

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



پایان نامه دوره کارشناسی ارشد مهندسی برق - الکترونیک

بازشناسی برون خط کلمات دست‌نوشته فارسی در یک مجموعه‌ای از لغات

الهام بایسته تاشک

استاد راهنما:

دکتر علیرضا احمدی فرد

استاد مشاور:

دکتر حسین خسروی

زمستان 1390

تعهد نامه

اینجانب الهام بایسته تاشک..... دانشجوی دوره کارشناسی ارشد رشته مهندسی برق گرایش الکترونیک..... دانشکده مهندسی برق و رباتیک دانشگاه صنعتی شاهرود نویسنده پایان نامه با عنوان: بازشناسی برون خط کلمات دست نوشته فارسی در یک مجموعه‌ای از لغات.....

تحت راهنمایی آقای دکتر علیرضا احمدی فرد..... متعهد می شوم:

- 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 Technology» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده‌اند 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:

امضاء دانشجو:

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزارها، تجهیزات ساخته شده است) متعلق به دانشگاه 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.



مدیریت تحصیلات تکمیلی

شماره :
تاریخ :
ویرایش :

بسمه تعالی

فرم صورتجلسه دفاع از پایان نامه تحصیلی دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عج) ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای الهام بایسته تاشک رشته مهندسی برق گرایش الکترونیک تحت عنوان بازشناسی برون خط کلمات دست‌نوشته فارسی در یک مجموعه‌ای از لغات
.....

که در تاریخ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می‌گردد:

قبول (با درجه : امتیاز :) دفاع مجدد مردود

1- عالی (19 - 20)
2- بسیار خوب (18 - 18/99)

3- خوب (16 - 17/99)
4- قابل قبول (14 - 15/99)

5- نمره کمتر از 14 غیر قابل قبول

| عضو هیأت داوران | نام و نام خانوادگی | مرتبه علمی | امضاء |
|---------------------------------|--------------------|------------|-------|
| 1- استاد راهنما | | | |
| 2- استاد مشاور | | | |
| 3- نماینده شورای تحصیلات تکمیلی | | | |
| 4- استاد ممتحن | | | |
| 5- استاد ممتحن | | | |

رئیس دانشکده :

تقدیم به

پدر و مادر عزیزم و

همسر مهربانم که همیشه مشوق بنده بوده‌اند.

تشکر و قدردانی

بر خود لازم می دانم مراتب سپاس و قدردانی خویش را از 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 کمک های ایشان تأثیر بسزایی در پیشبرد هر چه بهتر و سریعتر این پایان نامه داشته است. همچنین از تذکرات خردمندانه جناب آقای دکتر خسروی قدردانی می کنم.

چکیده

در این پایان‌نامه بازنشاسی برون خط کلمات دست‌نوشته فارسی در یک فرهنگ لغت محدود مورد مطالعه قرار می‌گیرد. برای این منظور یک روش دو مرحله‌ای پیشنهاد می‌گردد. در مرحله نخست توسط الگوریتم‌های خوشه‌بندی سلسله‌مراتبی و ISOCLUS کلمات موجود در فرهنگ لغت بر اساس تشابه خوشه‌بندی می‌شوند. ویژگی‌های تشابهی به کار رفته در این مرحله، بردارهای پروفایل بالا، پایین، پروژکشن عمودی و تعداد گذر از سیاه به سفید برای هر ستون تصویر است. برای کاهش ابعاد ویژگی‌های استخراج شده و محدود کردن آشفتگی این سیگنال‌ها از تبدیل موجک یک بعدی استفاده شده است. برای اندازه‌گیری تشابه بردارهای ویژگی‌های دو کلمه از معیار فاصله DTW استفاده می‌کنیم. میانگین هر خوشه در فضای ویژگی‌ها به عنوان نماینده آن خوشه و مدخل مشترک اعضای آن خوشه در فرهنگ لغت تصویری، در نظر گرفته می‌شود. تعداد کلمات موجود در فرهنگ لغت مورد مطالعه 16000 کلمه از 503 شهر ایران می‌باشد که "ایران شهر" نام دارد. در این مرحله کلمات دست‌نوشته در 62 خوشه قرار می‌گیرند. در مرحله شناسایی کلمه ورودی، با انتخاب 5 خوشه نزدیک به کلمه دست‌نوشته مورد آزمون با دقت 94% حدود 77% از کلمات مورد بررسی کاهش خواهد یافت.

در مرحله دوم بازنشاسی، یکی از کلمات کاندید بدست آمده از مرحله اول می‌بایست بعنوان کلمه مورد آزمون تشخیص داده شود. در این مرحله از ویژگی هیستوگرام گرادیان روشنایی محلی استفاده می‌کنیم. برای این منظور گرادیان تصویر کلمه ورودی بلوک‌بندی می‌شود. در این پایان‌نامه دو روش بلوک‌بندی تطبیقی برای بهبود عملکرد بازنشاسی پیشنهاد می‌گردد. در روش اول اندازه بلوک‌ها بر اساس توزیع پیکسل‌های سیاه (قلم) تنظیم می‌شوند و در روش دوم اجزاء اصلی کلمات دست‌نوشته به طور جداگانه بلوک‌بندی و سپس با هم ترکیب می‌شوند. بردارهای ویژگی مبتنی بر گرادیان کلمه

ورودی با بردارهای ویژگی حاصل از کلمات کاندید در فرهنگ لغت در یک طبقه بند مقایسه می شوند. برای این منظور از کلاسه‌بندهای K نزدیکترین همسایه و SVM چند کلاسه استفاده می‌شود. نتایج بازشناسی کلمات دست نوشته پایگاه داده "ایران شهر" نشان می‌دهد که مرحله کاهش کاندید ها در فرهنگ لغت باعث افزایش نسبی دقت و سرعت می‌شود. این به دلیل حذف کلمات نامتشابه در بازشناسی کلمه دست نوشته مورد بررسی است. همچنین روش پیشنهادی بلوک بندی تطبیقی در استخراج ویژگی هیستوگرام گرادیان محلی باعث بهبود 13 درصدی دقت سیستم بازشناسی می‌شود.

کلمات کلیدی: بازشناسی کلمه دست نوشته، کاهش فرهنگ لغت، الگوریتم خوشه بندی ISOCLUS، الگوریتم خوشه‌بندی سلسله مراتبی، الگوریتم DTW، هیستوگرام گرادیان محلی، کلاسه بند k نزدیکترین همسایه، کلاسه بند SVM چند کلاسه

فهرست مطالب

| صفحه | عنوان |
|---------|---|
| ث..... | فهرست شکل ها |
| خ..... | فهرست جدول ها |
| ۲..... | فصل اول: مقدمه |
| 2..... | 1-1 بازشناسی دست نوشته |
| 4..... | 2-1 سیستم بازشناسی کلمات دست نوشته برون خط |
| 8..... | 3-1 مشخصات زبان فارسی |
| 11..... | 4-1 هدف پایاننامه |
| 11..... | 5-1 ساختار پایان نامه |
| ۱۴..... | فصل دوم: مروری بر فعالیتهای انجام شده در زمینه بازشناسی کلمه دستنوشته |
| 14..... | 1-2 روشهای مبتنی بر ویژگیهای ساختاری |
| 18..... | 2-2 روشهای مبتنی بر ویژگیهای آماری |
| 23..... | 3-2 روشهای مبتنی بر ویژگیهای محلی |
| 32..... | 4-2 نتیجه گیری |
| ۳۵..... | فصل سوم: مباحث نظری |
| 35..... | 1-3 استخراج ویژگی |
| 35..... | 1-1-3 ویژگی های مبتنی بر شکل |
| 35..... | 2-1-3 روشهای توصیف شکل |
| 38..... | 3-1-3 شکل کلی کلمات |
| 39..... | 1-3-1-3 پروفایل بالا و پایین |
| 39..... | 2-3-1-3 پروفایل پروژکشن عمودی |
| 39..... | 3-3-1-3 پروفایل گذر سیاه / سفید |
| 41..... | 4-3-1-3 ویژگی هیستوگرام گرادیان روشنایی محلی |
| 43..... | 2-3 روش های خوشه بندی |
| 43..... | 1-2-3 تعریف خوشه بندی |
| 46..... | 2-2-3 الگوریتم های خوشه بندی |

| | |
|---------|--|
| 46..... | 1-2-2-3 الگوریتم های ترتیبی |
| 47..... | 2-2-2-3 الگوریتم های سلسله مراتبی |
| 49..... | 3-2-2-3 الگوریتم های خوشه بندی افزای |
| 52..... | 4-2-2-3 الگوریتم های مبتنی بر تراکم |
| 52..... | 5-2-2-3 الگوریتم های مبتنی بر شبکه |
| 53..... | 3-2-3 مقایسه الگوریتم های خوشه بندی |
| 55..... | 3-3 معیار فاصله |
| 55..... | 1-3-3 معیار فاصله مینکوسکی |
| 55..... | 2-3-3 معیار فاصله زاویه ای |
| 55..... | 3-3-3 الگوریتم DTW |
| 57..... | 4-3 ماشین بردار پشتیبان چند کلاسه |
| 58..... | 1-4-3 استراتژی یکی در مقابل همه |
| 58..... | 2-4-3 استراتژی یکی در مقابل یکی |
| 59..... | 3-4-3 BSVM |
| ۶۲..... | فصل چهارم: پیش پردازش |
| 62..... | 1-4 مقدمه |
| 63..... | 2-4 باینری کردن |
| 64..... | 1-2-4 روش پیشنهادی توسط Otsu |
| 66..... | 2-2-4 روش پیشنهادی توسط Manmatha و Wu |
| 67..... | 3-4 هموار سازی و حذف نویز |
| 68..... | 4-4 اصلاح انحراف خط اصلی |
| 69..... | 5-4 اصلاح کجی |
| 69..... | 4-5-1 تبدیل shear افقی |
| 70..... | 4-5-2 الگوریتم اصلاح کجی بر اساس پروفایل پروژکشن عمودی |
| 72..... | 4-5-3 الگوریتم اصلاح کجی بر اساس هیستوگرام زاویه |
| 75..... | 4-6 آنالیز زیر کلمات |
| 76..... | 4-6-1 الگوریتم کشیدن |
| 77..... | 4-7 نازک سازی |

| | |
|----------|--|
| 78..... | 1-7-4 الگوریتم پیشنهادی توسط Haung |
| 80..... | 2-7-4 الگوریتم پیشنهادی توسط Hilditch و همکارانش |
| 81..... | 8-4 نرمال سازی ابعاد |
| ۸۵..... | فصل پنجم: الگوریتم پیشنهادی و نتایج ارزیابی آن |
| 85..... | 1-5 مقدمه |
| 85..... | 2-5 مرحله اول: کاهش فرهنگ لغت |
| 89..... | 1-2-5 الگوریتم ISOCLUS |
| 91..... | 2-2-5 الگوریتم سلسله مراتبی |
| 93..... | 3-2-5 معیار عملکرد الگوریتم کاهش فرهنگ لغت |
| 94..... | 4-2-5 نتایج مرحله کاهش فرهنگ لغت |
| 98..... | 3-5 مرحله دوم: بازشناسی کلمه آزمون ورودی |
| 100..... | 1-3-5 روش پیشنهادی اول |
| 100..... | 2-3-5 روش پیشنهادی دوم |
| 106..... | 3-3-5 نتایج مرحله بازشناسی |
| ۱۱۵..... | فصل هفتم: نتیجه گیری و پیشنهادات |

فهرست شکل ها

| عنوان | صفحه |
|--|------|
| شکل 1-1 . یک مدل کلی برای سیستم های بازشناسی دست نوشته برون خط | ۶ |
| شکل 2-1 . زیر کلماتی که قطعه بندی آنها مشکل است [37]. | ۶ |
| شکل 1-2 . برخی از ویژگی های ساختاری در کلمات فارسی | ۱۴ |
| شکل 2-2 . روش هیستوگرام برای استخراج نواحی مختلف مانند پایین رونده ها و بالارونده ها [11] | ۱۶ |
| شکل 3-2 . تصویر نرمال شده و پروفایل بالایی و پروفایل پروژکشن تصویر [17] | ۱۹ |
| شکل 4-2 . الگوریتم ارائه شده در تحقیق [45] | ۲۰ |
| شکل 5-2 . الگوریتم ارائه شده در تحقیق [10] | ۲۱ |
| شکل 6-2 . سه تصویر کلمه (A), (B), (C) و در زیر هر کدام تصویر قطبی نرمالیزه اش [6] | ۲۲ |
| شکل 7-2 . الگوریتم روش [20] | ۲۴ |
| شکل 8-2 . نحوه نوشتن نقاط در دست نوشته ها [9] | ۲۵ |
| شکل 9-2 . نحوه استخراج ویژگی انتقال در تحقیق [9] | ۲۵ |
| شکل 10-2 . دو نوع ترتیب زیگزاگ برای استخراج ضرایب DCT [8] | ۲۷ |
| شکل 11-2 . نمای کلی سیستم پیشنهاد شده در [7] | ۲۹ |
| شکل 12-2 . نحوه بلوک بندی تصاویر در [7] | ۲۹ |
| شکل 13-2 . نحوه لغزش پنجره فرم بندی شده در امتداد تصویر [15] | ۳۰ |
| شکل 14-2 . نحوه استخراج ویژگی خمیدگی در [15] | ۳۱ |
| شکل 1-3 . طبقه بندی روش های نمایش و توصیف شکل گرفته شده از [23] | ۳۷ |
| شکل 2-3 . نحوه تطبیق کلمه ورودی با کلمات واژه نامه و تولید یک رتبه بندی از کلمات مشابه | ۳۸ |
| شکل 3-3 . تصویر کلمه دست نوشته، پروفایل پایین، پروفایل بالا، پروفایل پروژکشن عمودی، گذر سیاه به سفید | ۴۰ |
| شکل 4-3 . عملگر های سو بل | ۴۱ |
| شکل 5-3 . الف- تصویر ورودی ب- اندازه گرادیان پ- جهت گرادیان | ۴۲ |
| شکل 6-3 . کوانتیزه کردن با تجزیه بردار جهت گرادیان | ۴۲ |

- شکل 3-7. مراحل فرآیند خوشه بندی گرفته شده از [46]..... ۴۵
- شکل 3-8. نتیجه یک خوشه بندی بزرگ دو خوشه، و خوشه بندی کوچک چهار خوشه شده است..... ۴۶
- شکل 3-9. دندروگرام یک الگوریتم سلسله مراتبی جمع کننده و افراز های تشکیل شده از آن..... ۴۹
- شکل 3-10. محاسبه فاصله انطباق دو مجموعه زمانی..... ۵۶
- شکل 3-11. محدودیت های استفاده شده برای پیاده سازی الگوریتم DTW..... ۵۷
- شکل 4-1. به کار بردن الگوریتم OTSU بر روی تصویر دست نوشته..... ۶۶
- شکل 4-2. مقایسه دو روش OTSU و WU در یک تصویر با پیش زمینه پیچیده..... ۶۷
- شکل 4-3. نمونه ای از انحراف غیر یکنواخت یک متن دست نوشته..... ۶۸
- شکل 4-4. مثال هایی از کجی یکنواخت و غیر یکنواخت..... ۶۹
- شکل 4-5. تبدیل SHEAR..... ۷۰
- شکل 4-6. هیستوگرام عمودی تصاویر کج شده با زاویه های مختلف..... ۷۱
- شکل 4-7. محاسبه زاویه کجی با استفاده از معیار SD..... ۷۱
- شکل 4-8. مراحل اجرای الگوریتم اصلاح کجی تصویر بر اساس هیستوگرام زاویه..... ۷۳
- شکل 4-9. به کار بردن قوانین برای هموار کردن کردن لبه های تصویر اصلاح شده توسط تبدیل SHEAR..... ۷۵
- شکل 4-10. حذف فضای خالی بین زیر کلمات..... ۷۶
- شکل 4-11. کشیدن بعضی حروف باعث ایجاد خطا در پیدا کردن خط اصلی می شود..... ۷۷
- شکل 4-12. مراحل اجرای الگوریتم کشیدن..... ۷۷
- شکل 4-13. همسایه هشت تایی پیکسل p ۷۸
- شکل 4-14. الف- تصویر کلمه. ب- اسکلت کلمه با الگوریتم HILDITCH پ- اسکلت کلمه با الگوریتم HAUNG..... ۸۱
- شکل 4-15. الف- تصویر ورودی ب- تصویر نرمال در صفحه مربعی L..... ۸۲
- شکل 5-1. مراحل خوشه بندی تصاویر کلمات دست نوشته..... ۸۶
- شکل 5-2. (الف) یک نمونه از کلمه دست نوشته (ب) پروفایل پایین (پ) پروفایل بالا (ت) پروفایل پروژکشن عمودی (ث) پروفایل گذر سیاه به سفید..... ۸۸
- شکل 5-3. (الف) پروفایل بالا (ب) بردار ضرایب تقریب تبدیل موجک..... ۸۹
- شکل 5-4. نمودار دندروگرام کامل مربوط به یکی از آزمایشات انجام شده..... ۹۲

- شکل 5-5. تصاویر یک کلمه با دست نوشته های متفاوت ۹۲
- شکل 5-6. نمودار تعداد نمونه ها در هر خوشه ۹۵
- شکل 5-7. نمودار تعداد کلمه ها در هر خوشه ۹۵
- شکل 5-8. نمودار تعداد خوشه هایی که هر کلمه در آن قرار گرفته است ۹۶
- شکل 5-9. مقایسه عملکرد الگوریتم DTW با معیار فاصله اقلیدسی و بلوک شهری ۹۸
- شکل 5-10. (الف) و (ب) تصاویر اندازه گرادیان نمونه های یک کلاس و (پ) و (ت) تصاویر بلوک بندی شده با سایز بلوک های ثابت ۹۹
- شکل 5-11. تصویر گرادیان بلوک بندی شده با بلوک های تطبیقی برای نمونه هایی از یک کلمه ۱۰۰
- شکل 5-12. تصویر گرادیان اجزاء اصلی بلوک بندی شده کلمه دست نوشته ۱۰۲
- شکل 5-13. الف- میزان پراکندگی درون کلاس S_w و ب- بین کلاس ها S_b و پ- نسبت S_b به S_w ۱۰۴
- شکل 5-14. برخی از کلماتی که در جدا سازی مولفه های اصلی خطا داشته اند ۱۰۵
- شکل 5-15. تعدادی از تصاویر کلمات دست نوشته که باعث ایجاد خطا شده اند ۱۰۹
- شکل 5-16. تعدادی از تصاویر کلمات دست نوشته خوشه 22 ۱۱۰
- شکل 5-17. تعدادی از تصاویر کلمات دست نوشته خوشه 10 ۱۱۱
- شکل 5-18. تعدادی از تصاویر کلمات دست نوشته خوشه 15 ۱۱۲
- شکل 5-19. تعدادی از تصاویر کلمات دست نوشته خوشه 43 ۱۱۳

فهرست جدول ها

| صفحه | عنوان |
|------|---|
| ۹ | جدول 1-1 . شکل های مختلف حروف الفبای فارسی با توجه به محل قرار گرفتنش در کلمه |
| ۱۵ | جدول 1-2 . کد حروف الفبای انگلیسی |
| ۵۴ | جدول 1-3 . مقایسه الگوریتم های خوشه بندی |
| ۸۲ | جدول 1-4 . تابع های نگاشت برای نسبت ابعاد |
| ۸۲ | جدول 2-4 . نگاشت مختصات روشهای نرمال سازی مختلف |
| ۹۷ | جدول 1-5 . دقت، درجه کاهش و درجه تأثیر برای مقادیر مختلفی از N |
| | جدول 2-5 . عملکرد الگوریتم کاهش دهنده فرهنگ لغت پیشنهادی برای ترکیب های |
| ۹۷ | مختلفی از ویژگی ها |
| ۱۰۶ | جدول 3-5 . دقت بازشناسی برای ویژگی های مختلف |
| ۱۰۷ | جدول 4-5 . دقت و سرعت بازشناسی برای روش های مختلف |
| | جدول 5-5 . مقایسه دقت و زمان بازشناسی الگوریتم دو مرحله ای پیشنهادی با SVM چند |
| ۱۰۸ | کلاسه |

فصل اول

مقدمه

فصل اول: مقدمه

1-1 بازشناسی دست نوشته¹

بازشناسی متون دست نوشته منجر به تبدیل اتوماتیک تصویر اسکن شده متون دست نوشته به متون ماشینی می شود. این یکی از کاربردهای مهم بازشناسی الگوهای تصویری است که مورد توجه بسیاری از محققین بخصوص در زبان های لاتین و چینی قرار گرفته است. بازشناسی یا پردازش اسناد، خواندن فرم های مخصوص (فرم های ثبت نام، کوپن ها، اسناد و اوراق بهادار و...)، خواندن و مرتب کردن اتوماتیک مرسولات پستی، خواندن خودکار چک های بانکی، بازیابی و مرتب کردن تصاویر متون دست نوشته تاریخی اسکن شده که بعضی از آنها کیفیت پایینی دارند، از کاربردهای مهم بازشناسی دست نوشته هستند.

بازشناسی دست نوشته می تواند در سطح کلمه و یا در سطح حروف انجام شود که برای این کار متون دست نوشته را به کلمه و کلمات را به حروف تشکیل دهنده اش تجزیه می کنند.

سیستم های بازشناسی کلمه دست نوشته موجود به خصوص برای زبان های لاتین و چینی قادرند با دقتی در حدود 99%-90%، البته با یک سری محدودیت های آزمایشی، کلمات دست نوشته را شناسایی کنند. یکی از رایج ترین محدودیت های سیستم های موجود این است که بیشتر این سیستم ها دارای یک فرهنگ لغت محدود هستند و تنها قادرند کلماتی را شناسایی کنند که در این فرهنگ لغت وجود داشته باشند. این فرهنگ لغت² می تواند شامل 30 تا 1000 کلمه باشد [1].

هرچند که تعریف دقیقی برای بزرگی و یا کوچک بودن فرهنگ لغت وجود ندارد اما تعریف های زیر متداول است:

- فرهنگ لغت کوچک: دهها کلمه

¹ Handwritten Recognition
² lexicon

- فرهنگ لغت متوسط: صدها کلمه

- فرهنگ لغت بزرگ: هزارها کلمه

- فرهنگ لغت خیلی بزرگی: دهها هزار کلمه

هرچه تعداد کلمات موجود در فرهنگ لغت افزایش یابد، احتمال وجود کلمات مشابه در فرهنگ لغت بالا می‌رود و در نتیجه کار بازشناسی مشکل‌تر می‌شود. پیچیدگی محاسباتی سیستم بازشناسی نیز به اندازه فرهنگ لغت وابسته است و هرچه اندازه فرهنگ لغت افزایش یابد، با همان نسبت پیچیدگی محاسباتی بالا می‌رود.

در سیستم‌های شناسایی کلمه که در ارتباط با یک فرهنگ لغت بزرگ می‌باشند برای کاهش پیچیدگی محاسباتی و افزایش دقت و سرعت، تعداد کلمات مورد بررسی در فرآیند بازشناسی کلمه ورودی را کاهش می‌دهند که به این مکانیزم کاهش فرهنگ لغت^۱ یا هرس کردن^۲ گفته می‌شود.

البته این مرحله در برخی از سیستم‌های بازشناسی کلمه دست نوشته برون خط وجود دارد.

بعضی از سیستم‌ها مستقل از فرهنگ لغت می‌باشند یعنی قادرند هر کلمه که وارد سیستم شود را شناسایی کنند بدون اینکه متکی به فرهنگ لغت خاصی باشند ولی دقت این سیستم‌ها خیلی پایین‌تر از سیستم‌های با فرهنگ لغت محدود است. البته بعضی از روش‌ها که سعی در افزایش دقت نموده‌اند توجه‌ای به سرعت بازشناسی و پیچیدگی محاسبات نداشته‌اند.

