخراج کلمات متن انجام می‌شود. مولفه‌های متصل هر خط از نظر مکانی به ترتیب از سمت راست متن به سمت چپ ذخیره شده‌اند. از اولین مولفه متصل هر خط شروع کرده و به ترتیب برای ساختن ترکیب‌های دوتایی، دو تا مولفه متصل اول را در نظر گرفته و تصویری شامل هر دو مولفه متصل بدست می‌آید. ابعاد این تصویر با توجه به مختصات این دو مولفه متصل تعیین می‌شود. این روش تا آخر برای کل مولفه‌ها متصل این خط با در نظر گرفتن گام یک ادامه پیدا می‌کند. همین مراحل برای ترکیب سه تایی تا هفت تایی و روی تمام خطوط اجرا می‌شود و در نهایت تصویر تمام کلماتی که در متن وجود دارند ذخیره می‌شود. مولفه‌های متصلی که مربوط به یک کلمه مستقل در متن هستند به همان ترتیبی که در کلمه وجود دارند با این روش شناسایی شده‌اند.

این روش روی پایگاه داده‌ای شامل ۱۶۰ سند تصویری دست‌نوشته‌ی فارسی ارزیابی شده است. نرخ آشکارسازی در مرحله جداسازی خطوط برابر با ۹۹/۵٪ می‌باشد. درصد دقت در مرحله استخراج دنباله‌ای از مولفه‌های متصل، ۹۴/۳۶٪ برای کلمات و ۶/ ۹۷٪ برای حروف اضافه است. روش فوق در

متونی که دارای خطوط افقی و کج هستند و یا با فاصله کمی از هم نوشته شده‌اند عملکرد قابل قبولی دارد. یکی از مشکلات روش در متونی است که خطوط کج با فاصله‌ی کمی از هم نوشته شده‌اند و یا بین مولفه‌های متصل خطوط اتصال وجود دارد.

در نتیجه بعد از گام اول مجموعه‌ای از کلمات کاندید، ایجاد شده که در گام بعدی هدف یافتن نمونه‌های کلمات پرسشی در این مجموعه و مشخص کردن مکان آنها در متن می‌باشد. این مجموعه‌ی کاندید به همراه تصاویر کلمات پرسشی و رشته‌های متنی متناظر با کلمات پرسشی در یک فضای برداری مشترک تعبیه می‌شوند. در این فضا، تصویر و رشته‌ی متنی کلمه‌ی پرسشی، و تصاویری از کلمات مجموعه‌ی کاندید، که کلمه‌ی یکسانی را نشان می‌دهند نزدیک به یکدیگر قرار می‌گیرند. با روش طبقه‌بندی نزدیک‌ترین همسایگی عملیات تشخیص انجام می‌شود.

برای بازنمایی رشته‌های متنی، از مشخصه‌ای به نام هیستوگرام هرمی حروف استفاده می‌شود که رشته‌های متنی را در فضای باینری به نام فضای برچسب تعبیه می‌کند. به کارگیری این نوع مشخصه باعث می‌شود که شباهت بین کلمات مستقل از شیوه نگارش حفظ شود و منجر به روشی می‌شود که از دانش بین کلاس‌ها بهره می‌برد. در واقع الگوریتم قادر است برای کلماتی که آموزش ندیده‌اند در مرحله‌ی آزمون عملیات جستجو را انجام دهد و نتایج مناسبی داشته باشد. از بردارهای فیشر به عنوان نمایش پایه تصویر استفاده شده است و بردارهای ویژگی برای تصاویر استخراج می‌شوند. سپس این بردارهای ویژگی به همراه برچسب‌های PHOC متناظر برای آموزش مشخصه‌ها به کار می‌روند. با در نظر گرفتن الفبای فارسی و استفاده از روش PHOC، ۶۵۴ مشخصه وجود دارد که برای آموزش آنها از SVM خطی استفاده شده است. به عبارت دیگر برای آموزش مشخصه‌ی iأم ، مجموعه‌ای از نمونه‌های مثبت و منفی آن استفاده می‌شوند.(نمونه‌های مثبت، کلماتی هستند که دارای این مشخصه می‌باشند در واقع برای این نمونه‌ها در بُعد iأم هیستوگرام یک قرار گرفته و برای نمونه‌های منفی که دارای این مشخصه نیستند در بُعد iأم صفر قرار دارد). آموزش مشخصه‌ها منجر به ایجاد فضایی به نام فضای

مشخصه می‌شود که در این فضا تصاویر کلماتی که یک کلمه‌ی یکسان با دست‌خط‌های متفاوت‌اند نزدیک به یگدیگر قرار می‌گیرند.