به طور کلی بازشناسی کلمات دست نوشته^۳ را می‌توان به دو گروه تقسیم کرد:

1- بازشناسی کلمه دست نوشته برون خط^۴ 2- بازشناسی کلمه دست نوشته بر خط^۵

در سیستم‌های گروه اول فقط اطلاعات مکانی از تصاویر کلمات دست‌نوشته در دسترس است. اما در گروه دوم علاوه بر اطلاعات مکانی، اطلاعات زمانی و دینامیک حرکتی فرآیند نوشتن که از طریق قلم

^۱ Lexicon reduction

^۲ pruning

^۳ Handwritten word recognition(HWR)

^۴ Off line handwritten word recognition

^۵ On line handwritten word recognition

مغناطیسی به سیستم منتقل می‌گردد، نیز وجود دارد. این اطلاعات اضافی باعث شده که تحت شرایط آزمایشی مشابه سیستم‌های بر خط دقیق‌تر از سیستم‌های برون خط باشند. سیستم‌های بر خط با داشتن منبع اطلاعاتی اضافی می‌توانند از مدل‌های ساده‌تر که در بازشناسی گفتار نیز از آنها استفاده می‌شود، بهره ببرند. بیشتر این سیستم‌ها می‌توانند با فرهنگ لغت بزرگتری نسبت به سیستم‌های برون خط کار کنند [1].

این پایان نامه بر روی سیستم‌های بازشناسی کلمات دست نوشته برون خط متمرکز شده است.

2-1 سیستم بازشناسی کلمات دست نوشته برون خط

تکنیک‌های گسترده‌ای برای سیستم‌های بازشناسی کلمات دست نوشته برون خط ارائه شده است که همه آنها از یک مدل کلی پیروی می‌کنند که در شکل 1-1 نشان داده شده است.

این مدل در ابتدا تصویر کلمه دست نوشته را به عنوان ورودی دریافت می‌کند. برای تبدیل این تصویر به اطلاعات با معنی برای سیستم لازم است یک سری مراحل انجام شود. اولین مرحله پیش پردازش است. در این مرحله برای کاهش تغییرات نامناسب موجود در تصاویر کلمات یک سری عملیات شامل باینری کردن، اصلاح کجی و شیب در تصویر، حذف نویز، هموار سازی، نرمال کردن اندازه تصویر و... انجام می‌شود.

سیستم‌های بازشناسی دست نوشته برون خط را می‌توان به دو دسته تقسیم کرد:

1- سیستم‌های مبتنی بر قطعه بندی¹ 2- سیستم‌های مستقل از قطعه بندی²

روش‌های مبتنی بر قطعه بندی که معمولاً روش‌های آنالیزی³ گفته می‌شوند خود به دو گروه تقسیم می‌شوند. گروه اول روش‌هایی هستند که کلمات را به حروف تشکیل دهنده شان تجزیه می‌کنند و هر حرف را به طور مجزا، با استفاده از ویژگی‌های استخراج شده از آن بازشناسی می‌کنند. در این

¹ Segmentation based systems

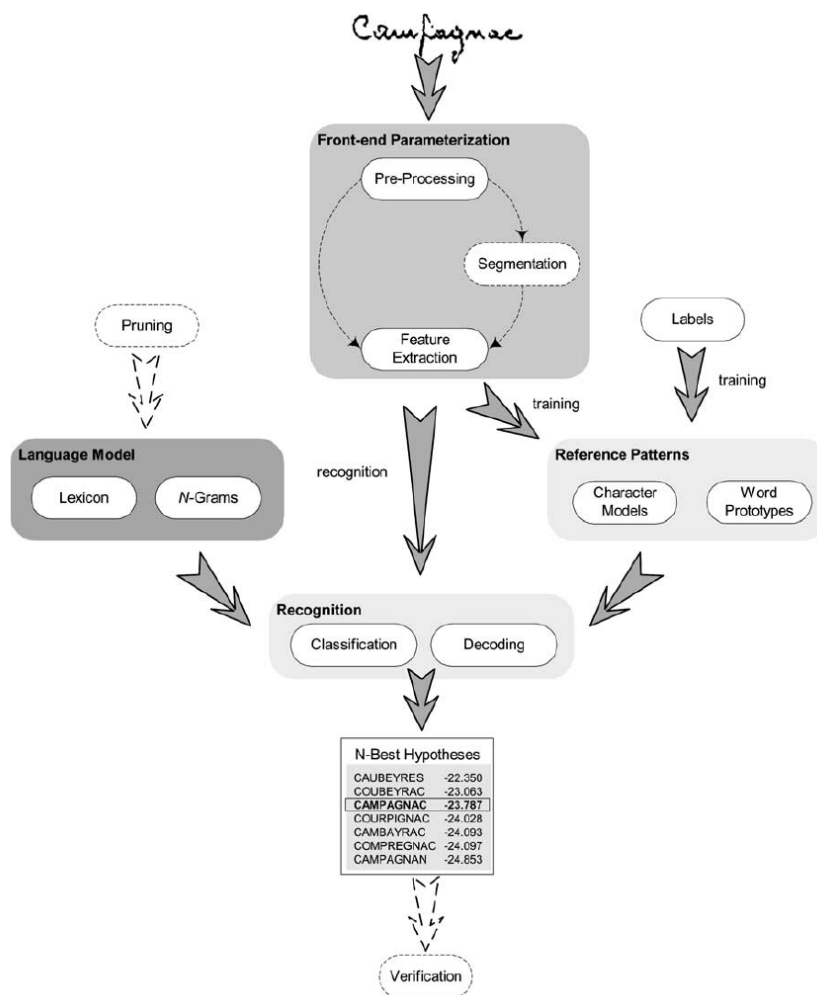
² Segmentation free systems

³ Analytical Approach

روش‌ها بازشناسی کلمه با استفاده از حروف شناخته شده و پس پردازش‌هایی مبنی بر فرهنگ لغت انجام می‌شود.

گروه دوم روش‌هایی هستند که کلمه را به واحدهای کوچکتر می‌شکنند و از هر واحد بدست آمده ویژگی استخراج می‌کنند. سپس با کنار هم قرار دادن این ویژگی‌ها یک توصیفگر کلمه استخراج می‌شود و به کمک آن در فرهنگ لغت بهترین گزینه را انتخاب می‌کنند [3].

در روش‌های بازشناسی کلمات دست نوشته فارسی مبتنی بر قطعه بندی، علاوه بر مشکلاتی مانند وجود نقاط، علائم و تنوع نحوه نوشتن، مشکل جداسازی بعضی حروف نیز وجود دارد. مثلاً در شکل 1-2 جدا کردن حروف نشان داده شده بسیار مشکل است.



شکل 1-1. یک مدل کلی برای سیستم های بازشناسی دست نوشته برون خط [1]

مناسب نبودن کیفیت متن نوشته شده، پایین بودن درجه روشنایی و کجی تصویر متن دست نوشته نیز مشکلاتی هستند که به سختی کار می افزایند.



شکل 2-1. زیرکلماتی که قطعه بندی آنها مشکل است [37].

در روش‌های بازشناسی مستقل از قطعه‌بندی که معمولاً روش‌های کلی^۱ نامیده می‌شوند ویژگی‌ها مستقیماً از تصویر کلمه دست‌نوشته بدون هیچ‌گونه قطعه‌بندی استخراج می‌شوند و با استفاده از یک فرهنگ لغت محدود، کلمه ورودی شناسایی می‌شود. سیستم‌های دسته دوم به دلیل اینکه از مشکلات ناشی از قطعه‌بندی رها شده اند ساده‌تر و دقیق‌تر هستند.

در مرحله آموزش سیستم بازشناسی دست‌نوشته برون خط تعدادی از نمونه‌های هر کلاس برای ایجاد یک نماینده یا الگوی آن کلاس، استفاده می‌شوند. الگوی نتیجه را الگوی مرجع یا نمونه اصلی کلاس می‌نامند.

در بعضی از سیستم‌ها برای هر کلمه در فرهنگ لغت یک کلاس در نظر گرفته می‌شود در حالیکه در برخی دیگر یک کلاس مطابق با یک زیر کلمه تعریف می‌گردد. در سیستم‌های نوع دوم از تکنیک‌های استاندارد الحاق زیر کلمات برای ساختن مدل یک کلمه در طی بازشناسی استفاده می‌شود.

در مرحله شناسایی، ویژگی‌های تصویر کلمه ورودی استخراج شده با ویژگی‌های نماینده کلاس‌ها در فرهنگ لغت مقایسه می‌شود و شباهت کلمه با نماینده هر کلاس محاسبه می‌گردد. از امتیاز شباهت برای تصمیم‌گیری اینکه نماینده کدام کلاس بهترین تطبیق را با نمونه ورودی دارد، استفاده می‌شود. عمل مقایسه در بازشناسی توسط یک کلاسه‌بند انجام می‌شود. در این رابطه روش‌های گوناگونی پیشنهاد شده از جمله: شبکه عصبی (NN) [52]، k نزدیکترین همسایه (KNN) [42]، مدل مخفی مارکوف (HMM)² [۱۶،۱۴،۴۰]، ماشین بردار پشتیبان (SVM) و ترکیبی از تکنیک‌ها [31، 35].

معمولاً در فرآیند بازشناسی یک لیستی از بهترین فرضیه‌ها برای کلمه مورد آزمون تولید می‌شود. مرحله بعدی در این سیستم‌ها پس پردازش یا بازبینی مجدد است که برای تولید یک لیست قابل اطمینان‌تر، یک سری تکنیک‌های مردود سازی استاندارد بر روی لیست اولیه انجام می‌شود.

اکثر روش‌های بازشناسی که برای زبان‌های لاتین و چینی دقت بالایی دارند، برای زبان‌های فارسی و عربی به دلیل خصوصیات خاص این زبان‌ها از دقت خوبی برخوردار نیستند. کارهای انجام شده در

^۱ Global Approach

^۲ Hidden Markov Model

زمینه بازشناسی متون دست نوشته زبان‌های فارسی و عربی به دلیل پیچیدگی‌های خاصی که این زبان‌ها دارند هنوز به در صد قابل قبولی برای تجاری شدن نرسیده‌اند.

3-1 مشخصات زبان فارسی

زبان فارسی یکی از گسترده‌ترین زبان دنیا است که بیش از 150 میلیون نفر از این زبان برای نوشتن استفاده می‌کنند. یکی از مهمترین مشخصه زبان فارسی ماهیت پیوستگی آن هم در متون تاییپی و هم در متون دست نوشته است که این مشخصه باعث شده دقت بازشناسی حتی در متون تاییپی کمتر از متون تاییپی لاتین باشد. الفبای فارسی دارای 32 حرف است که هرکدام از آنها بسته به موقعیتش در کلمه 2 تا 4 شکل دارد. جدول 1-1 شکل‌های مختلف حروف الفبای زبان فارسی را نشان می‌داد که وابسته به مکان حرف در کلمه (ابتدا، وسط، انتها، منفرد) است. مثلاً حرف - غ - که غین نامیده می‌شود دارای چهار شکل است. شکل آن در ابتدای کلمه به صورت - غ - و در وسط کلمه به صورت - غ - و در انتها به صورت - غ - و می‌باشد.

جدول 1-1. شکل های مختلف حروف الفبای فارسی با توجه به محل قرار گرفتنش در کلمه

| Character | Isolated | Initial | Middle | Final | Transliteration |
|-----------|----------|---------|--------|-------|-----------------|
| Alef | آ | آ | ا | ا | a |
| Beh | ب | ب | ب | ب | b |
| Peh | پ | پ | پ | پ | p |
| The | ت | ت | ت | ت | t |
| Theh | ث | ث | ث | ث | Th |
| Jeem | ج | ج | ج | ج | j |
| Cheh | چ | چ | چ | چ | ch |
| Heh | ح | ح | ح | ح | h |
| Kheh | خ | خ | خ | خ | kh |
| Dal | د | د | د | د | d |
| Thal | ذ | ذ | ذ | ذ | th |
| Reh | ر | ر | ر | ر | r |
| Zeh | ز | ز | ز | ز | z |
| Zheh | ژ | ژ | ژ | ژ | zh |
| Seen | س | س | س | س | s |
| Sheen | ش | ش | ش | ش | sh |
| Sad | ص | ص | ص | ص | s |
| Zad | ض | ض | ض | ض | th |
| Tah | ط | ط | ط | ط | t |
| Zah | ظ | ظ | ظ | ظ | z |
| Ain | ع | ع | ع | ع | a |
| Ghain | غ | غ | غ | غ | gh |
| Feh | ف | ف | ف | ف | f |
| Ghaf | ق | ق | ق | ق | gh |
| Kaf | ک | ک | ک | ک | k |
| Gaf | گ | گ | گ | گ | g |
| Lam | ل | ل | ل | ل | l |
| Meem | م | م | م | م | m |
| Noon | ن | ن | ن | ن | n |
| Waw | و | و | و | و | v |
| Heh | ه | ه | ه | ه | h |
| Yeh | ی | ی | ی | ی | y |

تعدادی حروف در الفبای فارسی وجود دارند که از سمت چپ قابل اتصال نیست و تنها دارای دو شکل متفاوت می باشند. مانند حرف - د - که دال نامیده می شود در ابتدای کلمه به صورت - د - و در وسط و انتهای کلمه به صورت - د - می باشند. این حروف اگر در ابتدا و وسط کلمه قرار بگیرند، باعث

جدا شدن کلمه و تشکیل زیر کلمات می‌شود. این مشخصه زبان فارسی باعث شده که مرحله قطعه بندی در سیستم‌های بازشناسی مبتنی بر قطعه بندی با مشکل روبه رو شود. به این دلیل که فاصله گذاری مناسب بین کلمات و زیر کلمات در بیشتر مواقع رعایت نمی‌شود. این حروف شامل {ر، ز، ژ، د، ذ، و، ا} می‌باشد. مثلاً در کلمه "مدرسه" وجود دو حرف "د" و "ر" در کلمه باعث شده که کلمه "مدرسه" شامل سه زیر کلمه شود.

در زبان عربی و فارسی کلماتی وجود دارد که شکل یکسانی دارند اما تلفظ و معنی متفاوتی دارند، برای نشان دادن تلفظ و معنی مورد نظر از یک سری علائم استفاده می‌شود. مثلاً کلمه فارسی "مرد" دارای دو تلفظ است که معنی‌های متفاوتی دارند: "مُرد" به معنی از دنیا رفتن و "مَرَد" به معنی آقا است. البته در بیشتر مواقعی که در متن آورده می‌شود معنی آن با توجه به متن مشخص می‌شود و نیازی به گذاشتن علائم بر روی حروف نیست. از دیدگاه بازشناسی کلمه بهتر است این علائم گذاشته نشود به دلیل اینکه دقت بازشناسی را کاهش می‌دهد.

یکی دیگر از مشخصات زبان فارسی وجود نقطه در بعضی حروف است. 18 حرف از 32 حرف زبان فارسی شامل یک یا دو یا سه نقطه در بالا یا پایین خط اصلی می‌باشند:

- ده حرف فارسی دارای یک نقطه می‌باشند {ب، ج، خ، ذ، ز، ض، ظ، غ، ف، ن}
- سه حرف شامل دو نقطه می‌باشند {ی، ق، ت}
- پنج حرف شامل سه نقطه می‌باشند {پ، چ، ژ، ش، ث}

تعدادی حروف در زبان فارسی دارای یک سری مکمل‌ها می‌باشند مانند سرکش حروف {ک و گ} یا آکلاد حرف {آ}.

وجود نقاط، مکمل‌ها و علائم در کلمات باعث می‌شود بازشناسی بخصوص در کلمات دست نوشته مشکل شود دلیل آن را می‌توان اینگونه بیان کرد که در متون دست‌نوشته اندازه و نحوه نوشتن نقاط، مکمل‌ها و علائم بسیار متغیر است و ممکن است بعضی از آنها آنقدر کوچک باشد که در مرحله پیش پردازش حذف شوند یا آنقدر بزرگ باشند که با بعضی حروف اشتباه گرفته شوند.

4-1 هدف پایان نامه

در این پایان نامه روشی برای سیستم بازشناسی کلمه دست نوشته فارسی ارائه شده است. این روش شامل دو مرحله اصلی است. در مرحله اول برای تسریع بازشناسی و کاهش محاسبات، روشی برای کاهش کلمات مورد بررسی در بازشناسی کلمه مورد آزمون ارائه شده است. در این مرحله از ویژگی‌هایی که شکل کلی کلمه را توصیف می‌کنند، استفاده شده است. برای این منظور کلمات در فرهنگ- لغت توسط الگوریتم‌های خوشه بندی، گروه بندی می‌شوند. برای هر خوشه تشکیل شده نماینده‌ای به عنوان مدخل اعضای آن خوشه در نظر گرفته می‌شود.

در مرحله دوم کلمه دست نوشته مورد آزمون از بین کلمات کاندید بازشناسی می‌شود. تعدادی از خوشه‌های نزدیک به کلمه دست نوشته ورودی توسط الگوریتم DTW¹ انتخاب می‌شود. کلمات موجود در این خوشه در لیست کاندید قرار می‌گیرند و از آنها ویژگی‌هایی که جزئیات شکل کلمه را بهتر بیان می‌کنند، استخراج می‌شود. در این مرحله از کلاسه بند k نزدیک ترین همسایه و SVM چند کلاسه برای انتخاب بهترین کاندید برای کلمه دست نوشته ورودی استفاده شده است.

در این پایان نامه از پایگاه داده "ایران شهر" که شامل 16000 تصویر کلمه دست نوشته از 503 نام شهرهای ایران است، برای ارزیابی روش پیشنهادی استفاده شده است. برای هر شهر حداقل 25 تصویر دست نوشته وجود دارد که توسط افراد مختلف نوشته شده است. 13500 تصویر برای فاز آموزش الگوریتم و 2500 تصویر برای فاز بازشناسی به صورت تصادفی انتخاب شده است.

5-1 ساختار پایان نامه

در فصل دوم کارهای انجام شده در زمینه بازشناسی کلمه دست نوشته برون خط مبتنی بر شکل کلی کلمه مرور شده است. فصل سوم به کلیات مربوط به روش‌های استخراج ویژگی، خوشه بندی، معیار-های فاصله و کلاسه بند SVM چند کلاسه اختصاص دارد. در این فصل هر یک از موارد ذکر شده به اختصار بررسی شده اند.

¹ Dynamic Time warping

فصل چهارم تکنیک‌های به کار رفته در مرحله پیش پردازش بیان شده است. در فصل پنجم مراحل الگوریتم پیشنهادی و نتایج ارزیابی آن ارائه شده است و فصل ششم به نتیجه گیری و ارائه پیشنهادها اختصاص دارد.

مروری بر فعالیت‌های انجام شده در زمینه
بازشناسی کلمه دست‌نوشته

فصل دوم: مروری بر فعالیت‌های انجام شده در زمینه بازشناسی کلمه دست‌نوشته

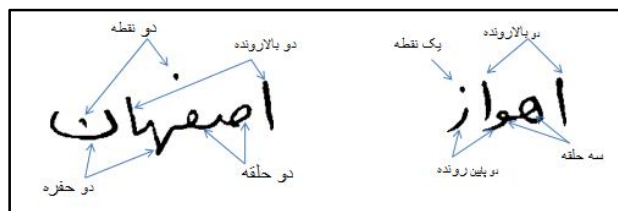
در این فصل تحقیقات انجام شده در زمینه بازشناسی کلمه مبتنی بر شکل کلی کلمات در زبان‌های لاتین، عربی و فارسی بررسی می‌شوند.

براساس ویژگی استفاده شده برای توصیف شکل کلی کلمه، روش‌ها و تکنیک‌های موجود را می‌توان به طور کلی به سه دسته تقسیم کرد: 1- روش‌های مبتنی بر ویژگی‌های ساختاری یا توپولوژیکی 2- روش‌های مبتنی بر ویژگی‌های آماری 3- روش‌های مبتنی بر ویژگی‌های محلی یا ناحیه‌ای البته بعضی روش‌ها از ترکیب این سه دسته بوجود آمده‌اند.

1-2 روش‌های مبتنی بر ویژگی‌های ساختاری

این ویژگی‌ها ساختار کلی و توپولوژی هر کلمه را توصیف می‌کنند. تعداد عناصر متصل، نقاط، حفره‌ها، بالارونده‌ها و پایین رونده‌ها و ... مثال‌هایی از ویژگی‌های ساختاری هستند. شکل 1-2 برخی از ویژگی‌های ساختاری در کلمات فارسی را نشان می‌دهد.

این ویژگی‌ها به نحوه نوشتن وابسته نیستند و نسبت به نویز حساس و استخراج آنها بسیار مشکل است. در ادامه تعدادی از روش‌های این دسته بیان خواهد شد.



شکل 1-2. برخی از ویژگی‌های ساختاری در کلمات فارسی

در تحقیق [12] با تخصیص یک عدد به هر یک از بالارونده‌ها¹، پایین رونده‌ها²، حفره‌ها و حلقه‌های موجود در کلمه، کد توصیف‌کننده شکل یک کلمه لاتین بدست می‌آید. در این تحقیق از روش تولید فرض³ برای تعیین کلمات با شکل کلی یکسان در یک فرهنگ لغت کلمات استفاده شده است. برای یک کلمه ورودی، در مرحله تولید فرض، کلمات کاندید آن در فرهنگ لغت جستجو می‌شود. در مرحله آزمون فرض، کلمه‌ای که بیشترین شباهت را به ورودی دارد با آزمایش ویژگی‌ها بدست می‌آید. ویژگی‌های استفاده شده برای کد کردن شکل کلمات عبارت است از:

کد صفر - فضای محدود شده در ابتدا یا انتهای کلمه، مثلاً فضای خالی طرف راست حرف C که بصورت عمودی محدود شده است.

کد یک - تکه کوتاه عمودی مانند طرف چپ r و طرف راست b

کد دو - تکه بلند عمودی که بالاتر از بدنه کلمه باشد مانند قسمت بالارونده b

کد سه - تکه بلند عمودی که پایین تر از بدنه کلمه باشد مانند قسمت پایین رونده p

کد چهار - نقطه روی تکه کوتاه عمودی مانند i

کد پنج - نقطه روی تکه بلند عمودی مانند j

در جدول 1-2 کد حروف الفبای انگلیسی آمده است. مثلاً حرف b به دلیل داشتن یک تکه عمودی بلند و یک تکه عمودی کوتاه به صورت 21 کد شده است.

جدول 1-2. کد حروف الفبای انگلیسی

| | | | | | | | |
|---|----|---|-----|---|----|---|---|
| a | 01 | h | 21 | O | 11 | v | 0 |
| b | 21 | i | 4 | P | 31 | w | 0 |
| c | 10 | j | 05 | Q | 13 | X | 0 |
| d | 12 | k | 20 | R | 10 | Y | 0 |
| e | 10 | l | 2 | S | 0 | Z | 0 |
| f | 20 | m | 111 | T | 1 | | |
| g | 11 | n | 11 | U | 11 | | |

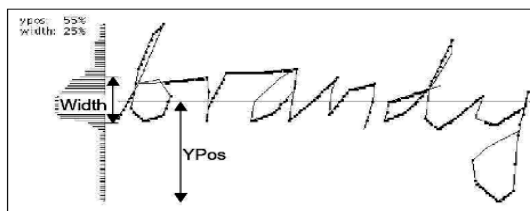
¹ Ascender

² Descender

³ Hypothesis generation

از این روش نمایش برای خوشه بندی فرهنگ لغت کلمات به زیرمجموعه هایی که اعضای هر زیرمجموعه توصیف دیداری یکسانی دارند، استفاده شده است. برای یک واژه نامه شامل 43264 کلمه، 28% از کلمات دارای نمایش یکسان هستند. برای یک مجموعه شامل 1500 تصویر کلمه، 85% کلمات درست خوشه بندی شده اند.

در روشی که توسط powalka و همکارش ارائه شده است [11] از ویژگی‌های ساختاری شکل کلمه برای بازشناسی کلمات دستنویس لاتین استفاده شده است. در این مقاله از اطلاعات وجود یا عدم وجود بالارونده‌ها و پایین‌رونده‌ها، مکان تقریبی آنها و ترتیب قرار گیری آنها برای توصیف شکل کلی کلمات دستنویس لاتین استفاده شده است. بالارونده‌ها و پایین‌رونده‌ها با استفاده از هیستوگرام افقی نقاط سیاه تصویر کلمه و بیشینه محلی تعداد نقاط سیاه تصویر کلمه استخراج می شوند. طول کلمه با تعداد دفعاتی که یک خط افقی فرضی در متراکم‌ترین قسمت، نوشته را قطع می کند، تخمین زده می شود. از روش هیستوگرام برای پیدا کردن نواحی مختلف استفاده شده است. چگالیترین قسمت هیستوگرام با استفاده از حد آستانه میانگین مقادیر آن بدست می آید. این ناحیه، ناحیه میانی است. اگر هیستوگرام افقی تصویر کلمه، در بالا یا پایین ناحیه میانی مقدار داشته باشد؛ بدین معنی است که تصویر کلمه بترتیب دارای نواحی بالا رونده یا پایین‌رونده است. نحوه انجام این روش در شکل 2-2 آمده است.



شکل 2-2. روش هیستوگرام برای استخراج نواحی مختلف مانند پایین رونده ها و بالارونده ها [11]

از روش بیشینه محلی برای تعیین محدوده نواحی استفاده شده است. در این روش، همانطور که در شکل 2-2 آمده است، با استخراج ماکزیمم‌های محلی در تصویر کلمه محل و محدوده نواحی مختلف بدست می آیند.

مقایسه شکل کلمه با لغات فرهنگ لغت در دو مرحله انجام شده است. ابتدا دنباله بالارونده ها و پایین‌رونده‌ها از شکل کلمه استخراج می‌شود. سپس این دنباله برای پیدا کردن لغات مشابه با آن با الگوهای کلمات فرهنگ لغت مقایسه می‌شود. برای یک مجموعه شامل 200 کلمه ، 80/1 % بازشناسی درست در 5 انتخاب اول گزارش شده است.

در تحقیق [33] سیستمی برای بازشناسی کلمات چاپی عربی با شبکه عصبی پیشنهاد شده است. بازشناسی کلمات در سیستم پیشنهادی به سه مرحله تقسیم می‌شود. در مرحله پیش پردازش، تصویر با یک روبشگر نوری¹ با درجه تفکیک 300 نقطه در اینچ تصویربرداری شده و تکه های جدای آن مشخص می‌شوند. سپس ویژگی‌های توپولوژیکی مانند تعداد زیر کلمات، تعداد بیشینه‌ها در هر زیرکلمه، تعداد و محل نقاط، علائم و ویژگی‌های هندسی دیگری از شکل کلی کلمه استخراج می‌شود. برای بازشناسی کلمات از یک شبکه عصبی سه لایه استفاده شده است. تعداد نوروها در ورودی شبکه 270 عدد گزارش شده است.

آقای حاجی و همکارانش در تحقیق [4] روشی برای بازشناسی کلمات دست نوشته فارسی ارائه داده‌اند. این روش براساس ویژگی‌های ساختاری و مدل مخفی مارکوف پیوسته می‌باشد. ویژگی‌های ساختاری استفاده شده در این تحقیق مستقل از خط اصلی کلمه و مقاوم در برابر چرخش و جابه جایی هستند و برای استخراج آنها اسکلت شکل کلمه به یک گراف تبدیل می‌شود و از هر گراف 10 ویژگی (مثل طول گراف، تعداد حلقه ها ، تعداد خمیدگی و...) استخراج می‌شود. بازشناسی بر اساس مدل‌های مخفی مارکوف پیوسته انجام می‌شود. نرخ بازشناسی این الگوریتم بر روی 100 کلمه از شهرهای ایران ، 82% گزارش شده است.

¹ Optical scanner

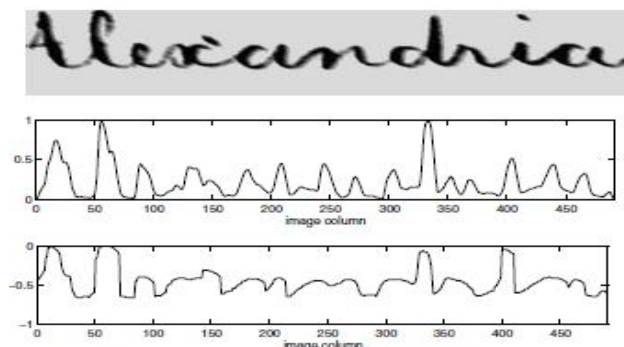
2-2 روش‌های مبتنی بر ویژگی‌های آماری

استخراج این ویژگی‌ها از تصویر ساده و نسبت به نویز مقاوم هستند. ضرایب تبدیل فوریه، گشتاورها و پروفایل‌ها از این نوع ویژگی هستند. در ادامه تحقیقاتی که از این نوع ویژگی استفاده کرده‌اند، آورده شده است.

در تحقیق [17] یک روشی برای بازشناسی متون دست نوشته تاریخی لاتین اسکن شده که دارای کیفیت پایینی می‌باشد ارائه شده است. در بازشناسی متون‌های تاریخی بیشتر از روش تطبیق کلمه استفاده می‌شود. البته این روش نیز به دلیل پایین بودن کیفیت تصاویر با مشکلات زیادی روبه رو می‌شود. در این تحقیق قبل از استخراج ویژگی‌ها از تصویر یک سری عملیات پیش پردازش برای بهتر شدن کیفیت تصاویر و کاهش اختلافات از قبیل حذف نویز، پیدا کردن خط اصلی، محاسبه کجی خط اصلی و اصلاح آن، نرمال کردن اندازه تصاویر و غیره انجام می‌شود. بعد از مرحله پیش پردازش از هر ستون تصویر دو ویژگی آماری: پروفایل بالا و پروفایل پروژکشن عمودی استخراج می‌شود. شکل 2-3 پروفایل بالایی و پروفایل پروژکشن تصویر کلمه را نشان می‌دهد که در رنج [0,1] نرمال شده است.

برای تطبیق کلمه از الگوریتم DTW^1 استفاده شده است. الگوریتم تطبیق کلمه پیشنهاد شده در این تحقیق بر روی دو مجموعه متون با کیفیت‌های متفاوت آزمایش شده است. هر مجموعه شامل 10 صفحه دست نوشته تاریخی است. نرخ بازشناسی به طور میانگین بر روی دو مجموعه 70/23 % می‌باشد.

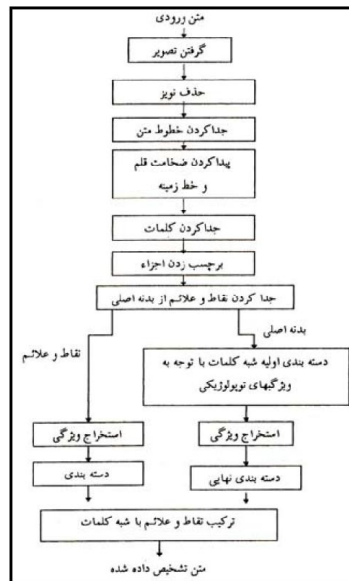
¹ Dynamic time warping



شکل 2-3. تصویر نرمال شده و پروفایل بالایی و پروفایل پروژکشن تصویر [17]

شیرعلی و همکارانش در [45] از ویژگی‌های آماری شبه زرنیکی¹ برای بازشناسی متون دست نویس و چاپی با شبکه عصبی احتمالاتی استفاده کرده است. روش ارائه شده در این تحقیق مبتنی بر شکل کلی زیرکلمات است. تعداد زیرکلمات فارسی 1000 زیرکلمه فرض شده است. شبکه عصبی استفاده شده، با در نظر گرفتن لایه ورودی، چهار لایه دارد. در این تحقیق از 28 ویژگی گشتاورهای شبه زرنیکی مرتبه 4 برای توصیف شکل زیرکلمات استفاده شده است. با آزمایش این روش بر روی مجموعه ای شامل 500 زیرکلمه دست نویس، نوشته شده با خط نسخ و یک صفحه متن چاپی، در هر دو مورد نرخ بازشناسی صحیح 96% گزارش شده است. الگوریتم ارائه شده در این تحقیق در شکل 2-4 آمده است.

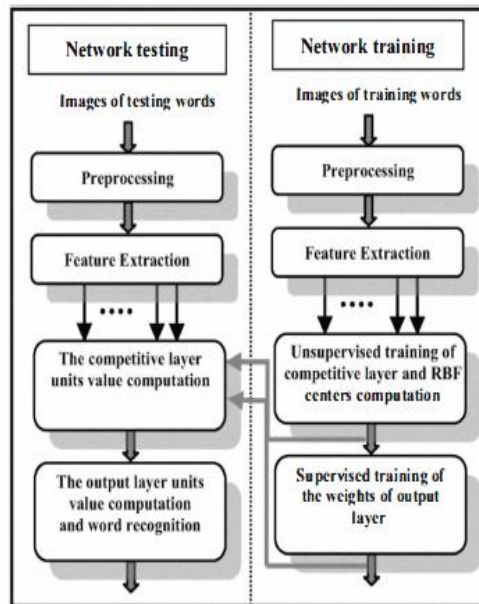
¹ Pseudo Zernike



شکل 2-4. الگوریتم ارائه شده در تحقیق [45]

روش ارائه شده در [10] از ضرایب موجک استخراج شده از پروفایل تصویر هموار شده در چهار جهت استاندارد به عنوان ویژگی استفاده شده است. دلیل انتخاب این ویژگی را ساختار ویژه زبان فارسی بیان کرده است. در زبان فارسی، شکل حروف بسته به موقعیتش در کلمه پروفایل‌های مختلفی تولید می‌کنند. بنابراین ویژگی آماری پروفایل تقریباً تمام ویژگی یک کلمه در زبان فارسی را بیان می‌کند. در مرحله بازشناسی از شبکه عصبی RBF¹ که برای معماری بهینه آن الگوریتم ژنتیک بکار رفته، استفاده شده است. بلوک دیاگرام الگوریتم پیشنهاد شده در این تحقیق در شکل 2-5 آمده است. این روش بر روی اسامی دست نویس 30 نام رایج فارسی آزمایش شده است. برای هر نام 110 نمونه وجود دارد. نرخ بازشناسی در اولین انتخاب 96% و در انتخاب دوم 97% است.

¹ Radial basis function



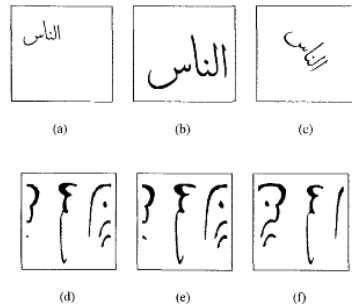
شکل 2-5. الگوریتم ارائه شده در تحقیق [10]

در تحقیق [6] یک ویژگی آماری برای بازشناسی متون دست نوشته معرفی شده است که نسبت به چرخش و اندازه تغییرناپذیر و مبتنی بر شکل کلی کلمات است. این تحقیق از m باند تبدیل موجک گسسته¹ تصویر قطبی کلمه برای استخراج بردار ویژگی برای بازشناسی 100 کلمه فارسی استفاده کرده است.

برای تغییر ناپذیر بودن ویژگی‌ها نسبت به چرخش و اندازه، از تصویر قطبی نرمالیزه شده تصویر کلمه بهره گرفته شده است. تبدیل قطبی نسبت به اندازه و جابه جایی مقاوم و با چرخش تصویر، به صورت ردیفی انتقال می یابد (شکل 2-6).

تصویر قطبی تولید شده از m باند تبدیل موجک گسسته که در برابر انتقال ردیفی مقاوم است، عبور می کند. از هر زیر باند ضرایب موجک، یک مجموعه‌ای از ویژگی‌های انرژی محلی استخراج می شود.

¹ M-band packet wavelet transform



شکل 2-6. سه تصویر کلمه (a), (b), (c) و در زیر هر کدام تصویر قطبی نرمالیزه اش [6]

فاصله ماهالانوبیس¹ برای اندازه‌گیری شباهت بین بردار ویژگی استخراج شده از کلمه ورودی و کلمات فرهنگ لغت استفاده شده است. نرخ بازشناسی برای این ویژگی 92% گزارش شده که نسبت به ویژگی‌های فوریه و گشتاورهای زرنیکی متغیر با چرخش 12% افزایش پیدا کرده است. آقای ابراهیمی و همکارانش روشی برای بازشناسی زیر کلمات چاپی فارسی ارائه داده است [46]. در این تحقیق از ویژگی‌های کلی و جزئی برای توصیف شکل کلی زیر کلمات در یک روش سلسله مراتبی استفاده شده است. تصاویر کامل زیر کلمات با استفاده از ویژگی مکان مشخصه² که اطلاعات کلی شکل زیر کلمات را نشان می‌دهد برای کاهش دامنه جستجو، خوشه‌بندی می‌شوند. تعداد مناسب خوشه‌ها برای تصاویر کامل زیر کلمات، با دو شاخص اعتبارسنجی خوشه‌بندی محاسبه شده است. میانگین هر خوشه به عنوان نماینده آن استفاده شده است. برای ارزیابی خوشه‌ها از طبقه بند حداقل میانگین فاصله بر روی 5000 زیر کلمه با 4 نوع قلم در 3 اندازه مختلف استفاده شده است. در مرحله دوم با استفاده از توصیفگرهای فوریه کانتور، کلمه ورودی از میان اعضای خوشه‌های نزدیک به آن شناسایی می‌شود. دقت بازشناسی در انتخاب نزدیکترین خوشه 78/71% و در پنج خوشه اول 99/1% و در ده خوشه اول 100% گزارش شده است.

¹ Mahalanobis
² Loci feature

3-2 روش‌های مبتنی بر ویژگی‌های محلی

در این روش‌ها تصاویر را به بلوک‌های کوچکتر تقسیم می‌کنند و از هریک از این بلوک‌ها، ویژگی‌های لازم استخراج می‌شود. ویژگی‌های محلی می‌توانند جزئیات شکل را بهتر از ویژگی‌های کلی توصیف کنند. در ادامه تعدادی از روش‌هایی که از این ویژگی‌ها استفاده کرده‌اند، بیان خواهد شد.

دهقان و همکارانش در مرجع [20] از مدل مخفی مارکف برای بازشناسی اسامی دست نوشته 198 شهر ایران استفاده نموده‌اند. در این روش از اطلاعات محلی کانتور استفاده شده است. تصویر کلمات در راستای افقی به پنج قسمت مساوی و در راستای عمودی به قسمت‌هایی که 50% با هم همپوشانی دارند تقسیم می‌شود. عرض تقسیمات عمودی دو برابر میانگین تکه‌های سیاه عمودی تصویر انتخاب می‌شود. برداری شامل هیستوگرام‌های جهتی کدهای زنجیره‌ای در هر یک از این پنجره‌ها بعنوان مدل کلمه انتخاب می‌شود.

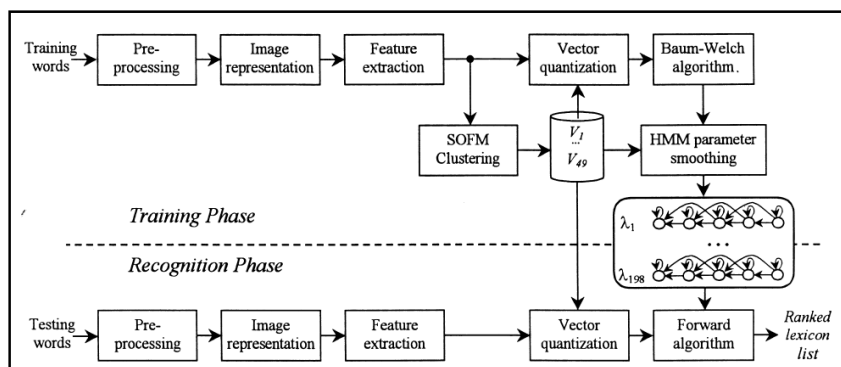
برای کم کردن تعداد مشاهده‌هایی که به مدل مخفی مارکف گسسته اعمال می‌شود، فضای ویژگی با استفاده از یک شبکه عصبی خود سامان کوهنن¹، کوانتیزه شده است. برای هر اسم شهر یک HMM گسسته² با الگوریتم بام-ولش³ بطور مجزا آموزش داده می‌شود. برای هر تصویر ورودی، اسامی شهرها بر حسب میزان شباهت مدل آنها به کلمه ورودی مرتب می‌شوند.

با آزمایش این روش بر روی یک مجموعه کلمات دستنویس شامل اسامی 198 شهر، درصد بازشناسی صحیح 65% در انتخاب اول و 95% در 20 انتخاب اول گزارش شده است. مجموعه تمرین شامل 17000 نام شهر است که بوسیله افراد مختلف نوشته شده‌اند. الگوریتم این روش در شکل 2-7 آمده است.

¹ Kohonen self organizing

² Discrete Hidden Markov Model

³ Baum Welch Algorithm



شکل 2-7. الگوریتم روش [20]

آقای دهقان در تحقیقی دیگر [21] از خوشه بندی FCM¹ و مدل مخفی مارکوف برای بازشناسی اسامی دست‌نویس شهرهای فارسی استفاده کرده است. مراحل پیش‌پردازش و استخراج ویژگی و مدل مخفی مارکف مانند تحقیق [20] است. خوشه‌بندی اولیه به روش FCM انجام شده است. این روش بر روی مجموعه‌ای شامل 198 شهر ایران انجام شده است که درصد بازشناسی در انتخاب اول 67% و در 20 انتخاب اول 96% گزارش شده است.

تحقیق [9] دارای یک مرحله کاهش فرهنگ لغت²، در بازشناسی کلمات دست‌نویس 200 شهر ایران است. مرحله کاهش فرهنگ لغت برای سیستم‌هایی با اندازه فرهنگ لغت بزرگ برای کاهش کلمات نامتشابه با کلمه ورودی در مرحله بازشناسی و افزایش سرعت و دقت سیستم، مناسب است. معمولاً هنگامی که فرهنگ لغت کوچک باشد دقت بازشناسی مهم‌تر از زمان بازشناسی است. اما در فرهنگ لغت بزرگ سرعت بازشناسی نیز یک معیار مهم تلقی می‌شود. در این تحقیق از تعداد و موقعیت نقاط در کلمه برای کاهش تعداد کلماتی که در بازشناسی کلمه آزمون ورودی بکار می‌رود، استفاده شده است.

نقاط کوچکترین واحد در یک متن فارسی و عربی است. تعدادی حروف در زبان فارسی و عربی موجود می‌باشد که دارای شکل یکسان هستند و وجه تمایز بین آنها تعداد و موقعیت نقاط است. شکل نقاط در دست‌نوشته‌ها بسیار متغیر است. در شکل 2-8 تعدادی از این شکل‌ها دیده می‌شود.

¹ Fuzzy C-means

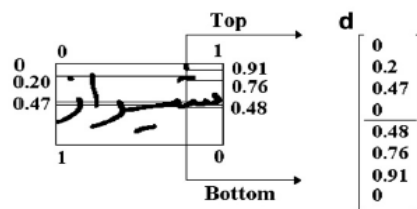
² Lexicon reduction

| | Printed | Handwritten |
|------------|---------|-------------|
| Single dot | . | • |
| Double dot | .. | •• |
| Triple dot | ... | ••• |

شکل 2-8. نحوه نوشتن نقاط در دست نوشته ها [9]

قبل از استخراج نقاط، خط اصلی¹ کلمه محاسبه می شود. خط اصلی محل اتصال حروف به یکدیگر می باشد. خط اصلی در این تحقیق با استفاده از نقاط می نیمم محلی کانتور کلمه و استفاده از رگرسیون خطی بر روی این نقاط تقریب زده می شود. سپس با استفاده از یک سری قواعد، نقاط استخراج شده و با تخصیص یک عدد و یک حرف به هر نقطه، برای هر کلمه یک کد توصیفگر تولید می شود. با استفاده از روش تطبیق رشته دمرا- لون اشتین² کد توصیفگر کلمه ورودی با تمام کدهای توصیفگر کلمات موجود در فرهنگ لغت مقایسه می شود و k تا از بهترین کلمات کاندید انتخاب می شود. مقدار k وابسته به نرخ دقت بازشناسی و درجه کاهش فرهنگ لغت است.

برای هر کلمه موجود در فرهنگ لغت یک مدل HMM گسسته با استفاده از ویژگی انتقال ساخته می شود. ویژگی انتقال، مکان و تعداد انتقال ها از پیکسل سیاه به سفید را در هر ستون، از بالا به پایین و از پایین به بالا، محاسبه می کند (شکل 2-9).



شکل 2-9. نحوه استخراج ویژگی انتقال در تحقیق [9]

در مرحله بازشناسی کلمه آزمون ورودی تنها مدل HMM کلماتی شرکت می کنند که در مرحله کاهش فرهنگ لغت انتخاب شده اند. این الگوریتم بر روی یک پایگاه داده شامل 17000 تصویر از اسامی 200 شهر ایران آزمایش شده است. نرخ بازشناسی در انتخاب اول بدون کاهش فرهنگ لغت 60/2% و با کاهش فرهنگ لغت 73/61% گزارش شده است.

¹ Base line

² Demerau levenshtein

آقای خسروی و همکارش در [47] دو روش برای بازشناسی متون تایپی فارسی با استفاده از شکل کلی ارائه داده‌اند. روش اول بر بازشناسی زیرکلمات، بدون حذف نقاط و علائم آنها استوار است و روش دوم مبتنی بر شکل بدنه زیرکلمات است که از حذف نقاط و علائم زیرکلمه حاصل می‌شود. در هر دو روش در مرحله‌ی آموزش، زیرکلمات مجموعه‌ی آموزش، خوشه‌بندی می‌شوند. برای خوشه‌یابی از الگوریتم ISODATA استفاده شده است. در روش اول بازشناسی طی دو مرحله صورت می‌گیرد: یافتن خوشه‌های نزدیک به ورودی و یافتن نزدیکترین زیرکلمه از بین خوشه‌های نزدیک. روش دوم علاوه بر این مراحل شامل یک مرحله‌ی اضافی برای یافتن زیرکلمه نهایی بر اساس الگوی نقاط نیز هست.

ویژگی که در مرحله آموزش برای خوشه‌بندی استفاده شده، میانگین بلوکی است. برای بدست آوردن این ویژگی، تصویر کلمه به بلوک‌هایی تقسیم می‌شود و از هر بلوک میانگین تعداد پیکسل‌های سیاه محاسبه می‌شود. این ویژگی کلیات شکل را بیان می‌کند. در مرحله بازشناسی با استفاده از فاصله بلوک شهری تعدادی از خوشه‌های نزدیک به کلمه ورودی انتخاب می‌شود. سپس با استفاده از ترکیب ویژگی‌های هیستوگرام گرادیان، تبدیل فوریه کانتور و مکان مشخصه، که جزئیات شکل کلمه را بهتر نشان می‌دهند، نزدیکترین زیر کلمات انتخاب می‌شود. هر دو روش نتایج قابل قبولی روی تصاویر تمیز ارائه می‌دهند به طوری که روش بانقطه دقتی حدود 94% و روش بدون نقطه دقتی حدود 93% در سطح کلمه ارائه می‌دهد. اما در برخورد با تصاویر کم کیفیت و نویزی دچار افت دقت می‌شوند که این کاهش در برخی موارد بسیار شدید است.

در تحقیق [8] تاثیر طبقه بندی $1NN$, KNN ² را در بازشناسی کلمات دست‌نویس عربی بررسی کرده است. الگوریتم پیشنهادی شامل سه مرحله است: پیش پردازش، استخراج ویژگی، کلاس بندی.

¹ Natural network

² K nearest neighbor

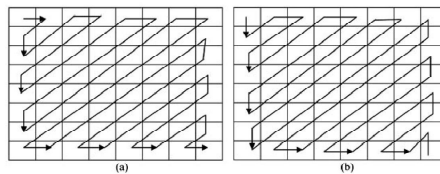
مرحله پیش پردازش شامل: آستانه‌گذاری برای باینری کردن تصاویر اسکن شده، حذف نویز برای بهبود تصاویر، تصحیح شیب و کجی¹ و نازک سازی² می باشد. در مرحله استخراج ویژگی، پنج نوع ویژگی محلی مبتنی بر شکل کلی کلمات استخراج می شود:

1- بلوک‌های هم پوشانی³ (OB): ابتدا لبه‌های تصویر توسط فیلتر بالا گذر عمودی و افقی پریویت⁴ استخراج می‌شود. تصویر فیلتر شده به بلوک‌هایی با اندازه 12×12 با دو پیکسل هم پوشانی تقسیم می‌شود. از هر بلوک میانگین تعداد پیکسل‌های سیاه محاسبه می‌شود.