در نهایت از یک روش کالیبره کردن به نام تحلیل هم‌بستگی کانونی برای تصحیح مقادیر ویژگی‌ها بهره برده که منجر به اصلاح بهتر نتایج شده است. با استفاده از این روش رتبه مشخصه‌ها و برچسب‌های PHOC در یک فضای برداری مشترک تعبیه می‌شوند که در این فضا همبستگی بین آنها ماکزیمم است و با روش طبقه‌بندی نزدیک‌ترین همسایگی عملیات تشخیص انجام می‌شود.

برای ارزیابی این الگوریتم از دو پایگاه داده فارسا که شامل تصاویر کلمات با دست‌خط‌های متفاوت است و پایگاه داده متن دست نوشته‌ی فارسی که شامل اسناد تصویری با دست‌خط‌های متفاوت است به ترتیب برای مرحله‌ی آموزش و آزمون استفاده شده است. آموزش فضای مشخصه و فضای برداری مشترک با استفاده از پایگاه داده‌ی فارسا انجام شد. و در مرحله‌ی آزمون با تایپ کلمه‌ی پرسشی یا تصویرآن به عنوان ورودی، مکان نمونه‌های آن کلمه در متون مشخص می‌شود.

بیشتر تحقیقات انجام شده در این زمینه بر روی متون زبان لاتین ارائه شده است و تحقیقات انجام شده در زبان فارسی و عربی بسیار کم می باشد که اکثر آنها هم بر روی اسناد چاپی هستند. با توجه به اینکه کار قابل توجهی در روی متون دست‌نوشته‌ی فارسی انجام نشده است، این پایان نامه نتایج مناسبی در این زمینه دارد. برای حالت پرسش با تصویر، بیشترین مقدار F_1 در آستانه‌ی ۰/۹۴۲۵ بدست آمده است که برابر با ۷۲/۵۴٪می‌باشد که در این آستانه میزان دقت ۷۵/۴۵٪ و میزان فراخوانی ۸۰/۴۲٪ است. این آستانه به عنوان بهترین آستانه انتخاب می‌شود. برای حالت پرسش با متن بیشترین مقدار F_1 در آستانه‌ی ۰/۹۴۲۵ بدست آمده است که برابر با ۷۵/۹۲٪ می‌باشد که در این آستانه میزان دقت ۸۰/۲۸٪ و میزان فراخوانی ۷۸/۹۲٪ است. در مقایسه با حالت پرسش با تصویر، F_1 به میزان ۳/۳۸ و دقت به میزان ۴/۸۳ افزایش و فراخوانی به میزان ۱/۵ کاهش یافته است.

با توجه به نتایج بدست آمده در این پایان نامه، برخی پیشنهادات برای ادامه کار در این زمینه ارائه می‌شود:

۱) استفاده از این روش برای متون چاپی که با توجه به اینکه متون چاپی از پیچیدگی‌های کمتری نسبت به متون دست‌نوشته برخوردارند، این روش بر روی آنها نتایج بهتری خواهد داشت.

۲) ساختار زبان فارسی به دلیل وجود علائم نشانه‌گذاری فراوان مانند سرکش‌ها، تنوع در تعداد و موقعیت نقاط در بالا و پایین حروف، تشکیل شدن کلمات از زیرکلمات به جای حروف و وجود فاصله بین زیرکلمات بسیار پیچیده‌تر از زبان انگلیسی می‌باشد، به همین دلیل ایجاد مشخصه‌ای که بتواند این موارد را نیز در نظر بگیرد، نتایج بهتری روی زبان فارسی خواهد داشت.

۳) در برخی کلمات در فارسی حروف تکراری وجود دارد که با روش مطرح شده در این پایان نامه موقعیت یک تکرار از آن قابل نمایش است، استفاده از مشخصه‌ای که تکرار حروف در یک کلمه را نیز در نظر بگیرد، تمیز دهنده‌ی بهتری برای کلمات فارسی است و نتایج بهتری خواهد داشت.