2- DCT⁵: ضرایب DCT به کاهش حشویات در تصویر کمک می‌کند. ضرایب DCT تصویر $f(x, y)$ توسط DCT دوبعدی محاسبه می‌شود:

$$f(u, v) = \alpha(u)\alpha(v) \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} f(i, j) \cos\left[\frac{(2i+1)u}{2I}\right] \cos\left[\frac{(2j+1)v}{2J}\right] \quad (1.2)$$

$f(u, v)$ تصویر کلمه است. بعد از به کار بردن DCT بر روی تصویر، 100 تا از ضرایب DCT با ترتیب زیگزاگ به صورت یک بردار یک بعدی استخراج می‌شود. به دلیل متمرکز شدن بیشترین انرژی در سمت چپ بالای تصویر ترتیب زیگزاگ انتخاب شده و به دو صورت انجام می‌شود (شکل 2-10)



شکل 2-10. دو نوع ترتیب زیگزاگ برای استخراج ضرایب DCT [8]

3- DCT بلوکی (b-DCT): ابتدا تصویر به بلوک‌هایی تقسیم می‌شود سپس از هر بلوک تعدادی از ضرایب DCT استخراج می‌شود.

¹ Slope and slant correction
² thinning
³ Overlapped blocks
⁴ prewitt
⁵ Discrete cosine time

4- گشتاورهای نامتغیر: از روی تصویر نازک شده لگاریتم گشتاورهای نامتغیر به صورت زیر استخراج می شوند.

$$\eta_{pq} = \sum_i \sum_j x^p y^q f(i, j) \quad (2.2)$$

پارامترهای p, q مرتبه گشتاور را نشان می دهند .

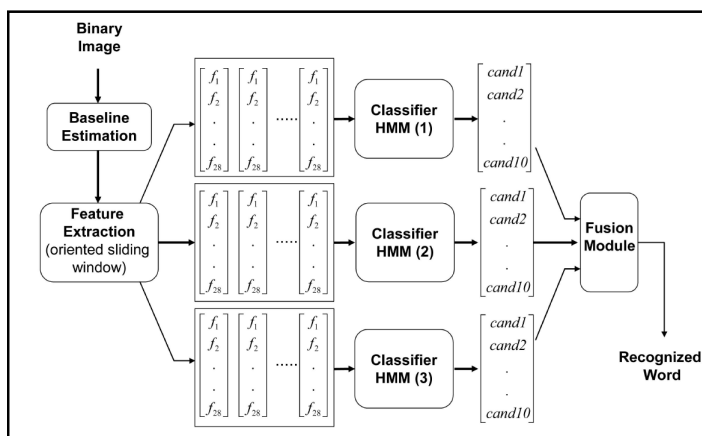
5- DWT : تبدیل موجک گسسته دو بعدی برای استخراج این ویژگیها استفاده می شود. ضرایب موجک اطلاعات خوبی در زمینه مشخصات مکانی و فرکانسی تصویر ارائه می دهند. در این روش هر تصویر به چهار سطح از تبدیل موجک تجزیه می شود.

در این تحقیق از دو کلاسه بند شبکه عصبی و k نزدیکترین همسایه بر روی پایگاه داده IFN/ENIT استفاده شده است. تعداد کلاسها در این پایگاه داده 937 متناسب با تعداد کلمات و شامل 32494 تصویر دستنوشته از این کلمات است. تعداد لایه های خروجی در شبکه عصبی، به دلیل زیاد بودن تعداد کلاسها، برابر با معادل باینری آنها می باشد. بهترین نرخ بازشناسی برای ویژگی DCT همراه با شبکه عصبی گزارش شده است .

Mohamd و همکارش در تحقیق [7] از ترکیب سه مدل HMM یکسان که با ویژگی های ناحیه ای متفاوت آموزش داده شده اند، برای بازشناسی کلمات دستنویس پایگاه داده IFN/ENIT استفاده شده است. نمای کلی سیستم پیشنهادی در این تحقیق در شکل 2-11 آمده است.

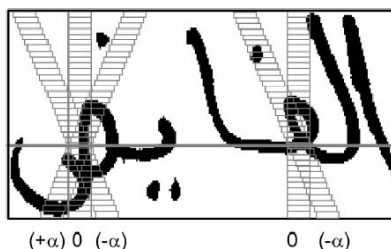
مجموعه ویژگی ها شامل ویژگی های وابسته به خط اصلی¹ و ویژگی های مستقل از خط اصلی می باشد.

¹ baseline



شکل 2-11. نمای کلی سیستم پیشنهاد شده در [7]

آنالیز خطاهای ایجاد شده در سیستم‌های بازشناسی متون دست نوشته نشان می‌دهد که منبع اصلی خطا در این سیستم‌ها انحراف و همپوشانی و جابه جایی علائم، نقاط و نشانه‌های حروف می باشند. برای کاهش خطای ناشی از این منبع برای بلوک‌بندی تصاویر از پنجره‌ها شیب داده شده استفاده شده است (شکل 2-12).

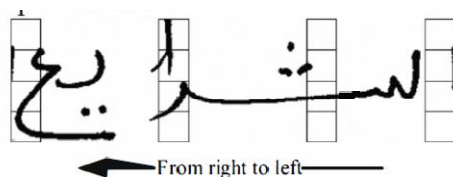


شکل 2-12. نحوه بلوک بندی تصاویر در [7]

قبل از استخراج ویژگی‌ها، خط اصلی تصویر کلمه با روش پروفایل افقی (البته با یک سری اصلاحات) محاسبه می‌شود. ویژگی‌ها از پنجره‌های عمودی و شیب داده شده استخراج می‌شوند. در بازشناسی از سه HMM یکسان که توسط ویژگی‌های استخراج شده از پنجره‌های با شیب‌های متفاوت آموزش داده شده‌اند، استفاده می‌شود. مدل HMM هر کلمه با به هم پیوستن مدل HMM حروف تشکیل دهنده‌اش تولید می‌شود. همه مدل‌های حروف دارای مشخصات یکسان زیر می‌باشند که به صورت تجربی بدست آمده‌اند: چهار حالت و سه انتقال حالت (دو انتقال به حالت‌های بعدی و یک انتقال به خودش) و احتمال مشاهده هر حالت توسط توزیع مخلوط گوسی با ماتریس کواریانس قطری بیان

می‌شود. به طور کلی 167 مدل برای شکل‌های مختلف حروف (ابتدا ، وسط ، انتها کلمه و به صورت منفرد) ساخته شده است. روش‌های مختلف ترکیب کلاسه‌بندها (مجموع آراء، اکثریت آراء، ترکیب MLP¹) و شیب‌های مختلف برای بلوک‌بندی بر روی پایگاه داده IFN/ENIT ارزیابی شده‌اند. بهترین نرخ بازشناسی در اولین انتخاب 90% گزارش شد که برای ترکیب MLP سه مدل HMM با شیب‌های پنجره 0,20,-20 می‌باشد .

روشی مبتنی بر شکل کلی کلمه در تحقیق [15] براساس مدل HMM پیشنهاد شده است. پنجره-گذاری یکی از رایج‌ترین تکنیک‌ها برای استخراج ویژگی‌های محلی از تصویر می‌باشد. عرض پنجره‌ها ثابت و به عنوان پارامتر سیستم است. پنجره‌ها به تعدادی بلوک تقسیم می‌شوند. این پنجره بلوک-بندی شده در امتداد عرض تصویر می‌لغزد و بردار ویژگی از هر بلوک استخراج می‌شود.

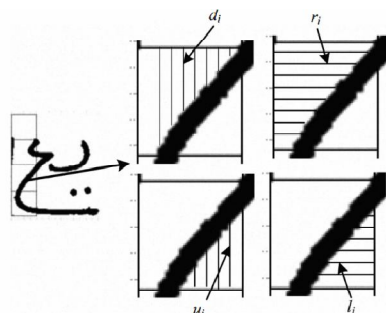


شکل 2-13. نحوه لغزش پنجره فرم بندی شده در امتداد تصویر [15]

از هر بلوک دو نوع ویژگی استخراج می‌شود:

- 1- ویژگی توزیع: که بر اساس چگالی پیکسل‌های پس زمینه در هر فرم می‌باشد.
- 2- ویژگی خمیدگی²: این ویژگی همانطور که در شکل 2-14 دیده می‌شود از محاسبه میانگین فاصله مرز تصویر تا اولین پیکسل سیاه حاصل می‌شود.

¹ Multilayer perceptron
² concavity



شکل 2-14. نحوه استخراج ویژگی خمیدگی در [15]

ویژگی‌های استخراج شده برای آموزش مدل HMM هر حرف به کار می‌روند. مدل هر کلمه از به هم پیوستن مدل HMM حروف تشکیل دهنده‌اش حاصل می‌شوند. آزمایشات انجام شده بر روی پایگاه داده IFN/ENIT نرخ بازشناسی 84% را نتیجه می‌دهد.

در تحقیق [5] الگوریتمی با استفاده از ترکیب ویژگی‌های محلی گرادیان و ویژگی آماری پروفایل و کلاس بند SVM¹ برای بازشناسی متون دست نوشته زبان اردو پیشنهاد شده است. با توجه به ماهیت پیوسته بودن متون دست نوشته اردو، در این تحقیق از ویژگی‌های آماری و محلی که مبتنی بر شکل کلی کلمه هستند و نیازی به جداسازی کلمات به اجزای سازنده‌اش نیست، استفاده شده است.

بعد از انجام پردازش‌های اولیه دو نوع ویژگی استخراج می‌شود:

1- ویژگی گرادیان محلی: یک ویژگی جهتی است که توسط گرادیان تصویر خاکستری شده² تولید می‌شود. نحوه استخراج آن از تصویر خاکستری شده به این صورت است که: ابتدا ابعاد تصاویر خاکستری شده نرمال می‌شود، سپس با اعمال فیلتر روبرتز³ یا سوبل⁴ یا کریش⁵، اطلاعات گرادیان برای هر پیکسل تصویر استخراج می‌شود. در این تحقیق از عملگر روبرتز استفاده شده است. تصویر بلوک بندی می‌شود و در هر بلوک طول گرادیان در جهت‌های کوانتره شده با هم جمع می‌شوند.

¹ Support vector machine

² Grayscale image

³ Roberts

⁴ Sobel

⁵ Krish

2- ویژگی آماری پروفایل: ویژگی‌های آماری شامل ویژگی‌های پروفایل، ویژگی پروژکشن و... است. که در این تحقیق فقط از ویژگی پروفایل بالایی که ساختار بیرونی کلمه را نشان می‌دهد، استفاده شده است.

برای کلاس‌بندی 57 کلمه از یک پایگاه داده به زبان اردو از کلاسه بند SVM استفاده شده است و نرخ بازشناسی 97% گزارش شده است.

2-4 نتیجه گیری

در این فصل، برخی از تحقیقات انجام شده در زمینه بازشناسی متون چاپی و دست نوشته لاتین، عربی، فارسی و اردو مبتنی بر روش‌های بدون قطعه‌بندی مرور شدند. روش‌های ارائه شده در زمینه بازشناسی متون دست‌نوشته مبتنی بر شکل کلی بر اساس ویژگی استفاده شده، به سه دسته تقسیم شدند. دسته اول روش‌های مبتنی بر ویژگی‌های توپولوژیکی است که به مدل کردن وجود یا عدم وجود تعدادی ویژگی‌های توپولوژیکی مانند بالارونده‌ها، پایین‌رونده‌ها، حفره‌ها و حلقه‌ها در تصویر کلمه محدود می‌شوند. تعدادی از تحقیقات که در دسته دوم قرار می‌گیرند مبتنی بر ویژگی‌های آماری هستند که این ویژگی‌ها نسبت به نویز مقاوم هستند و به راحتی استخراج می‌شوند. دسته سوم روش‌های مبتنی بر ویژگی‌های محلی هستند.

تحقیقات در زمینه بازشناسی متون چاپی و دست‌نوشته عربی و فارسی با استفاده از شکل کلی کلمات پیشینه زیادی ندارد. تحقیقات در این زمینه با توجه به نبود پایگاه داده استاندارد و پیچیدگی‌های زبان عربی و فارسی و زیاد بودن تعداد کلمات و کند بودن کامپیوترهای قدیمی، مورد توجه محققان قرار نگرفته بودند. امروزه با توجه به افزایش دانش بشری در زمینه آموزش ماشین و دستیابی به تکنیک‌های جدیدتر و موثرتر و افزایش روزافزون سرعت پردازنده‌ها و بالا رفتن حجم حافظه‌ها و ارزان شدن قیمت آنها پرداختن به این روش‌ها مفید به نظر می‌رسد. تعداد زیاد کلاس‌ها را می‌توان با روش‌های سلسله مراتبی کاهش داد و مقایسه‌ها را بجای یک مرحله در چندین مرحله انجام داد.

فصل سوم

مباحث نظری

فصل سوم: مباحث نظری

در این فصل مباحث نظری که در فصل‌های بعدی استفاده شده‌اند، ارائه می‌شود. روش‌های استخراج ویژگی، خوشه بندی، معیارهای فاصله و کلاسه بند SVM چند کلاسه در این فصل بررسی می‌شوند.

1-3 استخراج ویژگی

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 کلاسه بندی و بازشناسی به کار می‌رود.

1-1-3 ویژگی‌های مبتنی بر شکل

ویژگی‌های مبتنی بر شکل، از مهم‌ترین ویژگی‌های دیداری هستند که برای توصیف محتوای تصاویر استفاده می‌شوند. بدلیل اینکه اطلاعات شکل در اثر نویز یا عوامل دیگر کمتر از بین می‌روند، از توصیفگرهای شکل برای نمایش الگوهای مختلف استفاده می‌شود. با افزایش روزافزون پایگاه‌های داده تصویری، داشتن روش‌هایی برای جستجو در این پایگاه‌ها ضروری است. یکی از روش‌های بازیابی یا بازشناسی الگوها، روش مبتنی بر شکل است. در واقع هدف اصلی این است که در بانکی از تصاویر گوناگون، بتوانیم تصویر مورد نظر خود را در میان تصاویر این بانک جستجو کرده و تصاویر مشابه با آن را پیدا کنیم.

2-1-3 روش‌های توصیف شکل [23]

برای نمایش شکل در حالت کلی، معمولاً بدنبال ویژگی‌هایی بر پایه اطلاعات کانتور شکل یا اطلاعات ناحیه ای شکل هستیم. ویژگی‌های مختلفی مانند: امضاء شکل، ثابت‌های شکل، گشتاورها، انحناء و ویژگی‌های طیفی برای اینکار طراحی شده‌اند. نحوه ارزیابی این ویژگی‌ها بدین ترتیب است که با چه

دقتی می توانند شکل‌های مشابه را از یک پایگاه داده معین بازیابی کنند یا با چه نرخ کلمات را بازشناسی می نمایند.

روش‌های نمایش و توصیف شکل را بطور کلی می توان به دو دسته، روش‌های مبتنی بر کانتور و روش‌های مبتنی بر ناحیه¹، تقسیم کرد. روش‌های مبتنی بر کانتور از ویژگی‌های کانتور شکل برای تشخیص آنها استفاده می کند و در بسیاری از کاربردها بدلیل اینکه اطلاعات کانتور شکل مهم تر از اطلاعات داخل شکل است، این روش مورد توجه قرار گرفته است. با این حال، روش‌های مبتنی بر کانتور دارای مشکلاتی از جمله حساس بودن نسبت به نویز و تغییرات پیرامون شکل می باشند. روش‌های مبتنی بر ناحیه بدلیل استفاده از کل اطلاعات شکل در مقابل نویز و خرابی های شکل بهتر عمل می کنند و معمولاً نرخ بازیابی بهتری ارائه می دهند.

هر یک از این گروه‌ها را می توان به روش‌های سراسری² و روش‌های ساختاری دسته بندی کرد. در شکل 1-3 مثال هایی از روش‌های نمایش و توصیف شکل آمده است.

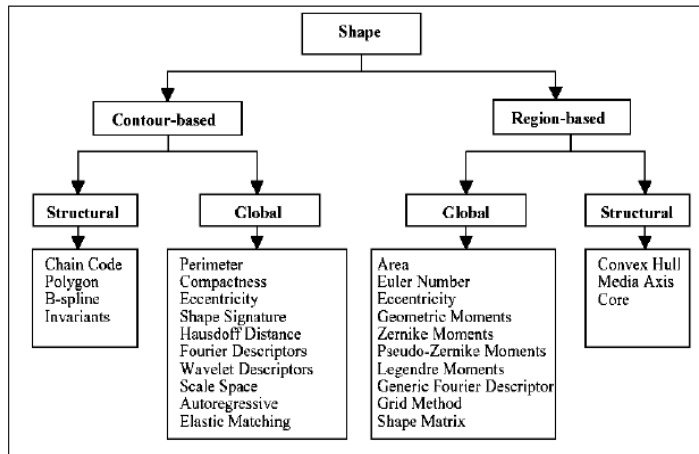
پیاده سازی روش‌های ساختاری از روش‌های سراسری پیچیده تر است. پیچیدگی شاخص گذاری و تطبیق ویژگی‌ها در روش‌های ساختاری، باعث ناپایداری این نوع روش‌های نمایش شکل شده است [46].

مزیت روش‌های ساختاری در اینست که می توان با آنها تطبیق جزء به جزء³ انجام داد. در مواردی که کانتور شکل یک منحنی بسته نباشد یا قسمتی از آن از بین رفته یا خراب شده باشد، تطبیق جزء به جزء کاربرد دارد [30].

¹ Region based methods

² Global

³ partial



شکل 3-1. طبقه بندی روش های نمایش و توصیف شکل گرفته شده از [23]

ویژگی‌ها در روش‌هایی که در حوزه مکان کار می‌کنند به نوبت حساسند و بعد بالایی دارند. این دو مسأله با روش‌های مختلفی مانند: هیستوگرام نقاط سیاه تصویر، گشتاورها، فضای مقیاس و تبدیلهای طیفی¹ حل می‌شوند. در بین این راه حل‌ها، روش تبدیل طیفی و بویژه تبدیل فوریه نتایج بهتری نشان می‌دهد [46].

روش‌های نمایش شکل با توصیفگرهای فوریه هم در فضای یک بعدی و هم در فضای دوبعدی محاسبات ساده‌ای دارند و مقاوم هستند. از آنها می‌توان در کاربردهای متنوعی استفاده کرد و همه این توصیفگرها تفسیر فیزیکی دارند. کارایی بازیابی با این توصیفگرها نشان می‌دهد که این روش در مقایسه با روش‌های مبتنی بر کانتور، راه حل مناسبی برای نمایش شکل‌ها در حالت کلی است [23].

بطور خلاصه می‌توان گفت: روش‌های ساختاری در مواردی که تطبیق جزء به جزء مورد نیاز باشد مناسب هستند. این روش‌ها، کاربردهای کمی دارند. برای کاربردهای عمومی نمایش شکل، روش‌های براساس گشتاورهای مختلط² و تبدیلهای طیفی مانند گشتاورهای زرنیکی³ بهترین انتخاب هستند. اگر در یک کاربرد، کاهش فضای ذخیره سازی مهم باشد، توصیفگر فوریه مناسب است.

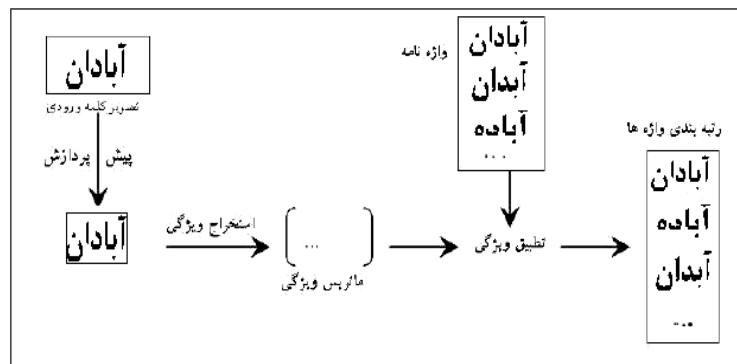
¹ Spectral transforms

² Complex moment

³ Zernike moment

3-1-3 شکل کلی کلمات

روش‌های توصیف و نمایش شکل را می‌توان برای بررسی شکل کلی کلمات نیز بکار برد. ویژگی‌هایی که شکل یک کلمه را توصیف می‌کنند استخراج شده و در یک توصیفگر شکل ذخیره می‌شوند. این توصیفگر شکل با توصیفگرهای کلمات داخل فرهنگ لغت تطبیق داده شده و یک رتبه بندی از آنها بر حسب میزان شباهت به کلمه ورودی تولید می‌شود. بلوک دیاگرام این روش در شکل 2-3 آمده است. مزیت این روش در این است که خطاهای بوجود آمده در اثر جداسازی نادرست حروف در بازشناسی کلمه تأثیری ندارد.



شکل 2-3. نحوه تطبیق کلمه ورودی با کلمات واژه نامه و تولید یک رتبه بندی از کلمات مشابه

از ویژگی‌های مختلفی مانند هیستوگرام‌ها، چگالی پیکسل‌ها، گشتاورهای هندسی، گشتاورهای زرنیکی و شبه زرنیکی¹، ضرایب تبدیل فوریه و مکان مشخصه برای توصیف شکل حروف، ارقام، زیرکلمات و کلمات استفاده شده است. به این ویژگی‌ها، ویژگی‌های آماری² گفته می‌شود. در کلمات فارسی اطلاعات ارزشمندی در بالارونده‌ها، پایین‌رونده‌ها، حلقه‌ها، حفره‌ها و نقاط وجود دارد. ویژگی‌هایی که این اطلاعات را توصیف کند، ویژگی ساختاری³ نامیده می‌شود. در ادامه ویژگی‌هایی که در توصیف شکل کلمات در این پایان نامه استفاده شده است، ذکر خواهد شد.

¹ Pseudo Zernike

² Statistical

³ Structural

1-3-1-3 پروفایل بالا و پایین¹:

ویژگی پروفایل بالا و پایین، نمای دیده شده از بالا و پایین از تصویر کلمه را نشان می دهند. این ویژگی ها با اندازه گیری فاصله اولین پیکسل کلمه با مرز تصویر در هر ستون محاسبه می شود. ابعاد هر کدام از این ویژگی ها برابر با تعداد ستون های تصویر کلمه است.

2-3-1-3 پروفایل پروژکشن عمودی²

پروفایل پروژکشن عمودی، توزیع پیکسل های کلمه را در جهت عمودی نشان می دهد. این ویژگی از مجموع تعداد پیکسل های سیاه در هرستون بدست می آید. ابعاد این ویژگی برابر با تعداد ستون های تصویر کلمه است.

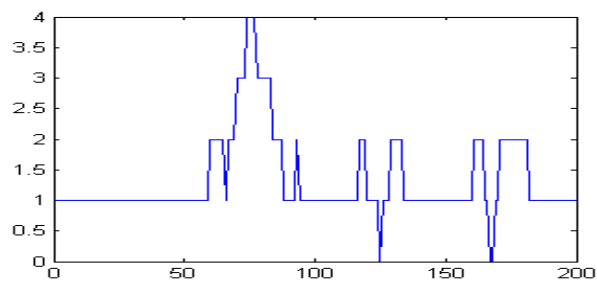
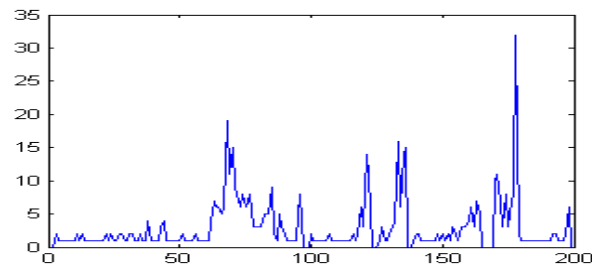
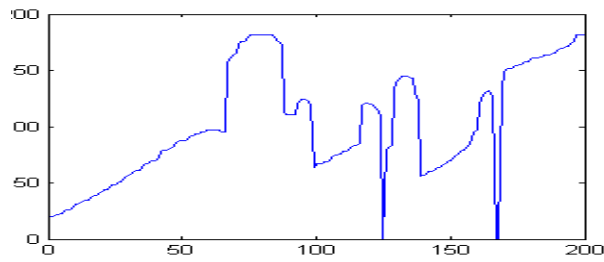
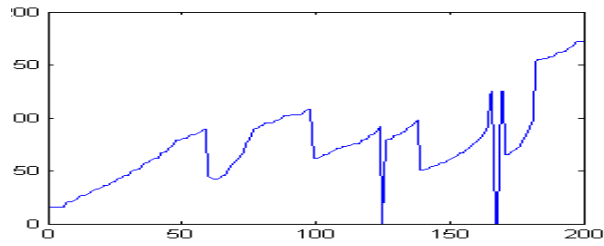
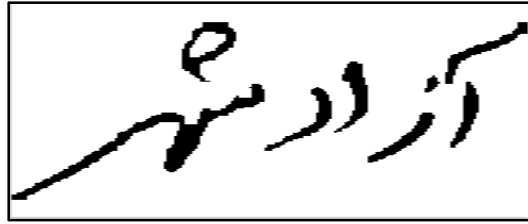
3-3-1-3 پروفایل گذر سیاه / سفید³

سه ویژگی پروفایل بالا و پایین و پروژکشن عمودی ساختار بیرونی شکل را نشان می دهد. برای نشان دادن ساختار درونی کلمه از ویژگی پروفایل گذر سیاه/سفید استفاده شده است. این ویژگی تعداد انتقال ها از پیکسل های پس زمینه (سفید) به پیکسل های پیش زمینه (سیاه) را محاسبه می کند. شکل 3-3 نمونه ای از ویژگی های پروفایل بالا، پایین، پروژکشن عمودی و گذر سیاه به سفید را برای یک تصویر کلمه دست نوشته نشان می دهد.