۴) استفاده از ویژگی‌های سریع‌تر از SIFT در مرحله‌ی استخراج ویژگی می‌تواند سرعت جستجو الگوریتم را بهبود دهد.

مراجع

[1] T. Kameshiro, T. Hirano, Y. Okada, and F. Yoda, "A document image retrieval method tolerating recognition and segmentation errors of OCR using shape-feature and multiple candidates," in *Document Analysis and Recognition, 1999. ICDAR'99. Proceedings of the Fifth International Conference on*, 1999, pp. 681-684.

[2] L. Li, S. Lu, and C. L. Tan, "A fast keyword-spotting technique," in *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, 2007, pp. 68-72.

[3] G. Jones, J. Foote, K. Sparck Jones, and S. Young, "Video mail retrieval: The effect of word spotting accuracy on precision," in *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95. International Conference on*, 1995, pp. 309-312.

[4] Y. Lu, L. Zhang, and C .L. Tan, "Retrieving imaged documents in digital libraries based on word image coding," in *Document Image Analysis for Libraries, 2004. Proceedings. First International Workshop on*, 2004, pp. 174-187.

[5] V. Frinken, A. Fischer, R. Manmatha, and H. Bunke,"A novel word spotting method based on recurrent neural networks," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, pp. 211-224, 2012.

[6] S. H. Kim, S. C. Park, C. B. Jeong, J. S. Kim, H. R. Park, and G. S. Lee, "Keyword spotting on Korean document images by matching the keyword image," in *Digital libraries: implementing strategies and sharing experiences*, ed: Springer, 2005, pp. 158-166.

[7] Y. Lu and C. L. Tan, "Chinese word searching in imaged documents," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, pp. 229-246, 2004.

[8] S. Belongie, G. Mori, and J. Malik, "Matching with shape contexts," in *Statistics and Analysis of Shapes*, ed: Springer, 2006, pp. 81-105.

[9] G. L. Scott and H. C. Longuet-Higgins, "An algorithm for associating the features of two images," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 244, pp. 21-26, 1991.

[10] T. M. Rath and R. Manmatha, "Word spotting for historical documents," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 9, pp. 139-152, 2007.

[11] T. Konidaris, B. Gatos, K. Ntzios, I. Pratikakis, S. Theodoridis, and S. J. Perantonis, "Keyword-guided word spotting in historical printed documents using synthetic data and user feedback," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 9, pp. 167-177, 2007.

[12] S. N. Srihari, H. Srinivasan, C. Huang, and S. Shetty, "Spotting words in Latin, Devanagari and Arabic scripts," *VIVEK-BOMBAY*, vol. 1, pp. 2-9, 2006.

[13] B. Zhang, S. N. Srihari, and C. Huang, "Word image retrieval using binary

features," in *Electronic Imaging 2004*, pp. 45-53.
[14] Y. POURASAD, H. HASSIBI, and A. GHORBANI, "A Farsi/Arabic Word Spotting Approach for Printed Document Images," *International Journal of Natural & Engineering Sciences,* vol. 6, pp.15-18, 2012.
[15] S. Lu and C. L. Tan, "Keyword spotting and retrieval of document images captured by a digital camera," in *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, 2007, pp. 994-998.
[16] S. Bai, L. Li, and C. L. Tan, "Keyword spotting in document images through word shape coding," in *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, 2009, pp. 331-335.
[17] A. Kefali and C. Chemmam, "A Semi-Automatic Approach of old Arabic Documents Indexing," in *CIIA*, 2011.
[18] Y. Pourasad, H. Hassibi, and A. Ghorbani, "A word spotting method for Farsi machine-printed document images," *Turkish Journal of Electrical Engineering & Computer Sciences,* vol. 21, pp.1-13, 2013.
[19] G. R. Ball, S. N. Srihari, and H. Srinivasan, "Segmentation-based and segmentation-free methods for spotting handwritten arabic words," in *Tenth International Workshop on Frontiers in Handwriting Recognition* ,2006.
[20] S. Srihari, H. Srinivasan, P. Babu, and C. Bhole, "Handwritten arabic word spotting using the cedarabic document analysis system," in *Proc. Symposium on Document Image Understanding Technology (SDIUT-05)*, 2005, pp. 123-132.
[21] G. Kim and V .Govindaraju, "A lexicon driven approach to handwritten word recognition for real-time applications," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 19, pp. 366-379, 1997.
[22] S. Marinai, "Text retrieval from early printed books ",*International Journal on Document Analysis and Recognition (IJDAR),* vol. 14, pp. 117-129, 2011.
[23] J. Almazán, A. Gordo, A. Fornés, and E. Valveny, "Segmentation-free word spotting with exemplar svms," *Pattern Recognition*, vol. 47, pp. 3967-3978, 2014.
[24] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *International journal of computer vision,* vol. 105, pp. 222-245, 2013.
[25] T. M. Rath and R. Manmatha, "Word image matching using dynamic time warping," in *Computer Vision and Pattern Recognition.Proceedings.IEEE Computer Society Conference on*, vol. 2, pp. II-521-II-527,2003.
[26] T. Adamek, N. E. O'Connor, and A. F. Smeaton, "Word matching using single closed contours for indexing handwritten historical documents," *International Journal of Document Analysis and Recognition (IJDAR),* vol. 9, pp. 153-165, 2007.
[27] J. A. Rodrıguez and F. Perronnin, "Local gradient histogram features for word spotting in unconstrained handwritten documents," in *Int. Conf. on Frontiers in Handwriting Recognition*, 2008.
[28] A. Fischer, A. Keller, V. Frinken, and H. Bunke, "HMM-based word spotting in handwritten documents using subword models," in *Pattern recognition (icpr), 2010 20th international conference on*, 2010, pp. 3416-3419.
[29] J. Rodriguez-Serrano and F. Perronnin, "A model-based sequence similarity with

application to handwritten word spotting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, pp. 2108-2120,2012.

[30] A. Abidi, A. Jamil, I. Siddiqi, and K. Khurshid, "Word spotting based retrieval of urdu handwritten documents," in *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*, 2012, pp. 331-336.

[31] Y. Pourasad, H .Hassibi, and A. Ghorbani, "Farsi Word Spotting and Font Size Recognition," *Procedia Technology*, vol. 1, pp. 372-377, 2012.

[۳۲] علی‌آبادی.م، صدری.ج، رضوی.م، "یک روش ترکیبی برای جستجوی کلمات در کتب تاریخی دست‌نویس فارسی "، اولین کنفرانس بازشناسی الگو و تحلیل تصویر ایران، ۱۳۹۱، ص ۴۱۸-۴۲۳.

[33] R. Manmatha, C. Han, E. M. Riseman, and W. B. Croft, "Indexing handwriting using word matching," in *Proceedings of the first ACM international conference on Digital libraries*, 1996, pp. 151-159.

[34] S. Kane, A. Lehman, and E. Partridge, "Indexing George Washington's handwritten manuscripts," *Center for Intelligent Information Retrieval, Computer Science Department, University of Massachusetts, Amherst, MA*, vol. 1003, 2001.

[35] Y. Lu, C. L. Tan, W. Huang, and L. Fan, "An approach to word image matching based on weighted Hausdorff distance," in *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on*, 2001, pp. 921-925.

[36] Z. S. Harris, "Distributional structure," vol. 10, pp. 146-162, 1954.

[37] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learning in computer vision, ECCV*, 2004, pp. 1-16.

[38] F. Perronnin, "Universal and adapted vocabularies for generic visual categorization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, pp. 1243-1256, 2008.

[39] R. Shekhar and C. Jawahar, "Word image retrieval using bag of visual words," in *Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on*, 2012, pp. 297-301.

[40] M. W. Sagheer, N. Nobile, C. L. He, and C. Y. Suen, "A novel handwritten urdu word spotting based on connected components analysis," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 2013-2016.

[41] S. Wshah, G. Kumar, and V. Govindaraju, "Script independent word spotting in offline handwritten documents based on hidden markov models," in *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*, 2012, pp. 14-19.

[42] Z. Bahmani and R. Azmi, "Farsi/Arabic Document Image Retrieval through Sub-Letter Shape Coding for mixed Farsi/Arabic and English text" IJCSI International Journal of Computer Science Issues, Vol. 8, pp. 166-172, September 2011

[43] R. Manmatha, C. Han ,and E. M. Riseman, "Word spotting: A new approach to indexing handwriting," in *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, 1996, pp.