¹Top and bottom profile

²Vertical projection profile

³Black/white transitions



شکل 3-3. تصویر کلمه دست نوشته، پروفایل
 پایین، پروفایل بالا، پروفایل پروژکشن عمودی، گذر
 سیاه به سفید

3-1-3 ویژگی هیستوگرام گرادیان روشنایی محلی

از ویژگی گرادیان روشنایی نتایج بسیار خوبی در سیستم‌های بازشناسی ارقام و حروف گزارش شده است. ویژگی هیستوگرام گرادیان روشنایی با استفاده از جهت‌های محلی، به خوبی مشخصه‌های نمونه‌های داخل یک کلاس و تغییرات بین کلاسی را مدل می‌کند. این ویژگی از گرادیان محلی تصاویر خاکستری شده بدست می‌آید. در این روش تصویر ورودی ابتدا به بلوک‌هایی تقسیم شده و از هر بلوک ویژگی‌های مربوط به جهت گرادیان استخراج می‌شود [47].

مراحل استخراج ویژگی هیستوگرام گرادیان محلی به صورت زیر می‌باشد:

- برای تبدیل تصاویر باینری به تصاویر خاکستری از فیلتر میانگین 3×3 استفاده شده است. در فیلتر میانگین مقدار هر پیکسل با مقدار میانگین شدت همسایگی 3×3 به مرکز آن پیکسل جایگزین می‌شود.

- برای بدست آوردن اطلاعات گرادیان هر پیکسل از عملگرهای سوبل یا روبرتز یا کریش استفاده می‌شود. عملگر کریش از این نظر که مولفه گرادیان را در چهار راستای افقی، عمودی، قطر اصلی و قطر فرعی محاسبه می‌کند، از بقیه پیچیده‌تر است. عملگرهای روبرتز و سوبل تنها دو مولفه گرادیان در اختیار می‌گذارند. در اینجا از عملگر سوبل استفاده شده است. این عملگر نسبت به عملگر کریش پیچیدگی محاسباتی کمتری دارد و با توجه به این که نسبت به عملگر روبرتز، همسایگی بزرگتری برای هر پیکسل در نظر می‌گیرد تخمین بهتری از گرادیان را نتیجه می‌دهد. کرنل عملگرهای سوبل در شکل 3-4 آمده اند.

| | | | | | |
|----|---|---|----|----|----|
| -1 | 0 | 1 | 1 | 2 | 1 |
| -2 | 0 | 2 | 0 | 0 | 0 |
| -1 | 0 | 1 | -1 | -2 | -1 |

شکل 3-4. عملگرهای سوبل

- مولفه‌های گرادیان عمودی g_y و افقی g_x تصویر ورودی f در موقعیت (x, y) به صورت زیر محاسبه می‌شود:

$$g_x(x, y) = f(x+1, y-1) + 2f(x+1, y) + f(x+1, y+1) - f(x-1, y-1) - 2f(x-1, y) - f(x-1, y+1)$$

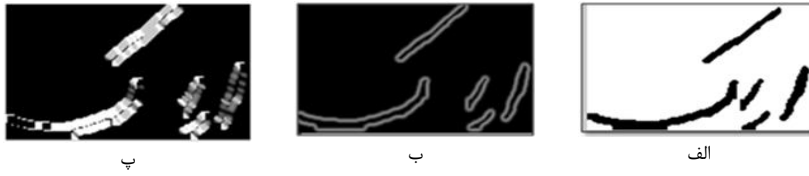
$$g_y(x, y) = f(x-1, y+1) + 2f(x, y+1) + f(x+1, y+1) - f(x-1, y-1) - 2f(x, y-1) - f(x+1, y-1)$$

- به کمک g_x و g_y جهت و اندازه گرادیان هر پیکسل به صورت زیر بدست می‌آید:

$$\text{اندازه} \quad s(x, y) = \sqrt{g_x^2 + g_y^2} \quad (1.3)$$

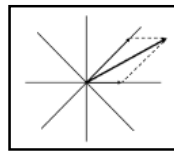
$$\text{جهت} \quad \theta(x, y) = \tan^{-1}(g_y / g_x) \quad (2.3)$$

در شکل 3-5 تصاویر گرادیان یک نمونه از تصویر کلمه را نشان می‌دهد.



شکل 3-5. الف-تصویر ورودی ب-اندازه گرادیان پ-جهت گرادیان

- برای استخراج ویژگی، جهت‌های گرادیان به 32 جهت با فاصله‌های $\pi/16$ (جهت‌های پایه) کوانتیزه می‌شوند. هر زاویه را می‌توان به دو روش کوانتیزه کرد. در روش اول هر جهت به دو جهت پایه تجزیه می‌شود. خطای کوانتیزه کردن در این روش کم است و هر جهت گرادیان به دو جهت پایه برچسب می‌خورد (شکل 3-6).



شکل 3-6. کوانتیزه کردن با تجزیه بردار جهت گرادیان

در روش دیگر هر جهت گرادیان به نزدیکترین جهت پایه برچسب می‌خورد. این روش نسبت به روش اول خطای کوانتیزه بیشتری دارد ولی پیاده سازی آن ساده‌تر است.

- برای بدست آوردن هیستوگرام محلی جهت گرادیان، گرادیان تصویر کلمه به بلوک‌هایی تقسیم می‌شود. اندازه هر بلوک در تمام تصاویر فرهنگ لغت می‌تواند ثابت یا متغیر باشد. تعداد بلوک‌ها در تمام تصاویر کلمات ثابت و به میزان دقت شناسایی و پیچیدگی الگوریتم وابسته است.

یعنی تعداد بلوک‌ها به گونه‌ای انتخاب می‌شود که در عین حال که به دقت مورد نظر نزدیک می‌شویم، پیچیدگی محاسباتی قابل قبول باشد.

- در هر بلوک در هر یک از 32 جهت، اندازه گرادیان‌ها با هم جمع می‌شوند. بنابراین از هر بلوک یک بردار 32 تایی استخراج می‌شود.
- اگر تعداد بلوک‌ها در جهت افقی n و درجهت عمودی m باشد. ابعاد ویژگی برابر با $32 \times n \times m$ خواهد شد. برای کاهش ابعاد ویژگی، می‌توان تعداد بلوک‌ها را با بیان تصویر به صورت هرمی¹ و استفاده از فیلتر گوسی کاهش داد و برای کاهش بردار ویژگی‌های مربوط به جهت‌های پایه، بردار ویژگی را با فیلتر [1 4 6 4 1] کانوال نموده و سپس به کمک عملیات کاهش نرخ نمونه برداری تعداد ویژگی‌ها را کاهش می‌دهیم [5].

2-3 روش‌های خوشه بندی

هدف از تحلیل خوشه، شناسایی گروه‌هایی از اشیاء مشابه و آشکار کردن توزیع الگوها و همبستگی‌های موجود در مجموعه‌های بزرگ داده‌ها است. خوشه‌یابی زمینه تحقیقاتی گسترده‌ای در کاربردهای مهندسی، اقتصاد و علوم اجتماعی دارد. در سالهای اخیر به دلیل فراهم بودن مجموعه داده‌های تجربی بزرگ و نیاز فزاینده به کاوش این داده‌ها، وجود الگوریتم‌های کارآمد و قابل اعمال در حوزه‌های مختلف ضروری بنظر می‌رسد.

در این بخش مفاهیم اولیه خوشه‌یابی در کنار روش‌های شناخته شده خوشه‌بندی مرور می‌شود. برای تهیه این بخش از مراجع [46] و [48] استفاده شده است.

1-2-3 تعریف خوشه بندی

در دنیای امروز با اطلاعات و داده‌های فراوانی روبه‌رو می‌شویم که قبل از پردازش و آنالیز آنها، لازم است ذخیره شوند. یکی از مهمترین کارهایی که در رابطه با این داده‌ها باید انجام داد گروه بندی و طبقه بندی آنهاست. طبقه بندی نقشی مهم در رشد و پیشرفت سیستم‌های هوشمند دارد. هنگامی

¹ Pyramid

که یک الگو وارد سیستم شناسایی می‌شود در ابتدا ویژگی‌هایی که آن را توصیف می‌کند، استخراج می‌شود و سپس با ویژگی‌های الگوهای معلوم مقایسه می‌شود و براساس شباهت‌ها و یک سری قواعد، هویت الگوی نامعلوم مشخص می‌شود.

به طور کلی طبقه بندی داده‌ها به دو دسته تقسیم می‌شود: طبقه بندی با نظارت، طبقه بندی بدون نظارت. در طبقه بندی با نظارت یک مجموعه‌ای از داده‌های استخراج شده از الگوها با کلاس معین در دسترس است. در طبقه بندی بدون نظارت که خوشه‌بندی نیز نامیده می‌شود چنین مجموعه‌ای از داده‌های برچسب خورده وجود ندارد. در این دسته، داده‌ها بر اساس معیار از پیش تعیین شده به تعدادی خوشه تقسیم می‌شوند. داده‌های موجود در هر یک از این خوشه‌ها بیشترین شباهت را نسبت به هم و نسبت به داده‌های موجود در خوشه‌های دیگر کمترین شباهت را دارند.

هدف اصلی فرآیند خوشه بندی، تفکیک کردن داده‌ها به گروه‌های معنادار است، به نحوی که شباهت‌ها و تضادها قابل مشاهده باشند و بتوان نتیجه گیری مناسبی در مورد آنها انجام داد. این ایده در حوزه‌های مختلفی مانند علوم زیستی، علوم پزشکی و مهندسی کاربرد دارد. خوشه‌بندی بسته به زمینه کاربرد آن اسامی متفاوتی مانند یادگیری بدون نظارت (در بازشناسی الگو)، طبقه بندی عددی¹ (زیست شناسی² و بوم شناسی³) و گونه شناسی⁴ (در علوم اجتماعی) دارد.

فرآیند خوشه‌بندی شامل مراحل است که در شکل 3-7 آمده است .

- استخراج یا انتخاب ویژگی: استخراج ویژگی از الگوهای مورد بررسی و انتخاب ویژگی‌های مناسب از مجموعه ویژگی‌های استخراج شده برای کاهش ابعاد از مراحل اساسی خوشه بندی است.
- معیار شباهت: معیاری است که میزان شباهت بردارهای ویژگی استخراج شده از الگوهای مختلف را اندازه می‌گیرد. معیار شباهت مستقیماً بر روی شکل خوشه‌های حاصل تأثیر می‌گذارد.

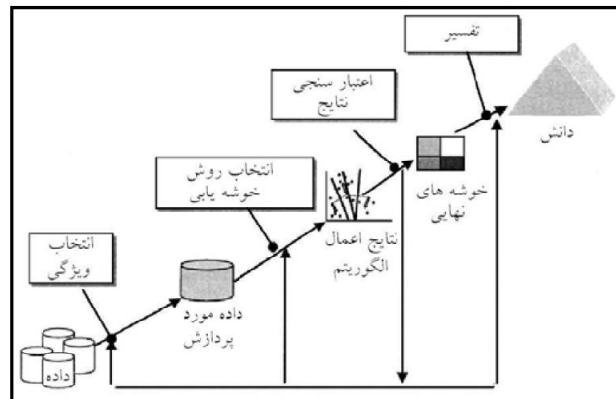
¹ Numerical taxonomy

² Biology

³ Ecology

⁴ Typology

- معیار خوشه بندی: این معیار به تفسیر فرد از نوع خوشه های مشهود بر روی مجموعه داده های ورودی وابسته است. مثلاً خوشه های فشرده با یک معیار و خوشه های خطی با معیار دیگری حاصل می شود. معیار خوشه بندی توسط تابع هزینه یا بعضی قواعد دیگر بیان می شود.



شکل 3-7. مراحل فرآیند خوشه بندی گرفته شده از [46]

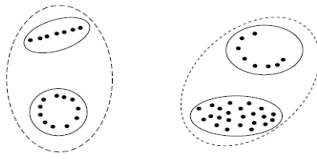
- اعتبار سنجی¹ نتایج: صحت نتایج یک الگوریتم خوشه بندی با استفاده از معیارها و روشهای مناسب سنجیده می شود. از آنجائیکه الگوریتم های خوشه بندی، صرف نظر از روش به کار رفته، خوشه هایی ناشناخته تعریف می کنند، دستیابی به افراز نهایی مجموعه داده در اکثر کاربردها مستلزم استفاده از نوعی ارزیابی است.

- تفسیر² نتایج: هدف نهایی خوشه بندی این است که فرد یک بینشی مناسب از مجموعه داده های مورد بررسی پیدا کند و از آن برای آنالیزهای بعدی استفاده کند.

انتخاب های متفاوت ویژگی، معیار شباهت، معیار خوشه بندی و الگوریتم خوشه بندی، نتایج متفاوتی خواهد داشت که وابسته به نوع کاربرد است. برای مثال شکل 3-8 به دو صورت خوشه بندی انجام شده است که هر دو درست و معتبر هستند و بسته به نوع کاربرد و دانش فرد می توان از آنها استفاده کرد.

¹Validation

²Interpretation



شکل 3-8. نتیجه یک خوشه بندی بزرگ دو خوشه، و خوشه بندی کوچک چهار خوشه شده است

2-2-3 الگوریتم های خوشه بندی

روش های خوشه بندی متعددی در مقالات ارائه شده اند. این روش ها را از سه منظر می توان دسته بندی کرد:

- نوع داده ورودی الگوریتم
 - معیار خوشه بندی که بیانگر شباهت بین اعضای مجموعه داده است
 - مفاهیم تئوری که روش های تحلیل خوشه بر آنها استوار هستند، مانند تئوری فازی
- با توجه به چگونگی تعریف خوشه، این الگوریتم ها را می توان به پنج دسته عمده تقسیم کرد:

- الگوریتم های ترتیبی¹
- الگوریتم های سلسله مراتبی²
- الگوریتم های خوشه بندی افرازی³
- الگوریتم های مبتنی بر شبکه⁴
- الگوریتم های مبتنی بر تراکم⁵

در ادامه هر کدام از این الگوریتم ها بررسی می شود.

1-2-2-3 الگوریتم های ترتیبی

¹ Sequential algorithm
² Hierarchical algorithm
³ clustering Partitional
⁴ Density-based
⁵ Grid-based

در این نوع الگوریتم‌ها بردارهای ویژگی نمونه‌ها به ترتیب وارد می‌شوند. فاصله هر بردار ورودی با بردار ویژگی نماینده خوشه‌های موجود محاسبه می‌شود. در صورتی که فاصله بردار ورودی از نماینده خوشه‌های موجود کمتر از حد یک آستانه باشد بردار به خوشه‌ای تعلق دارد که کمترین فاصله را با آن داشته باشد و نماینده آن خوشه به روز می‌شود. در غیر این صورت خوشه جدیدی تولید می‌شود و بردار ورودی در آن قرار می‌گیرد. یکی از فاکتورهای مهمی که در نتیجه نهایی الگوریتم‌های خوشه بندی ترتیبی تأثیر دارد ترتیب بردارهای ورودی است. فاکتور دیگری که می‌تواند در نتیجه تعداد خوشه نهایی تأثیر داشته باشد حد آستانه است و به طوری که اگر حد آستانه بزرگ گرفته شود تعداد خوشه‌های نهایی کمتر و اگر خیلی کوچک گرفته شود تعداد خوشه‌ها زیادتر از حد مناسب خواهد شد.

یکی از معایب این الگوریتم این است که بعد از این اینکه تمام بردارهای ویژگی نمونه‌ها وارد شدند تعداد و شکل نهایی خوشه‌ها مشخص می‌شود. برای غلبه بر این مشکل الگوریتم دو مرحله‌ای پیشنهاد شده است به طوری که در مرحله اول تعداد خوشه‌ها با برچسب زدن بعضی از بردارها مشخص می‌شود و در مرحله دوم بردارهایی که در مرحله اول برچسب نخورده بودند، با توجه به فاصله آنها با خوشه‌های موجود برچسب می‌خورند. این روش حساسیت الگوریتم را به ترتیب ورود بردارها کمتر می‌کند.

این نوع از الگوریتم‌ها بسته به نوع معیار فاصله که استفاده می‌شود تمایل به تشکیل خوشه‌های کروی شکل و یا بیضوی شکل دارند.

3-2-2-2 الگوریتم‌های سلسله مراتبی

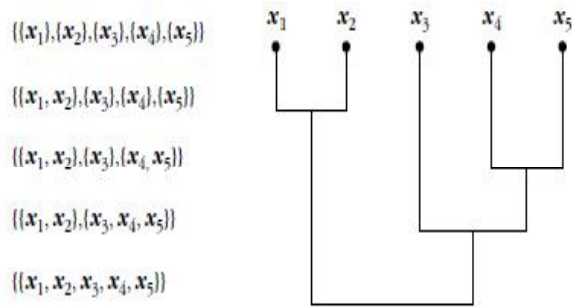
این نوع از الگوریتم‌ها به جای تولید یک خوشه بندی تنها، یک دنباله از خوشه بندی را تولید می‌کند. این الگوریتم‌ها در بسیاری از زمینه‌ها مثل باستان‌شناسی، زیست شناسی مدرن، پزشکی و.... کاربرد دارد و به دو دسته تقسیم می‌شوند: الگوریتم‌های سلسله مراتبی جمع‌کننده و الگوریتم‌های سلسله مراتبی تقسیم‌کننده.

- الگوریتم‌های جمع کننده¹: برنامه پایه این الگوریتم‌ها به این صورت است که در ابتدا هر بردار ویژگی تشکیل یک خوشه می دهد. در مرحله بعدی هر دو خوشه‌ای که خیلی به هم نزدیک است با هم ادغام می‌شوند و به همین صورت ادامه می‌یابد تا همه بردارهای ویژگی در یک خوشه قرار بگیرند. الگوریتم‌های جمع کننده نیز به دو دسته تقسیم می‌شوند: الگوریتم‌های جمع کننده براساس مفاهیم تئوری ماتریس و الگوریتم‌های جمع کننده بر اساس مفاهیم تئوری گراف

- الگوریتم‌های سلسله مراتبی تقسیم کننده²: این الگوریتم‌ها بصورت معکوس روش الگوریتم‌های جمع کننده عمل می‌کنند. در مرحله اول تمام بردارهای ورودی در یک خوشه قرار می‌گیرند. در مرحله بعد بردارهای ویژگی به دو گروه تقسیم می‌شوند. در این رابطه می‌بایست تمام تقسیم‌بندی‌های ممکن بررسی شود و بهترین تقسیم‌بندی بر اساس معیاری خاص انتخاب شود. این روال برای هر یک از دو خوشه تشکیل شده تکرار می‌شود تا اینکه هر یک از بردارها در یک خوشه قرار بگیرند (به تعداد بردارهای ویژگی خوشه ایجاد می‌شود).

نتایج الگوریتم‌های سلسله مراتبی را می‌توان در درختی به نام دندروگرام³ مشاهده کرد که ارتباط بین خوشه را نشان می‌دهد. با قطع دندروگرام در یک سطح دلخواه، یک افراز از مجموعه داده به خوشه‌های مجزا بدست می‌آید. در مواردی که یک خوشه‌بندی مناسب برای یک مجموعه دیتای مشخص لازم است باید بررسی شود که کدام یک از افرازهای تولید شده توسط الگوریتم سلسله مراتبی برای مجموعه دیتا بهتر است و همان‌جا الگوریتم پایان داده شود. شکل 3-9 یک دندروگرام از یک الگوریتم سلسله مراتبی جمع کننده همراه با افرازهای تشکیل شده از آن را نشان می‌دهد.

¹ Agglomerative
² Divisive
³ Dendrogram



شکل 3-9. دندروگرام یک الگوریتم سلسله مراتبی جمع کننده و افراز های تشکیل شده از آن

الگوریتم‌های سلسله مراتبی با پیچیدگی محاسباتی $O(N^3)$ (تعداد نمونه‌ها) برای مجموعه داده‌های بزرگ مناسب نیستند. در ادامه برخی از الگوریتم‌های سلسله مراتبی مناسب برای مجموعه داده‌های بزرگ، آورده شده است.

CURE¹: یک الگوریتم سلسله مراتبی جمع کننده می‌باشد که برای نمایش هر خوشه از یک تعدادی نمایشگر استفاده می‌کند. با استفاده از این نمایشگرها شکل هر خوشه مشخص می‌شود. برای کاهش بی‌نظمی در مرز خوشه‌ها، نمایشگرهای انتخاب شده به سمت میانگین خوشه‌ها حرکت می‌کنند. به این عمل جمع کردن² گفته می‌شود چون حجم فضای تعیین شده توسط نمایشگرها به سمت میانگین خوشه جمع می‌شود. این عمل اثر نویزها را کاهش می‌دهد. این روش از ترکیب نمونه برداری اتفاقی با خوشه بندی افرازی برای مجموعه داده‌های بزرگ استفاده می‌کند.

ROCK³: الگوریتمی است که برای داده‌های بولی و قیاسی مناسب است. به دلیل اینکه برای این داده‌ها نمی‌توان میانگین را به عنوان نماینده خوشه‌ها در نظر گرفت و فاصله بین جفت نقاط را نمی‌توان با معیارهای رایج محاسبه کرد بنابراین الگوریتم ROCK ایده اتصال را به جای فاصله برای ادغام خوشه‌ها در نظر گرفته است.

3-2-2-3 الگوریتم های خوشه بندی افرازی

¹ Clustering Using Representative

² Shrinking

³ Robust Clustering using link

در این روش‌ها، مجموعه داده مستقیماً به مجموعه‌ای از خوشه‌های جدا از هم تجزیه می‌شود. به عبارت بهتر، در این روشها تعداد مناسب افزازها با بهینه کردن یک تابع هزینه تعیین می‌شوند. تابع هزینه ممکن است بر ساختار کلی یا محلی داده‌ها تاکید داشته باشد و فرآیند بهینه‌سازی آن یک فرآیند تکراری است.

k - میانگین^۱: هدف الگوریتم k - میانگین بهینه کردن یک تابع هدف است که با معادله زیر قابل توصیف است:

$$E = \sum_{i=1}^K \sum_{x \in C_i} d(x, m_i) \quad (3.3)$$

در رابطه بالا m_i مرکز خوشه C_i و $d(x, m_i)$ فاصله اقلیدسی دو نقطه x و m_i است. در این روش تابع هزینه با کمینه کردن فاصله هر نقطه از مرکز خوشه مربوطه، بهینه می‌شود. الگوریتم با مقدار دهی اولیه مراکز k خوشه شروع می‌شود. سپس هر نقطه از مجموعه داده به نزدیکترین مرکز خوشه نسبت داده می‌شود و مراکز خوشه‌ها با محاسبه میانگین آنها مشخص می‌شوند. این کار تا زمانی که مراکز خوشه تغییر نکنند، ادامه پیدا می‌کند.

الگوریتم‌های ISODATA و ISOCLUS از الگوریتم k - میانگین ناشی شده‌اند. در این الگوریتم‌ها تعداد خوشه در هر تکرار با حذف خوشه‌های کوچک و یا تقسیم خوشه‌های بزرگ به روز می‌شود. الگوریتم دیگر این گروه PAM^۲ است. در این الگوریتم، هدف پیدا کردن یک عضو به عنوان میانه^۳ برای هر خوشه است، به عبارت دیگر هدف پیدا کردن مرکزی‌ترین عضو هر خوشه است. الگوریتم با انتخاب اتفاقی C عضو از مجموعه داده به عنوان میانه هر یک از C خوشه شروع می‌شود. سپس اعضای انتخاب نشده مجموعه داده با نزدیکترین میانه برای تشکیل خوشه‌ها ترکیب می‌شوند. الگوریتم PAM میانه‌ها را با اعضای غیر میانه، عوض می‌کند و این کار را تا زمانی ادامه می‌دهد که همه اعضا به عنوان

^۱ K-means

^۲ Partitioning Around Medoids

^۳ Medoid

میانه آزمایش شده باشند. واضح است که این روش بخاطر مقایسه هر نمونه با تمام مجموعه داده برای انتخاب میانه، یک الگوریتم پر هزینه است.

¹CLARA: پیاده سازی الگوریتم PAM در یک زیر مجموعه از مجموعه داده است. در این روش نمونه‌هایی از مجموعه داده به تصادف انتخاب می‌شوند و الگوریتم PAM روی آنها اعمال شده و خوشه بندی مناسب استخراج می‌شود.

² CLARANS: این الگوریتم روش‌های نمونه برداری را با PAM ترکیب می‌کند. فرآیند خوشه‌ابی را می‌توان بصورت جستجوی یک گراف در نظر گرفت که در آن هر گره به منزله یک جواب است. خوشه‌بندی که پس از تعویض یک میانه بدست می‌آید، یک همسایه خوشه بندی فعلی خوانده می‌شود.

CLARANS: این روش یک گره را انتخاب کرده و با تعدادی از همسایه‌هایش برای یافتن یک کمینه محلی مقایسه می‌کند. تعداد همسایه‌ها توسط کاربر تعیین می‌شود. اگر همسایه مناسبتری، با مربع خطای کمتر پیدا شد، الگوریتم به آن گره ارجاع داده شده و فرآیند از اول شروع می‌شود. در غیر این صورت، خوشه بندی فعلی یک کمینه محلی است. اگر کمینه بهینه پیدا شود، CLARANS با یک مجموعه گره‌های اتفاقی جدید شروع به جستجوی کمینه بهینه می‌کند. الگوریتم‌های k-mode و k-prototype بر اساس ایده‌های مشابه الگوریتم k-mode میانگین کار می‌کنند و در خوشه بندی داده‌های قیاسی کاربرد دارند. در الگوریتم k-mode مراکز خوشه‌ها با مدها³ تعویض شده‌اند و یک معیار عدم شباهت برای تطبیق با داده‌های قیاسی طراحی شده است.

خوشه‌بندی فازی: الگوریتم‌هایی که تاکنون مرور شدند، خوشه‌های سخت⁴ ایجاد می‌کنند. به این مفهوم که یک نمونه از فضای داده، متعلق به یک خوشه است یا متعلق به آن نیست. خوشه‌ها همپوشانی ندارند و این نوع افراز را خوشه‌بندی سخت می‌نامند. استفاده از عدم قطعیت در الگوریتم-

¹ Clustering Large Applications

² Clustering Large Applications based on Randomized Search

³ Mode

⁴ Crisp

های خوشه بندی با کاربرد مفاهیم منطق فازی در روند آنها امکان پذیر می شود. معمول ترین الگوریتم خوشه بندی فازی C-میانگین فازی¹ یا FCM است. که بسط یافته الگوریتم k-میانگین کلاسیک است. هدف FCM پیدا کردن نقطه مشخصه هر خوشه به عنوان مرکز آن خوشه و درجه عضویت نمونه ها در خوشه است.

3-2-2-4 الگوریتم های مبتنی بر تراکم

این الگوریتم ها برای مجموعه داده های بزرگ مناسب است. ایده اصلی این روش ها در ترکیب نقاط همسایه در مجموعه داده برای تشکیل خوشه ها بر اساس شرایط تراکم است. به عبارت دیگر در این الگوریتم ها، خوشه ها بصورت نواحی با تراکم بالا که با نواحی با تراکم پایین از هم جدا شده اند، تعریف می شوند.

یکی از شناخته شده ترین الگوریتم های این گروه DBSCAN² است. ایده اصلی این روش این است که برای هر نقطه از یک خوشه، همسایگی با شعاع داده شده باید حتما شامل حداقل تعداد نمونه ها باشد. این الگوریتم در برابر نویز مقاوم است و قابلیت ایجاد خوشه هایی با شکل های دلخواه را دارد.

الگوریتم دیگر این گروه DENCLUE³ است. این الگوریتم برای خوشه بندی پایگاه های داده بزرگ چند رسانه ای ارائه شده است. ایده اصلی این روش مدل کردن تحلیلی چگالی نقاط به صورت مجموع توابع تاثیر⁴ نقاط مجموعه داده است. تابع تاثیر تابعی است که فشردگی یک نقطه را در همسایگی خودش نشان می دهد. بنابراین خوشه ها با جذب کننده های چگالی⁵ مشخص می شوند. جذب کننده های چگالی بیشینه های محلی تابع چگالی کلی هستند.

3-2-2-5 الگوریتم های مبتنی بر شبکه

این الگوریتم ها برای کاوش داده های مکانی ارائه شده اند و مشخصه اصلی آنها، تقسیم تصویر به تعداد محدودی خانه است، که عملیات بعدی بر روی این تصویر چندی شده انجام می شود. این الگوریتم ها

¹ Fuzzy C Mean

² Density-Based Spatial Clustering of Applications with Noise

³ DENsity of CIUstEring

⁴ Influence functions

⁵ Density attractors

دارای دو مرحله اصلی هستند: 1- شناسایی خانه هایی که احتمالاً در آن خوشه وجود دارد 2- تعیین خوشه های هر یک از این خانه ها

STING یکی از این الگوریتم ها است. در این الگوریتم سطح مکانی با استفاده از یک ساختار سلسله مراتبی به خانه های مستطیلی تقسیم می شود. سپس پارامترهای آماری مجموعه داده مانند میانگین، واریانس، کمینه و بیشینه اعضای داخل خانه ها محاسبه می شوند. یک ساختار سلسله مراتبی از خانه های شبکه برای توصیف اطلاعات خوشه بندی در سطوح مختلف تولید می شود. بر اساس این ساختار، در الگوریتم STING می توان از اطلاعات خوشه بندی برای تخصیص مناسب اعضا جدید به خوشه ها استفاده کرد.

3-2-3 مقایسه الگوریتم های خوشه بندی

در این بخش مشخصات اساسی الگوریتم های خوشه بندی بصورت مقایسه ای ارائه شده است [47]. مقایسه ها از چهار دیدگاه: نوع داده ورودی، شکل خوشه ها، مقاومت در برابر نویز و پیچیدگی انجام شده است. در این مقایسه پارامترهای ورودی الگوریتم ها نیز ارائه شده اند.

جدول 3-1. مقایسه الگوریتم های خوشه بندی

| الگوریتم | نوع داده ورودی | پیچیدگی | شکل خوشه | مقاوم در برابر نویز | پارامتر های ورودی | نتایج |
|----------|----------------|---|----------|---------------------|---|-------------------------------------|
| ترتیبی | عددی | زمانی $O(N)$ | محدب | خیر | تعداد ماکزیمم خوشه ها ، حد آستانه فاصله | تعداد خوشه ها ، مراکز خوشه ها |
| CURE | عددی | زمانی $O(N^2 \log N')$ حافظه $O(N')$ | دلخواه | بله | تعداد خوشه ها - تعداد نماینده های خوشه ها | نسبت دادن مقادیر داده ها به خوشه ها |
| ROCK | قیاسی | $O(N^2 + Nm_m m_a + N^2 \log N)$ که در آن حداکثر تعداد همسایگی و m_a میانگین تعداد همسایگی یک نقطه است . | دلخواه | بله | تعداد خوشه ها | نسبت دادن مقادیر داده ها به خوشه ها |
| K-means | عددی | $O(N)$ | محدب | خیر | تعداد خوشه ها | مراکز خوشه ها |
| K-mode | قیاسی | $O(N)$ | محدب | خیر | تعداد خوشه ها | مراکز خوشه ها |
| PAM | عددی | $O(K(N-K)^2)$ | محدب | خیر | تعداد خوشه ها | مدیود های خوشه ها |
| CLARA | عددی | $O(K(40+K)^2 + K(N-K))$ | محدب | خیر | تعداد خوشه ها | مدیود های خوشه ها |
| CLARAS | عددی | $O(KN^2)$ | محدب | خیر | تعداد خوشه ها و تعداد همسایه ها | مدیود های خوشه ها |
| FCM | عددی | $O(N)$ | محدب | خیر | تعداد خوشه ها | مراکز خوشه ها و وزن ها |
| DBSCAN | عددی | $O(N \log N)$ | دلخواه | بله | شعاع خوشه ها و حداقل تعداد اعضا | نسبت دادن مقادیر داده ها به خوشه ها |
| DENCLUE | عددی | $O(N \log N)$ | دلخواه | بله | شعاع خوشه ها و حداقل تعداد اعضا | نسبت دادن مقادیر داده ها به خوشه ها |

N تعداد نمونه های مجموعه داده و k تعداد خوشه ها.

3-3 معیار فاصله

در این بخش برخی معیارهای فاصله را که در فصل های بعدی برای محاسبه فاصله بین دو بردار استفاده خواهد شد، ارائه می شود. از این معیارها برای اندازه گیری فاصله بین یک نمونه و یک خوشه، فاصله بین دو نمونه و فاصله بین دو خوشه استفاده خواهد شد.

1-3-3 معیار فاصله مینکوسکی¹

فاصله مینکوسکی از مشهورترین فواصلی است که از آن در بسیاری از کاربردها استفاده می شود. این فاصله از رابطه (4.3) بدست می آید. دو گونه مشهور از فاصله مینکوسکی، که به ازای $M = 1$ و $M = 2$ استخراج می شود، به ترتیب فاصله قدر مطلق یا L_1 و فاصله اقلیدسی یا L_2 نام دارد.

$$d(F, G) = L_M(F, G) = \left(\sum_{i=1}^n |f_i - g_i|^M \right)^{\frac{1}{M}}, 1 \leq M \leq \infty \quad (4.3)$$

$$d(F, G) = L_1(F, G) = \sum_{i=1}^n |f_i - g_i| \quad (5.3)$$

$$d(F, G) = L_2(F, G) = \left(\sum_{i=1}^n |f_i - g_i|^2 \right)^{\frac{1}{2}} \quad (6.3)$$

2-3-3 معیار فاصله زاویه ای

فاصله زاویه ای دو بردار F و G با رابطه زیر تعریف می شود. اگر $F = G$ باشد، این فاصله مقدار کمینه خود یعنی صفر و اگر $F = -G$ باشد، مقدار بیشینه خود یعنی 2 را بر می گرداند.

$$d(F, G) = \frac{2}{\pi} \cos^{-1} \left(\frac{F^T G}{|F||G|} \right) \quad (7.3)$$

3-3-3 الگوریتم DTW

یکی از الگوریتم هایی که برای مقایسه و اندازه گیری میزان شباهت کلمات دست نوشته می توان به کار برد، الگوریتم انطباق زمانی پویا² (DTW) است. این الگوریتم برای تطبیق سگینال های یک بعدی به

¹ Minkowski distance measure

² Dynamic time warping

کار می‌رود. برای به کار بردن این الگوریتم در تطبیق تصاویر، باید تصاویر به سیگنال‌های یک بعدی تبدیل شوند.

الگوریتم DTW برای محاسبه فاصله بین دو مجموعه زمانی استفاده می‌شود. یک مجموعه زمانی شامل یک لیستی از نمونه‌های گرفته شده از سیگنال یک بعدی به ترتیب زمانی که نمونه‌ها بدست آمده‌اند، می‌باشد.

یک روش ساده برای محاسبه فاصله انطباق¹ بین دو مجموعه زمانی، مقایسه نمونه به نمونه است. اما این روش نتیجه شهودی و قابل درکی نمی‌دهد چون ممکن است نمونه‌هایی که با هم مقایسه می‌شوند، مشابه هم نباشند (شکل 3-10 - الف).

الگوریتم DTW با تنظیم بهینه نمونه‌ها در دو مجموعه زمانی، اختلاف بین نمونه‌های متناظر و محاسبه فاصله انطباق را کاهش می‌دهد. این تنظیم بهینه با می نیم کردن یک مجموعه‌ای از فاصله‌های محلی نمونه‌ها بدست می‌آید. (شکل 3-10 - ب)



شکل 3-10 . محاسبه فاصله انطباق دو مجموعه زمانی

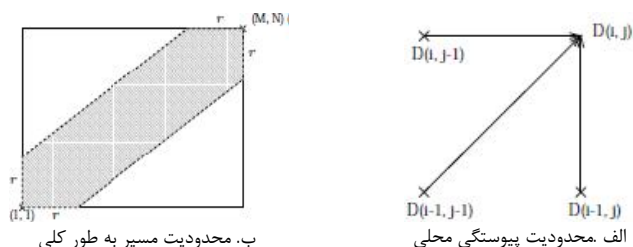
فاصله DTW بین دو مجموعه زمانی x_1, \dots, x_M و y_1, \dots, y_N ، $D(M, N)$ ، با استفاده از روش برنامه نویسی پویا² به صورت زیر محاسبه می‌شود:

$$D(i, j) = \min \left\{ \begin{array}{l} D(i, j-1) \\ D(i-1, j) \\ D(i-1, j-1) \end{array} \right\} + d(x_i, y_j). \quad (8.3)$$

انتخاب تساوی بازگشتی و تابع فاصله محلی $d(.,.)$ به نوع کاربرد بستگی دارد. با استفاده از $D(i, j-1), D(i-1, j), D(i-1, j-1)$ یک قید پیوستگی در محاسبه $D(i, j)$ ایجاد می‌شود که باعث

¹ Matching
² Dynamic programming

هموار شدن انطباق زمانی می شود (یعنی هیچ نمونه ای از سمت چپ در انطباق جا نمی افتد) (شکل 11-3 الف). مسیر انطباق، DTW، با دنبال کردن جفت شاخص های هزینه می نیمم شده $(i, j)_k$ بدست می آید. برای اینکه مسیر انطباق نزدیک به قطر ماتریس شامل $D(i, j)$ باشد، از محدودیت باند ساکو-چیبا¹ استفاده می شود (شکل 11-3 ب). اطلاعات بیشتر در مرجع [24] موجود می باشد.



شکل 11-3. محدودیت های استفاده شده برای پیاده سازی الگوریتم DTW

4-3 ماشین بردار پشتیبان چند کلاسه

در یادگیری با ناظر فرض بر این است که یک مجموعه از نمونه های ورودی $\{x_i\}_1^n$ به همراه جواب مطلوب هر نمونه $\{t_i\}_1^n$ از قبل آماده است. هدف یافتن مدل مناسبی است که بتواند خروجی داده های مشاهده نشده را به درستی پیش بینی کند. این پیش بینی بر اساس نگاشت $y(x)$ از فضای ورودی به فضای خروجی صورت می پذیرد. می توان مدل خطی زیر را برای این تابع در نظر گرفت:

$$y(x) = \sum_{i=1}^N w_i \phi_i(x) + w_0 \quad (9.3)$$

به این ترتیب فرآیند یادگیری، در یافتن پارامترهای مدل به کمک داده های آموزشی خلاصه می گردد. الگوریتم های کلاسیک یادگیری مانند شبکه عصبی RBF بر پایه حداقل کردن خطا بر روی مجموعه داده های آموزشی می باشد. برخلاف این روش ها ماشین بردار پشتیبانی مبتنی بر حداقل کردن خطای ساختاری است. ماشین بردار پشتیبان SVM دارای قدرت تعمیم پذیری بالاتری نسبت به بقیه روش -

¹ Sakoe-Chiba band constraint

های یادگیری کلاسک برای داده‌های آزمون می باشد. قدرت تعمیم پذیری فوق العاده آن ناشی از کمینه کردن خطای داده‌های آموزشی و بیشینه کردن حاشیه بین دو کلاس به صورت همزمان می باشد.

ماشین بردار پشتیبان به طور کلی یک طبقه بند دو کلاسه است. اما روش‌هایی برای استفاده از طبقه بندهای دو کلاسه برای حل مسائل چند کلاسه وجود دارد. به طور کلی دو استراتژی برای استفاده از طبقه بند دو کلاسه برای مسائل چند کلاسه وجود دارد: استراتژی یکی در مقابل همه¹ و استراتژی یکی در مقابل یکی²

در مسائل چند کلاسه، مجموعه دیتای یادگیری (x_i, c_i) شامل N نمونه متعلق به M کلاس است .

1-4-3 استراتژی یکی در مقابل همه

در این استراتژی برای هر کلاس یک SVM ساخته می‌شود به طوری که هر کلاس را از کلاس‌های دیگر جدا می‌کند. تعداد SVM‌های ساخته شده برابر با تعداد کلاس ها M است. یک نمونه آزمون بر اساس استراتژی winner-takes-all کلاسه‌بندی می‌شود. یعنی نمونه آزمون متعلق به کلاسی است که بالاترین مقدار را در تابع مشخص‌کننده³ داشته باشد. همه دیتاهای آموزش برای ساختن هر SVM استفاده می‌شود.

برای ساختن SVM کلاس k خروجی مطلوب برای داده‌های یادگیری x_i به صورت زیر تعیین می‌شود:

$$y_i = \begin{cases} +1 & \text{if } c_i = k \\ -1 & \text{if } c_i \neq k \end{cases} \quad (10.3)$$

2-4-3 استراتژی یکی در مقابل یکی

¹ One versus all

² One versus one

³ Discriminant function

در این روش، برای هر جفت کلاس از M کلاس، یک SVM ساخته می‌شود. بنابراین تعداد SVM ها برابر با $M(M-1)/2$ است. برای ساختن SVM دو کلاس (k, m) ، تنها داده‌های آموزش این دو کلاس استفاده می‌شود و خروجی مطلوب برای این داده‌های آموزش x_i به صورت زیر تعیین می‌شود:

$$y_i = \begin{cases} +1 & \text{if } c_i = k \\ -1 & \text{if } c_i = m \end{cases} \quad (11.3)$$

برای تعیین کلاس یک نمونه آزمون معمولاً از استراتژی ماکزیمم آرا¹ استفاده می‌شود. اگر مقدار تابع مشخص‌کننده SVM جفت کلاس (k, m) ، $f_{km}(x)$ ، مثبت باشد، یک رای به کلاس k در غیر این صورت یک رای به کلاس m داده می‌شود. به همین ترتیب از همه SVM ها برای دادن رای به کلاس ها استفاده می‌شود. در نهایت نمونه مورد آزمون متعلق به کلاسی است که بیشترین رای را داشته باشد.

BSVM 3-4-3

در [18] یک روشی برای مسائل چند کلاسه با حل فقط یک تابع بهینه سازی ارائه شده است. این ایده شبیه به روش یک در مقابل همه است. M قانون تصمیم‌گیری دو کلاسه که تابع $w_m^T \phi(x) + w_0$ داده‌های آموزش کلاس m را از بقیه کلاس‌ها جدا می‌کند. بنابراین M تابع تصمیم‌گیری وجود دارد. SVM مسئله بهینه سازی را به صورت زیر حل می‌کند:

$$\min_{W, B, \xi} \frac{1}{2} \sum_{m=1}^M w_m^T w_m + C \sum_{i=1}^l \sum_{m \neq y_i} \xi_i^m \quad (12.3)$$

که

$$w_{y_i}^T \phi(x_i) + b_{y_i} \geq w_m^T \phi(x_i) + b_m + 2 - \xi_i^m$$

$$\xi_i^m \geq 0 \quad \text{for } i=1, \dots, l$$

¹ majority voting

$$m \in \{1, \dots, M\} \setminus y_i$$

تابع تصمیم‌گیری به صورت زیر است:

$$\arg \max_{m=1, \dots, M} (w_m^T \varphi(x) + b_m) \quad (13.3)$$

که یک نمونه از روش یک در مقابل همه است.

فصل چهارم

پیش پردازش

فصل چهارم: پیش پردازش

1-4 مقدمه

تبدیل متون دست‌نویس نوشته شده بر روی کاغذ به فرمت تصویر دیجیتالی می‌تواند مهم‌ترین فرآیند در سیستم‌های کامپیوتری برای عرضه متون دیجیتالی، خواندن پرسش‌نامه‌ها، اتوماسیون مرسولات پستی و در نهایت کمک به دولت الکترونیک باشد.

فرآیند تبدیل متون دست‌نویس به متون دیجیتالی شامل اسکن دست‌نویس، ارتقاء کیفیت تصویر دست‌نویس، استخراج ویژگی از تصویر حاصله و نهایتاً بازشناسی متون است.

بعد از اسکن متون دست‌نویس یک دنباله‌ای از تکنیک‌ها برای بهبود کیفیت تصویر اسکن شده بر روی تصویر انجام می‌شود. این تکنیک‌ها می‌تواند بر روی دقت نهایی سیستم‌های بازشناسی متون دست‌نویس تأثیر بسزایی داشته باشند. برای ساختن یک سیستم دقیق و کاربردی در دنیای واقعی تکنیک‌های پیش پردازش باید به دقت طراحی شوند. در این بخش تکنیک‌هایی که در مرحله پیش پردازش برای بهبود کیفیت تصویر و کاهش اختلافات و پیچیدگی‌های موجود در تصاویر، در سیستم بازشناسی دست‌نویس انجام شده، بیان خواهد شد.

تکنیک‌های پیش پردازش می‌تواند شامل: ¹ باینری کردن ¹ تصویر رنگی اسکن شده، هموار سازی ² و حذف نویز، آشکارسازی و اصلاح شیب خط اصلی ³ و کجی تصویر ⁴، آنالیز مولفه‌های متصل ⁵، نرمال سازی ابعاد تصاویر ⁶ و نازک سازی ⁷ باشد.

¹ Binarization

² Smoothing

³ Detection and correction slope baseline

⁴ Slant normalization

⁵ Connected component

⁶ Size normalization

⁷ Thinning

2-4-2 باینری کردن

ممکن است متون دست‌نوشته توسط قلم‌های رنگی و یا بر روی کاغذهای رنگی نوشته شده باشند. اما پردازش متون دست‌نوشته اغلب بر روی تصاویر دوسطحی انجام می‌شود. بنابراین اولین عمل در مرحله پیش پردازش باینری کردن تصاویر رنگی اسکن شده می‌باشد. تصاویر دوسطحی متون دست‌نوشته، تصاویری هستند که پیکسل‌های مربوط به کلمه یا پیش زمینه¹ دارای سطح صفر (سیاه) و پیکسل‌های مربوط به پس زمینه² دارای سطح یک (سفید) می‌باشند. این کار توسط الگوریتم‌های آستانه‌گذاری انجام می‌شود.

روش‌های آستانه‌گذاری به طور کلی به دو دسته تقسیم می‌شوند: آستانه‌گذاری کلی و محلی. در روش‌های کلی تنها یک آستانه برای کل تصویر استفاده می‌شود. این روش برای تصاویر خاکستری که هیستوگرام چند مده یا پس زمینه غیر یکنواخت دارند، مناسب نیستند. در روش‌های آستانه‌گذاری محلی برای هر پیکسل یا گروه کوچکی از پیکسل‌ها بر اساس اطلاعات پیکسل‌های مجاور یک آستانه مجزا محاسبه می‌شود. این روش در مقایسه با روش اول کندتر و دارای محاسبات بیشتری است.

فرض پایان نامه این است که تصاویر متنی دارای پس زمینه‌ای ساده می‌باشند. هیستوگرام روشنایی این تصاویر دو مده³ است و روشنایی‌های پیش زمینه با پس زمینه نسبتاً از هم جدا هستند. بنابراین بکار بردن یک روش آستانه‌گذاری کلی نتایج مناسبی دربردارد. اما اگر تصاویر، نویزی با پیش زمینه پیچیده باشند استفاده از روش‌های آستانه‌گذاری کلی مناسب نمی‌باشد و باید از روش‌های آستانه‌گذاری محلی استفاده کرد. در ادامه دو روش شناخته شده آستانه‌گذاری دو سطحی تصاویر دست‌نوشته، بیان خواهد شد.