631-637.
[44] J. Almazán, A. Gordo, A. Fornés, and E. Valveny, "Word spotting and recognition with embedded attributes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, pp. 2552-2566, 2014.
[45] J. A. Rodriguez-Serrano, F. Perronnin, and F. Meylan, "Label embedding for text recognition," in *Proc. BMVC*, 2013.
[46] J. F. G. Srikantan and S. Srihari, "Handprinted character/digit recognition using a multiple feature/resolution philosophy," in *Proc. Fourth Int'l Workshop Frontiers in Handwriting Recognition*, 1994, pp. 57-66.
[47] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International journal of computer vision*, vol. 60, pp. 63-86, 2004.
[48] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, 1999, pp. 1150-1157.
[۴۹] عابدینی.ع، صمدزادگان.ف، "استخراج عوارض متنـاظر در تصـویر هـوایی و داده‌هـای LIDAR براساس الگوریتم SIFT"، همایش ژئوماتیک ۸۷، ۱۳۸۷.
[50] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91-110, 2004.
[51] M. Ziaratban, K. Faez, and F. Bagheri, "FHT: An unconstraint Farsi handwritten text database," in *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, 2009, pp. 281-285.
[52] H. Angrish and S. Kaur, "A Survey on Feature Description Techniques," *International Journal of Science and Research*, vol. 4, pp. 38-41, 2015.
[۵۳] نیکومرام.ه، پورزمانی. ز، "بررسی رابطه بـین الگوهـای پـیش‌بینـی بحـران مـالی"، پژوهشـنامه حسابداری مالی و حسابرسی"، دوره۱، شماره۱، ص ۳۳-۴۲، ۱۳۸۸.
[54] J. Almazan, A. Gordo, A. Fornés, and E. Valveny, "Handwritten word spotting with corrected attributes," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, 2013, pp. 1017-1024.
[۵۵] بایسته تاشک.الف، احمدی‌فرد.ع، "بازشناسی برون‌خط کلمات دست نوشته فارسی در یک مجموعه محدود لغات"، پایان نامه کارشناسی ارشد مهندسی برق-الکترونیک، دانشگاه شاهرود، زمستان ۱۳۹۰.
[56] M. Ziaratban and K. Faez, "An adaptive script-independent block-based text line extraction," in *2010 International Conference on Pattern Recognition*, 2010, pp. 249-252.
[57] Y. Li, Y. Zheng, D. Doermann, and S. Jaeger, "Script-independent text line segmentation in freestyle handwritten documents," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, pp. 1313-1329, 2008.

Abstract

Word spotting is a way to search for text information in document images. This process includes finding and locating a query word within a document image. In this context, more researches are provided on Latin documents. A Few researches in Arabic and Farsi has been reported, the majority of these researches are on printed documents. In this thesis we present a Farsi handwritten word spotting system. The proposed system is based on two steps. In the first stage, all existing connected components in the text image have been extracted, then the first connected component of each line is determined. Therefore by applying the nearest neighbor method, all the connected components of each line obtained according to their order in the text. As a result, after this stage, each of the connected components in the text devoted to one of the lines. Actually each line can be demonstrated with its connected components independent of other lines. The result of this stage is decomposing the text image into separated lines. In the next step with regard to the rule that the majority of prepositions and Persian words are combination of one to seven connected components, the words in the text are extracted. Actually the connected components of an independent word in the text are indentified in the same order of appearance in the word. For each of the word images extracted, the coordinates of their location in the text is saved.

As a result, after the first stage a set of candidate words is obtained. In the second stage, the goal is to find all instances of a query word in this set then their location in the text is specified. The query word may be a text string or may also be an image. In this stage, for word spotting we use attributes from Persian alphabet. The use of this kind of attributes makes word spotting independent of the writing style.

This approach has been applied on a dataset consisting of 55 query words and 100 Farsi handwritten document images. Average precision and recall for query by image is 75.45% and 80.42% and for query by text is 80.28% and 78.92%.

***Keywords**: :* Keyword spotting, Handwritten Farsi document images, Attributes, Label Embedding, Separating lines of text, Connected Components

Shahrood University of Technologhy

Faculty of Electrical Engineering

Handwritten Farsi WordSpotting Using Attributes embedding Technique

Bahare Asadi

Supervisor:

Dr. Alireza AhmadiFard

Adviser:

Majid Iranpour

February 2016