¹ Foreground
² background
³ Biomodal

4-2-1 روش پیشنهادی توسط Otsu

یکی از رایج‌ترین روش‌های آستانه‌گذاری کلی که در کاربردهای ماشین بینایی بسیار استفاده می‌شود، توسط Otsu و همکارانش ارائه شده است [38]. این روش را می‌توان یکی از آسان‌ترین و سریع‌ترین روش آستانه‌گذاری کلی برای کاربردهای دو کلاسه بیان کرد. الگوریتم آستانه‌گذاری مذکور بر اساس تقسیم بندی پیکسل‌های تصویر به دو کلاس شی و پس زمینه با یک آستانه بهینه t می‌باشد. انتخاب آستانه بهینه با می‌نیم کردن واریانس درون کلاس¹های بدست آمده توسط یک آستانه شناور انجام می‌شود. برای این منظور، ابتدا هیستوگرام روشنایی تصویر محاسبه می‌شود. در هیستوگرام روشنایی، مقدار $P(i)$ برابر با احتمال وجود سطح خاکستری i ام در تصویر است. برای تصویر خاکستری 8 بیتی، 256 سطح خاکستری ارزیابی می‌شود.

$$P(i) = \text{تعداد کل پیکسل ها} / \text{تعداد پیکسلها با سطح خاکستری } i$$

اگر هیستوگرام دو مُده باشد، بهترین آستانه، مقداری است که دو قله هیستوگرام را از هم جدا کند. در غیر این صورت آستانه، مقداری خواهد بود که مجموع وزن‌های واریانس‌های درونی دو کلاس تشکیل شده را حداقل کند. واریانس یک کلاس معیاری از همگنی² روشنایی در آن کلاس است. یک کلاس با واریانس بالا دارای همگنی پایین و کلاسی با واریانس پایین دارای همگنی بالا خواهد بود. بنابراین آستانه‌ای که به وسیله روش Otsu و همکارانش حاصل می‌شود نتیجه همگنی بالای کلاس‌ها را به دنبال دارد.

آستانه بهینه در این روش به صورت زیر بدست می‌آید:

اگر $\sigma_1^2(t)$ و $\sigma_2^2(t)$ به ترتیب واریانس کلاسی با مقدار کمتر یا مساوی آستانه t و واریانس کلاسی با مقدار بیشتر از t باشد و $\mu_1(t), q_1(t)$ به ترتیب احتمال و میانگین کلاس اول و $\mu_2(t), q_2(t)$ احتمال و میانگین کلاس دوم در نظر گرفته شود. واریانس درون کلاس‌ها σ_w^2 ، به صورت زیر محاسبه می‌شود:

¹Within class variance
²homogeneity

$$\sigma_w^2(t) = q_1(t) \cdot \sigma_1^2(t) + q_2(t) \cdot \sigma_2^2(t) \quad (1.4)$$

که

$$q_1(t) = \sum_{i=0}^t P(i) \quad (2.4)$$

$$q_2(t) = \sum_{i=t+1}^{255} P(i) \quad (3.4)$$

$$\mu_1(t) = \sum_{i=0}^t i \cdot P(i) / q_1(t) \quad (4.4)$$

$$\mu_2(t) = \sum_{i=t+1}^{255} i \cdot P(i) / q_2(t) \quad (5.4)$$

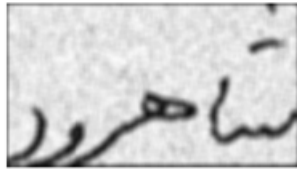
$$\sigma_1^2(t) = \sum_{i=0}^t (i - \mu_1(t))^2 P(i) / q_1(t) \quad (6.4)$$

$$\sigma_2^2(t) = \sum_{i=t+1}^{255} (i - \mu_2(t))^2 P(i) / q_2(t) \quad (7.4)$$

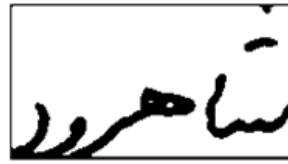
بهترین آستانه t ، مقداری است که $\sigma_w^2(t)$ را می نیمم کند .

در شکل 4-1 الگوریتم مذکور بر روی تصویر خاکستری دست نوشته پیاده شده است.

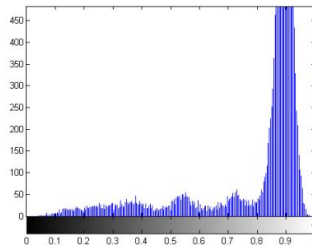
همانطور که گفته شد تصاویر استفاده شده در این پایان نامه ، تصاویری با پس زمینه یکنواخت می باشند. بنابراین الگوریتم پیشنهاد شده توسط Otsu نتیجه قابل قبولی داد. اما در شرایط واقعی ممکن است پس زمینه پیچیده و یا نویزی باشد. بنابراین در ادامه یک روش دیگری که برای تصاویر متنی با پس زمینه های پیچیده و کنتراست پایین و نویزی طراحی شده است، بیان خواهد شد.



الف. تصویر اصلی



ب. تصویر دوسطحی شده



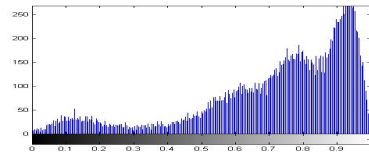
پ. هیستوگرام تصویر اصلی

شکل 4-1. به کار بردن الگوریتم Otsu بر روی تصویر دست نوشته

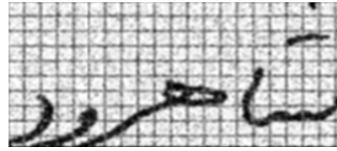
2-2-4 روش پیشنهادی توسط Manmatha و Wu

این روش یک الگوریتم دو مرحله ای است [4]. ابتدا روشنایی تصویر توسط فیلتر پایین گذر گوسی هموار می شود. این مرحله برای تصاویری که پس زمینه و پیش زمینه آن به راحتی جدا پذیر نیستند، لازم است. در تصاویر متنی، فراوانی روشنایی ها در پس زمینه بیشتر از پیش زمینه است. بنابراین فیلتر هموارسازی بیشتر بر روی پس زمینه تاثیر می گذارد و سعی می کند تا پس زمینه را یکنواخت کند. در مرحله بعد، آستانه مناسب از هیستوگرام شدت روشنایی تصویر هموار شده انتخاب می شود. در تصاویری که متن دست نوشته تاریکتر از پس زمینه است آستانه برابر با اولین دره سمت چپ هیستوگرام است و برای تصاویری که پیش زمینه تاریکتری دارد آستانه برابر با آخرین دره در نظر گرفته می شود. در هر دو مورد قبل از انتخاب آستانه، هیستوگرام شدت با کرنل گوسی برای حذف می نیم های محلی هموار می شوند.

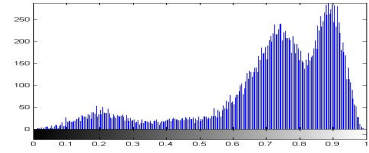
برای نشان دادن تاثیر روش Manmatha و Wu در تصاویر با پس زمینه پیچیده نسبت به روش Otsu، این دو روش بر روی تصویر با پیش زمینه شطرنجی اجرا شده اند. در روش Manmatha و Wu تصویر اصلی توسط فیلتر گوسی 5×5 با واریانس $1/44$ هموار شده است.



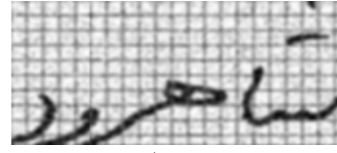
ب. هیستوگرام تصویر اصلی



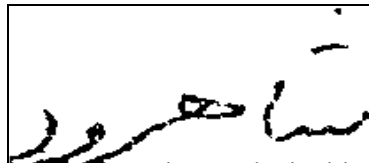
الف. تصویر دست نوشته



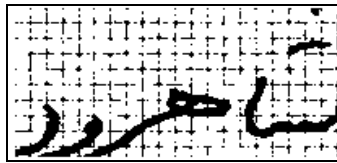
ت. هیستوگرام تصویر هموار شده



پ. تصویر هموار شده



ج. تصویر باینری شده توسط الگوریتم Wu



ث. تصویر باینری شده توسط الگوریتم Otsu

شکل 2-4. مقایسه دو روش Otsu و Wu در یک تصویر با پیش زمینه پیچیده

3-4 هموار سازی و حذف نویز

مرحله بعدی برای کاهش نویز و اصلاح لبه‌های تصاویر باینری انجام می‌شود. این مرحله برای حذف نویزهای نمک و فلفل، پر کردن شکاف‌ها، دره‌های کوچک و حذف دست اندازهای بوجود آمده در لبه‌ها مناسب می‌باشد. این مرحله توسط اعمال یک فیلتر مناسب بر روی همسایگی هر پیکسل انجام می‌شود. در این صورت مقدار هر پیکسل در تصویر خروجی تابعی از روشنایی پیکسل‌های همسایه آن در تصویر ورودی می‌باشد.

در این تحقیق برای حذف نویز و هموارسازی تصویر از فیلتر میانه استفاده شده است. گاهی اوقات چگالی نویز فلفل نمکی بوجود آمده در تصویر بزرگ است و با به کار بردن یک بار فیلتر میانه بر روی تصویر نتیجه مطلوبی حاصل نمی‌شود و اگر این فیلتر چندین بار بر روی تصویر ورودی پیاده شود ممکن است لبه‌های حروف نیز محو شده و باعث کاهش دقت شناسایی متن خواهد شد.

یکی از الگوریتم‌های موثر برای کاهش نویز به صورت زیر پیشنهاد شده است [39]:

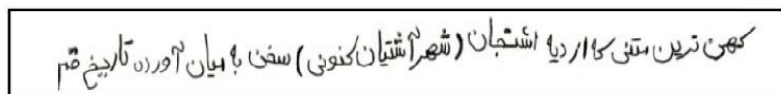
- 1- تصویر ورودی توسط فیلتر میانگین 3×3 هموار می‌شود.
- 2- اندازه گرادینان تصویر هموار شده به صورت تصویر G محاسبه می‌شود. (به کمک اپراتور سوبل)

- 3- همه پیکسل های منفرد از تصویر دو سطحی حذف می شوند.
- 4- با استفاده از تصویر G ، برای هر مولفه متصل C_i در تصویر دو سطحی میانگین اندازه گردایان پیکسل های لبه، M_i محاسبه می شود.
- 5- چنانچه میانگین M_i ها را T بنامیم. این پارامتر می تواند معیاری برای میانگین قدرت مولفه های متصل تصویر باشد.
- 6- مولفه هایی که میانگین اندازه گردایان پیکسل های لبه اش کمتر از kT باشد، حذف می شوند. (k یک پارامتر وابسته به تصویر است)

4-4 اصلاح انحراف خط اصلی

خطوط متون نوشته شده در اثر اسکن کردن یا کپی کردن ممکن است نسبت به خط افق کج شوند که به این کجی انحراف خط اصلی متن گفته می شود. زوایه این انحراف نسبت به محور افق زاویه انحراف نامیده می شود.

در حالت کلی انحراف خطوط در متن نوشته شده را می توان به سه نوع تقسیم کرد: انحراف کلی هنگامی که تمام خطوط یک متن در یک جهت باشند، انحراف چند گانه هنگامی است که تعدادی خطوط در جهتی متفاوت از خطوط دیگر باشند، انحراف غیر یکنواخت هنگامی که جهت در یک خط متنی نوسان داشته باشد که بیشتر در متون دست نوشته دیده می شود (شکل 3-4).



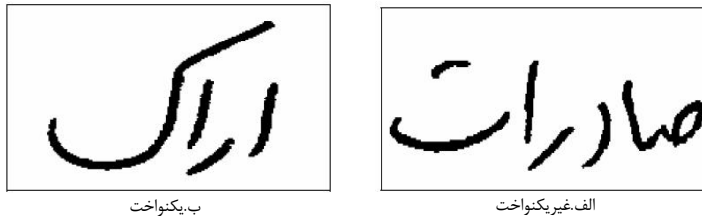
شکل 3-4. نمونه ای از انحراف غیر یکنواخت یک متن دست نوشته

در این پایان نامه بازشناسی در سطح کلمه انجام شده است. انحراف خط اصلی از خط افق در طول کلمه، زیاد محسوس نیست و بکار بردن الگوریتم های اصلاح انحراف خط اصلی در سطح کلمه باعث ایجاد انحراف های بیشتر در بعضی از کلمات کوتاه شد. بنابراین این زیرمرحله در پیش پردازش کلمات دست نوشته در این تحقیق انجام نمی شود.

4-5 اصلاح کجی

کجی، انحراف میانگین پاره‌خط‌های^۱ پیش فرض عمودی نسبت به راستای عمود می باشد. این زیر مرحله در مرحله پیش‌پردازش برای کاهش محدوده اختلافات، هم در متون دست‌نوشته و هم در متون تایپی، لازم است. کجی در متون دست‌نوشته به سبب نحوه نوشتن خاص یک فرد و در متون تایپی به دلیل خصوصیات ذاتی برخی فونت‌ها بوجود می آید. گاهی اوقات کجی‌ها بسته به ویژگی استخراج شده باعث کاهش نرخ بازشناسی می‌شود. بنابراین این مرحله در پیش‌پردازش برای حذف یا کاهش تأثیر کجی لازم است.

کجی در یک کلمه می‌تواند به دو صورت باشد: کجی یکنواخت هنگامی که پاره‌خط‌های نزدیک به خط عمودی زاویه کجی تقریباً یکسانی داشته باشند. کجی غیر یکنواخت هنگامی است که زاویه کجی در پاره‌خط‌های نزدیک به خط عمودی متفاوت باشند که بیشتر در متون دست‌نوشته دیده می‌شود (شکل 4-4).



شکل 4-4. مثال‌هایی از کجی یکنواخت و غیر یکنواخت

در تکنیک‌های اصلاح کجی یکنواخت، ابتدا زاویه کجی تخمین زده می‌شود، سپس توسط تبدیل shear افقی، تصویر نسبت به خط عمودی اصلاح می‌شود.

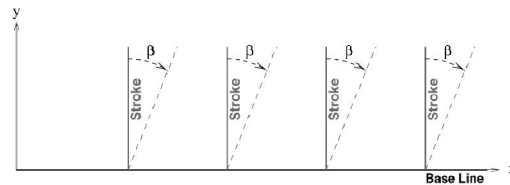
4-5-1 تبدیل shear افقی

در این تبدیل خطی هر پیکسل (x, y) به مختصات جدید (x_s, y_s) با تساوی زیر تبدیل می‌شود:

$$\begin{cases} x_s = x - (y_c - y) \cdot \tan(\beta) \\ y_s = y \end{cases} \quad (8.4)$$

^۱ stroke

y_c مرکز مختصات y و β زاویه انتقال است (شکل 4-5). با این تبدیل ارتفاع تصویر تغییر نخواهد کرد اما عرض تصویر ممکن است تغییر کند.



شکل 4-5. تبدیل shear

برای تخمین میانگین زاویه کجی، دو الگوریتم پیاده سازی شده است که هر دو الگوریتم بر روی پایگاه داده استفاده شده نتایج تقریباً یکسانی داشتند.

2-5-4 الگوریتم اصلاح کجی بر اساس پروفایل پروژکشن¹ عمودی

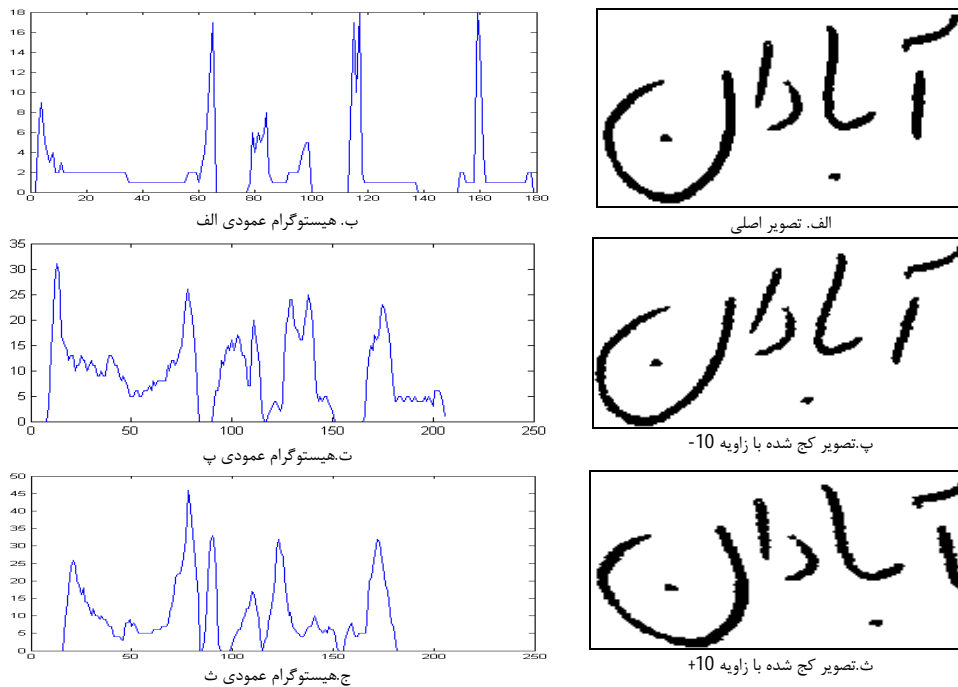
این الگوریتم بر این ایده استوار است که کلمه نوشته شده بدون کجی دارای پروفایل پروژکشن عمودی با پیکهای بلند و دره‌هایی عمیق است. پروفایل پروژکشن عمودی برابر با هیستوگرام تعداد پیکسل‌های پس زمینه در هر ستون از تصویر است. بر اساس این ایده، هر تصویر با زاویه‌های مختلف کجی با تبدیل shear چرخانده می‌شود و پروفایل پروژکشن عمودی آن کشیده می‌شود، تصویری که پیکهای بلند تر و دره‌های عمیق‌تر داشته باشد به عنوان تصویر اصلاح شده و زاویه کجی آن به عنوان زاویه اصلاح کجی در نظر گرفته می‌شود. معیارهای مختلفی برای مقایسه پروفایل‌ها پیشنهاد شده است. در [19] از معیاری به نام SD استفاده شده است. این معیار با مجموع مربع اختلاف نقاط متوالی هیستوگرام پروژکشن عمودی تصویر اصلاح شده I بدست می‌آید:

$$slant_angle = \arg \max_{\theta_{min} \leq \theta \leq \theta_{max}} SD(vertical_projection_profile(horizontal_shear(I, \theta)))$$

بازه زاویه اصلاح کجی مورد بررسی [45, -45] در نظر گرفته شده است. این روش هم برای متون دست‌نوشته و هم برای متون تایپی مناسب است و نسبت به نویز و شیوه نوشتن مقاوم است و فقط

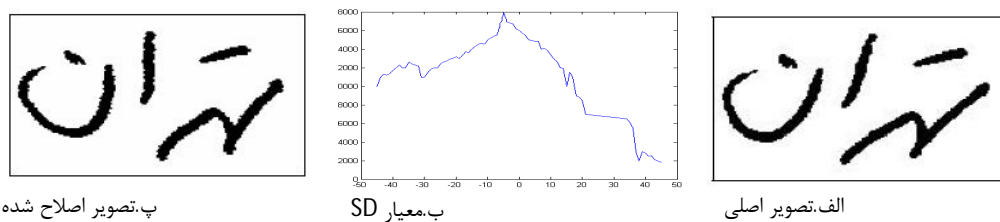
¹ projection

برای تصاویر که شامل یک کلمه هستند کاربرد دارد. شکل 4-6 هیستوگرام عمودی تصاویر با زاویه-های مختلف کجی را نشان می‌دهد.



شکل 4-6. هیستوگرام عمودی تصاویر کج شده با زاویه‌های مختلف

شکل 4-7 نشان می‌دهد که زاویه اصلاح کجی تصویر برابر با زاویه‌ای است که در آن معیار SD ماکزیمم شود.



شکل 4-7. محاسبه زاویه کجی با استفاده از معیار SD

3-5-4 الگوریتم اصلاح کجی بر اساس هیستوگرام زاویه

این الگوریتم بر اساس هیستوگرام زاویه لبه‌های تصویر کلمه می باشد [22]. مراحل اجرای این الگوریتم به صورت زیر است:

- 1- لبه‌های تصویر توسط فیلتر سوبل به ابعاد 3×3 پیدا می شود. (شکل 4-8-ب)
- 2- برای حذف خطوط افقی، هر ردیف از تصویر اسکن می شود و هر پیکسلی که به τ تا از پیکسل های مجاورش متصل باشد حذف می شود ($\tau = 3$). (شکل 4-8-پ)
- 3- مولفه‌های متصلی که تعداد پیکسل‌هایش کمتر از آستانه T است، حذف می شود. (شکل 4-8-ت)

4- بهترین خط منطبق بر هر مولفه متصل باقیمانده با استفاده از مقادیر مختصات محورهای x و y پیکسل‌های تشکیل دهنده (شکل 4-8-ث) طبق فرمول زیر محاسبه می شود:

$$y = ax + b \quad (9.4)$$

a شیب خط و b محل تقاطع با محور y می باشد. مقدار a, b توسط رگرسیون خطی به صورت زیر به دست می آید:

$$b = \frac{\left(n \sum_{i=1}^n x_i y_i \right) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\left(n \sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2} \quad (10.4)$$

$$a = \frac{\left(n \sum_{i=1}^n y_i \right) - \left(b \sum_{i=1}^n x_i \right)}{n} \quad (11.4)$$

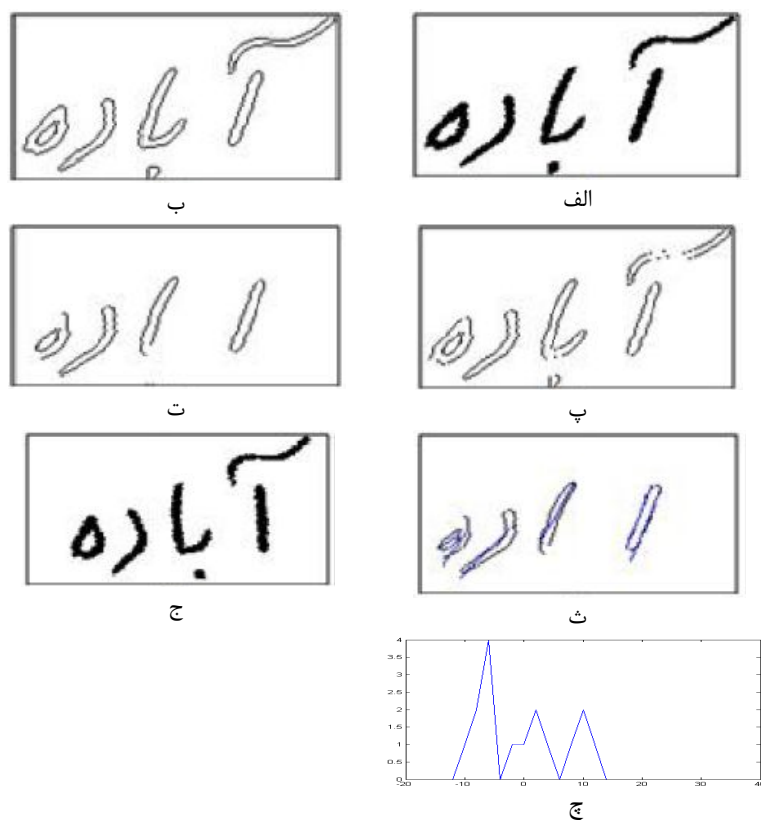
n تعداد پیکسل‌های مولفه متصل مورد بررسی و x_i مقدار مختصات محور x و y_i مقدار مختصات محور y پیکسل i ام از مولفه متصل می باشد.

5- مقدار زاویه خط منطبق بر هر مولفه متصل برابر است با:

$$\theta = \tan^{-1}(a) \quad (12.4)$$

6- هیستوگرام زاویه خطوط منطبق بر مولفه‌های متصل رسم می شود (شکل 4-8-چ).

7- مقدار زاویه اصلاح کجی تصویر برابر با زاویه‌ای که مقدار ماکزیمم را در هیستوگرام داشته باشد. در این الگوریتم زاویه کجی ذاتی بعضی حروف مثل " ر ، ز ، ژ ، و " و سرکش حروف " ک ، گ " در انتخاب مقدار زاویه کجی نهایی تأثیر می‌گذارد. برای کاهش اثر زاویه خطوط منطبق بر این حروف، تنها زاویه‌های نزدیک به خط عمودی در محاسبه هیستوگرام زاویه شرکت می‌کنند.



شکل 4-8. مراحل اجرای الگوریتم اصلاح کجی تصویر بر اساس هیستوگرام زاویه

اصلاح کجی تصویر توسط تبدیل shear معمولاً باعث ناهمواری و دنداندار شدن لبه‌های تصویر می‌شود که این ممکن است باعث ایجاد خطا شود. برای کاهش خطای ناشی از این مسئله لبه‌های تصویر با استفاده از چند قانون هموار می‌شوند [51]: (پنجره 3×3 سمت چپ، پیکسل‌های تصویر اصلی و پنجره سمت راست پیکسل‌های تصویر اصلاح شده توسط قوانین را نشان می‌دهند. 0 نشان دهنده پس زمینه و 1 نشان دهنده پیکسل متن و X نشان دهنده بی‌اهمیت است)

- قانون یک :

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

 \rightarrow

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 1 | 0 |
| 1 | 1 | 1 |

قانون دو، سه و چهار با چرخیدن ماسک قانون یک با زاویه‌های 90، 180 و 270 حاصل می‌شوند.

- قانون پنج :

| | | |
|---|---|---|
| 1 | 0 | 0 |
| 1 | 1 | 0 |
| 1 | 0 | 0 |

 \rightarrow

| | | |
|---|---|---|
| 1 | 0 | 0 |
| 1 | 0 | 0 |
| 1 | 0 | 0 |

قانون شش و هفت و هشت با چرخیدن ماسک قانون پنج به ترتیب با زاویه‌های 90، 180 و 270 حاصل می‌شوند.

- قانون نه :

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |

 \rightarrow

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 1 | 1 | 1 |

قانون ده از مکمل ماسک قانون نه بدست می‌آید.

- قانون یازده :

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | X |
| 1 | 1 | 1 |

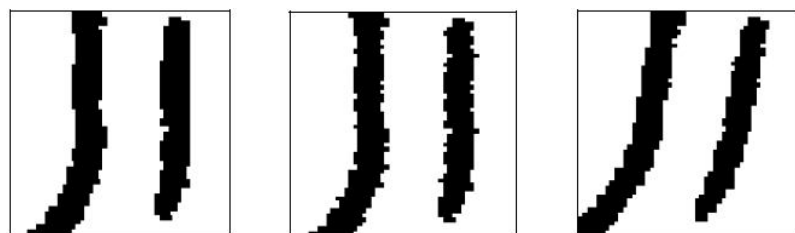
 \rightarrow

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 1 | X |
| 1 | 1 | 1 |

قانون دوازده تا هجده از چرخاندن ماسک قانون با زاویه‌های 45، 90، 135، 180، 225، 270 و 315 حاصل می‌شوند.

هر قانون، یکی پس از دیگری بر روی تمام پیکسل‌های تصویر اصلاح شده اجرا می‌شود (یعنی قانون n بر روی تصویر هموار شده توسط قانون n-1 انجام می‌شود). هیچ یک از این قانون‌ها باعث اتصال

یا شکسته شدن عناصر متصل نمی‌شوند. شکل 4-9 قسمتی از تصویر اصلاح شده با قوانین گفته شده را نشان می‌دهد.



الف. قسمتی از تصویر اصلی
ب. تصویر اصلاح شده توسط تبدیل shear
پ. تصویر هموار شده توسط قوانین
شکل 4-9. به کار بردن قوانین برای هموار کردن کردن لبه های تصویر اصلاح شده توسط تبدیل shear

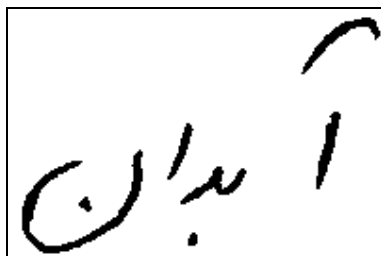
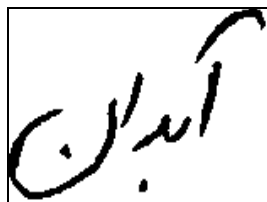
6-4 آنالیز زیرکلمات

درمتون دست نوشته تنوع در شیوه نوشتن بسیار زیاد است و این باعث پیچیده شدن فرآیند بازشناسی کلمات می‌شود. مرحله پیش پردازش با انجام یک سری عملیات بر روی تصاویر این تغییرات و تنوع را کاهش می‌دهد. یکی از این عملیات، آنالیز زیرکلمات است.

در بعضی از دست‌نوشته‌ها فاصله بین زیر کلمات بسیار زیاد است. در این مرحله برای یکسان کردن فاصله بین زیرکلمات، در صورت وجود فضای خالی در بین زیرکلمات، فضای خالی حذف می‌شود (شکل 4-10).

در متون دست‌نوشته، بعضی از حروف یا زیرکلمات در زیر حروف یا زیرکلمات دیگر قرار می‌گیرند بدون اینکه به هم متصل شوند و این ممکن است در استخراج بعضی ویژگی‌ها خطا ایجاد کند، برای حل این مشکل از الگوریتم کشیدن¹ استفاده شده است [3].

¹ stretching



ب

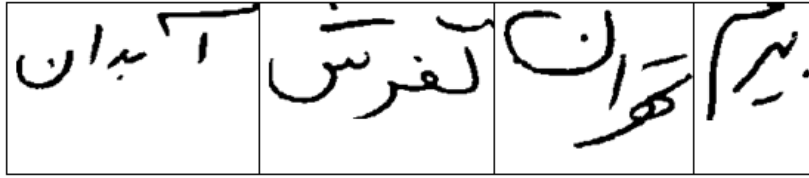
الف

شکل 4-10. حذف فضای خالی بین زیر کلمات

4-6-1 الگوریتم کشیدن

هدف از اجرای این الگوریتم حذف هم‌پوشانی‌های ایجاد شده در بعضی از زیر کلمات تصویر دست-نوشته می باشد. اولین مرحله این الگوریتم پیدا کردن خط اصلی کلمه دست نوشته است. خط اصلی یک کلمه، خطی فرضی است که بر روی آن حروف تشکیل دهنده کلمه به یکدیگر متصل می‌شوند. یکی از رایج ترین روش‌ها برای پیدا کردن خط اصلی یک کلمه بر اساس پروفایل پروژکشن افقی است [9]. به این صورت که خطی که بیشترین مقدار را در پروفایل پروژکشن افقی داشته باشد، به عنوان خط اصلی شناخته می شود. بیشترین مقدار در پروفایل پروژکشن افقی ردیفی با بیشترین تعداد پیکسل را نشان می‌دهد. البته در برخی دست‌نوشته‌ها بعضی حروف یا علامت‌ها مثل کلاه روی الف " آ " یا نقاط کشیده شده‌اند (شکل 4-11). کشیدن این حروف باعث خطا در پیدا کردن خط اصلی در روش مبتنی بر پروفایل پروژکشن افقی می‌شود. برای کاهش این خطا از یک سری شرایط و قواعد در پیدا کردن خط اصلی همراه با ماکزیمم مقدار پروفایل پروژکشن افقی استفاده شد.

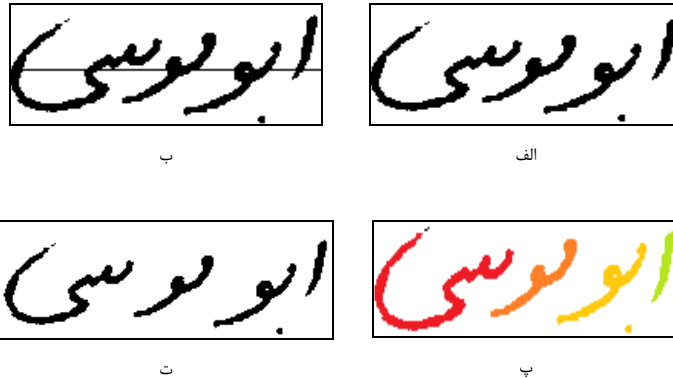
مرحله بعدی این الگوریتم تعیین مولفه‌های متصل اصلی کلمه که همان زیرکلمات می باشند، است. برای این کار بر روی خط اصلی حرکت کرده و مولفه‌های متصلی که با خط اصلی یا خطوط نزدیک به خط اصلی برخورد و تعداد پیکسل‌هایشان بزرگتر از میانگین تعداد پیکسل‌های عناصر متصل آن کلمه است، به عنوان مولفه‌های متصل اصلی کلمه در نظر گرفته می شوند.



شکل 4-11. کشیدن بعضی حروف باعث ایجاد خطا در پیدا کردن خط اصلی می شود .

مولفه‌های متصل غیر اصلی (مثل نقاط، علائم و مکمل‌ها) به آن مولفه اصلی که مرکز ثقل‌اش کمترین فاصله را با مرکز ثقل مولفه غیر اصلی دارد، تعلق دارد.

مرحله بعدی گذاشتن فاصله بین مولفه‌های متصل اصلی است. شکل 4-12 مراحل اجرای الگوریتم کشیدن را نشان می دهد.



شکل 4-12. مراحل اجرای الگوریتم کشیدن

7-4 نازک سازی

یکی از مهمترین مراحل در پیش پردازش مرحله نازک سازی یا اسکلت بندی است. نازک سازی یعنی تبدیل شکل به منحنی‌ها و خطوطی با ضخامت یک پیکسل می باشد به گونه‌ای که ویژگی‌های ساختاری شکل حفظ شود. نازک سازی با یکسان کردن ضخامت قلم، در دقت سیستم‌های بازشناسی کلمه تاثیر دارد.

تکنیک‌های نازک سازی به طور کلی به دو دسته تقسیم می شوند: روش‌های تکراری و غیر تکراری. در روش تکراری، اسکلت تصویر با حذف پیکسل‌های قابل حذف به صورت تکراری بدست می‌آید. در این الگوریتم‌ها، همه پیکسل‌ها در تمام پریودهای تکرار با یک ترتیب ثابت بررسی می‌شوند. حذف

یک پیکسل در یک پرپود تکرار به تصویر بدست آمده از پرپود قبلی و عملیات‌های انجام شده بر روی آن پیکسل وابسته است.

در روش‌های غیر تکراری، اسکلت تصویر از اتصال پیکسل‌هایی که دارای شرایط از پیش تعیین شده-ای هستند، حاصل می‌شود.

الگوریتم‌های نازک سازی بر اساس این فرض هستند که پیکسل‌های درون هر ناحیه از تصویر به هم متصل‌اند و هیچ پراکندگی در میان آنها نیست. این الگوریتم‌ها در تصاویری با ناحیه‌های پراکنده ضعیف عمل می‌کنند. پراکندگی در درون ناحیه‌های تصویر می‌تواند ناشی از کهنگی تصویر، روشنایی نامنظم یا آستانه گذاری باشد.

برای نازک سازی تصاویر کلمه دست نوشته زبان لاتین الگوریتم‌های زیادی پیشنهاد شده است که بعضی از آنها برای کلمات دست نوشته فارسی مناسب می‌باشند. در این پایان نامه دو روش پیاده سازی و مقایسه شده است:

1-7-4 الگوریتم پیشنهادی توسط Haung

الگوریتم پیشنهاد شده توسط Haung و همکارانش، یک الگوریتم نازک سازی تکراری می‌باشد [50]. این الگوریتم از اطلاعات بدست آمده از ماسک‌های 3×3 ، 4×3 ، 3×4 ، 4×4 استفاده می‌کند. این الگوریتم در مقابل نویز مقاوم است.

پیکسل‌های ناحیه پس زمینه (پیکسل‌های سفید) مقدار صفر و پیکسل‌های پیش زمینه (پیکسل‌های سیاه) مقدار یک هستند.

هم جواری هشت گانه پیکسل p توسط ماسک 3×3 در شکل 4-13 آورده شده است.

| | | |
|-------|-------|-------|
| n_3 | n_2 | n_1 |
| n_4 | p | n_0 |
| n_5 | n_6 | n_7 |

شکل 4-13. همسایه هشت تایی پیکسل p

قانون‌های زیر به طور همزمان برای هر پیکسل تصویر برای حذف یا باقی ماندن آن در عمل نازک سازی اجرا می شود:

- اگر تعداد پیکسل‌های سیاه در همسایگی p ، صفر یا یک یا هشت باشند، پیکسل p حذف نمی‌شود.
- اگر پیکسل p دارای دو پیکسل سیاه در همسایه‌اش باشد در صورتی حذف می شود که دو پیکسل کنار هم باشند مثلاً n_0 و n_1 یا n_1 و n_2 یا ... سیاه باشند.
- چنانچه سه پیکسل سیاه متصل در همجواری هشت تایی پیکسل p و یا مانند الگوهای زیر باشند، پیکسل p حذف می شود.

| | | | | | | | | | | | |
|---|-----|---|---|-----|---|---|-----|---|---|-----|---|
| 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| 1 | p | 0 | 0 | p | 1 | 0 | p | 1 | 1 | p | 0 |
| 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

- چنانچه در همسایگی هشت تایی پیکسل p ، چهار یا پنج یا شش پیکسل سیاه باشد در صورتی که این پیکسل‌ها پشت سر هم و یا مانند الگوهای زیر باشند، پیکسل p حذف می‌شود.

| | | | | | |
|---|-----|---|---|-----|---|
| 1 | 1 | 0 | 0 | 1 | 1 |
| 0 | p | 1 | 1 | p | 0 |
| 0 | 0 | 1 | 1 | 0 | 0 |

- اگر پیکسل p ، هفت پیکسل سیاه در همجواری هشت گانه‌اش و یک پیکسل سفید در همجواری چهارگانه‌اش باشد، پیکسل p حذف می شود.

با این قوانین اشکال مستطیلی با عرض دو پیکسل حذف می‌شوند. برای جلوگیری از دست دادن اطلاعات اصلی، در صورتی که همسایگی پیکسل p مانند یکی از الگوهای زیر باشند، پیکسل p حذف نمی‌شود.

| | | |
|---|---|---|
| x | 0 | x |
| 1 | p | 1 |
| 1 | 1 | 1 |
| x | 0 | x |

| | | |
|---|---|---|
| x | 0 | 0 |
| 1 | 1 | 0 |
| 0 | p | 0 |
| 0 | 0 | x |

| | | |
|---|---|---|
| x | 0 | 0 |
| 0 | p | 0 |
| 0 | 1 | 1 |
| 0 | 0 | x |

| | | | |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | p | 1 | 0 |
| 0 | 1 | 1 | 0 |
| 0 | 0 | 0 | 0 |

| | | | |
|---|---|---|---|
| x | 0 | 0 | 0 |
| 0 | p | 1 | 0 |
| 0 | 0 | 1 | x |

| | | | |
|---|---|---|---|
| x | 1 | 1 | x |
| 0 | p | 1 | 0 |
| x | 1 | 1 | x |

| | | | |
|---|---|---|---|
| 0 | 0 | 0 | x |
| 0 | 1 | p | 0 |
| x | 1 | 0 | 0 |

بعد از اینکه یکبار این قوانین بر روی تمام پیکسل‌های تصویر اجرا شد، الگوریتم دوباره تکرار می‌شود تا اینکه که هیچ پیکسل دیگری را نتوان حذف نمود.

2-7-4 الگوریتم پیشنهادی توسط Hilditch و همکارانش

بر روی هر پیکسل p از تصویر کلمه، قوانین زیر اجرا می‌شود [30]. این قوانین بر اساس پیکسل‌های همجواری هشت‌گانه پیکسل p و دو تابع زیر تعیین می‌شوند. پیکسل‌های ناحیه پس زمینه (پیکسل‌های سفید) مقدار صفر و پیکسل‌های پیش زمینه (پیکسل‌های سیاه) مقدار یک هستند.

$$A(p) = \text{تعداد انتقال‌ها از صفر به یک در دنباله } n_2, n_3, n_4, n_5, n_6, n_7, n_0, n_1, n_2$$

$$B(p) = \text{تعداد پیکسل‌های سفید}$$

پیکسل‌های شی‌ای که چهار شرط زیر را داشته باشند حذف می‌شوند:

- $2 \leq B(p) \leq 6$: این شرط از دو زیر شرط تشکیل شده است که شرط اول ($2 \leq B(p)$) باعث می‌شود که پیکسل‌های تنها و انتهایی حذف نشود و زیر شرط دوم ($B(p) \leq 6$) نشان می‌دهد که پیکسل p یک پیکسل مرزی است.
- $A(p) = 1$: این شرط اتصال را بررسی می‌کند. اگر $A(p) > 1$ باشد با تغییر ارزش p از سیاه به سفید ناپیوستگی ایجاد خواهد شد.
- $A(n_2) \neq 0$ یا $n_2, n_4, n_0 = 0$: این شرط نشان می‌دهد که خطوط عمودی با پهنای دو پیکسل ساییده نمی‌شود.
- $A(n_4) \neq 1$ یا $n_2, n_4, n_6 = 0$: این شرط نشان می‌دهد که خطوط افقی با پهنای دو پیکسل ساییده نمی‌شود.

این شرایط بر روی تمام پیکسل‌های تصویر اجرا می‌شود و پیکسل‌هایی که همه چهار شرط را داشته باشند سفید (پس زمینه) می‌شوند. این فرآیند تکرار می‌شود تا اینکه مقدار پیکسلی تغییر نکند. شکل 4-14 نتایج پیاده سازی دو الگوریتم را بر روی تصویر کلمه نشان می‌دهد.



شکل 4-14. الف- تصویر کلمه. ب- اسکلت کلمه با الگوریتم hilditch پ- اسکلت کلمه با الگوریتم haung

8-4 نرمال سازی ابعاد

نرمال سازی ابعاد تصاویر یکی از مهمترین مراحل پیش پردازش می‌باشد. در این مرحله ابعاد همه تصاویر کلمات یکسان می‌شوند تا بردار ویژگی استخراج شده از هر کلمه ابعاد یکسانی داشته باشد. روش‌های نرمال سازی ابعاد تصاویر به دو دسته تقسیم می‌شوند: نرمال سازی خطی و نرمال سازی غیر خطی

اگر عرض تصویر w و ارتفاع آن h باشد نسبت ابعاد¹ تصویر برابر است با :

$$R = \frac{\min(w, h)}{\max(w, h)} \quad (13.4)$$

بر اساس استراتژی نرمال سازی تطبیقی نسبت ابعاد²[30]، نسبت ابعاد تصویر نرمال R_2 به طور تطبیقی بر اساس نسبت ابعاد تصویر ورودی R_1 با استفاده از تابع‌های نگاشت جدول 4-1 محاسبه می‌شود.

تصویر ورودی به صفحه‌ای مربعی با طول از پیش تعیین شده L نگاشت می‌یابد. تصویر نرمال در وسط صفحه قرار می‌گیرد به طوری که بزرگترین بعد تصویر نرمال برابر با L و بعد دیگر توسط نسبت ابعاد R_2 بدست می‌آید (شکل 4-15)

¹ Aspect ratio

² Aspect ratio adaptive normalization(ARAN)

جدول 4-1. تابع های نگاشت برای نسبت ابعاد

| Method | Function |
|-----------------------------|---|
| Fixed aspect ratio | $R_2 = 1$ |
| Aspect ratio preserved | $R_2 = R_1$ |
| Square root of aspect ratio | $R_2 = \sqrt{R_1}$ |
| Cubic root of aspect ratio | $R_2 = \sqrt[3]{R_1}$ |
| Sine of aspect ratio | $R_2 = \sqrt{\sin\left(\frac{\pi}{2} R_1\right)}$ |



ب

الف

شکل 4-15. الف - تصویر ورودی ب - تصویر نرمال در صفحه مربعی L

مختصات تصویر ورودی توسط تبدیل های نگاشت جدول زیر به مختصات تصویر نرمال تبدیل می شود.

جدول 4-2. نگاشت مختصات روشهای نرمال سازی مختلف

| Method | Forward mapping | Backward mapping |
|--------|---|---|
| linear | $x' = \alpha x$ $y' = \beta y$ | $x = x' / \alpha$ $y = y' / \beta$ |
| moment | $x' = \alpha(x - x_c) + x'_c$ $y' = \beta(y - y_c) + y'_c$ | $x = (x' - x'_c) / \alpha + x_c$ $y = (y' - y'_c) / \beta + y_c$ |
| slant | $x' = x - (y - y_c) \tan \theta$ $y' = y$ | $x = x' + (y - y_c) \tan \theta$ $y = y'$ |

(x, y) مختصات و (h_1, w_1) عرض و ارتفاع تصویر ورودی، (x', y') مختصات و (h_2, w_2) عرض و ارتفاع

تصویر نرمال می باشد و :

$$\alpha = w_2 / w_1 \text{ و } \beta = h_2 / h_1 \quad (14.4)$$

$$x_c = m_{10} / m_{00} \text{ و } y_c = m_{01} / m_{00} \quad (15.4)$$

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (16.4)$$

$$\tan \theta = \frac{\mu_{11}}{\mu_{02}} \quad (17.4)$$

$$\mu_{pq} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q f(x, y) \quad (18.4)$$

با توجه به نتایج پیاده سازی روش‌های مختلف نرمال سازی، روش گشتاور با تبدیل نسبت ابعاد سینوسی بهترین جواب را داده است.

الگوریتم پیشنهادی و نتایج ارزیابی آن

فصل پنجم: الگوریتم پیشنهادی و نتایج ارزیابی آن

1-5 مقدمه

در این فصل الگوریتم پیشنهاد شده برای بازشناسی کلمات دست‌نوشته فارسی به تفصیل بیان خواهد شد. این الگوریتم بر روی پایگاه داده ایران شهر مورد ارزیابی قرار گرفته است. این پایگاه داده شامل 16000 تصویر دست‌نوشته از 503 نام شهرهای ایران می‌باشد. برای هر کلمه حداقل 25 نمونه از دست‌نوشته‌های افراد مختلف وجود دارد. برای هر کلمه از این پایگاه داده 80% از نمونه‌هایش برای مجموعه آموزش و 20% آن برای آزمون به صورت تصادفی انتخاب شده‌اند.

الگوریتم پیشنهاد شده در این تحقیق شامل دو مرحله اصلی می‌باشد. در مرحله اول شهرهای نامتشابه با کلمه مورد آزمون، از مجموعه کلمه‌های فرهنگ لغت حذف می‌گردد. برای این منظور در مرحله آموزش کلمه‌ها خوشه‌بندی شده تا کلمات مشابه و نامتشابه تفکیک شوند. به این مرحله کاهش فرهنگ لغت¹ گفته می‌شود. مرحله دوم بازشناسی کلمه آزمون ورودی از زیر مجموعه کلمات کاندید بدست آمده در مرحله اول است.

در ادامه نحوه انجام هر یک از این مراحل بیان خواهد شد.

2-5 مرحله اول: کاهش فرهنگ لغت

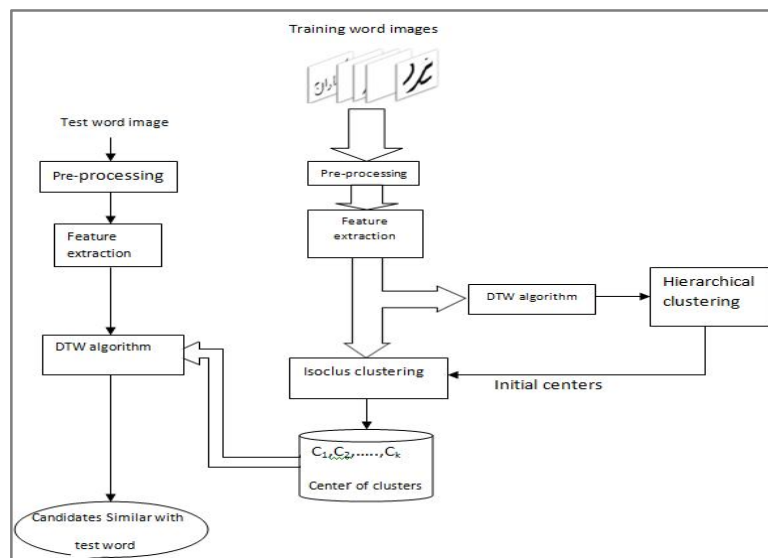
همان طور که در فصل اول بیان شد هر چه تعداد کلمات موجود در فرهنگ لغت بیشتر باشد به همان اندازه دقت و سرعت سیستم‌های بازشناسی کلمه کاهش می‌یابد. بعضی از سیستم‌ها برای کاهش مشکلات ناشی از فرهنگ لغت بزرگ، در ابتدا تعداد کلمات مورد بررسی برای بازشناسی کلمه تست ورودی را کاهش می‌دهند. این کار نه تنها سرعت سیستم را بالا می‌برد بلکه دقت را نیز افزایش می‌دهد. در این تحقیق، برای کاهش فرهنگ لغت از خوشه بندی تصاویر کلمات دست‌نوشته

¹ Lexicon reduction

استفاده شده است. ابتدا ویژگی‌هایی که شکل کلی کلمه را توصیف می‌کنند از تصاویر استخراج می‌شوند. سپس از توالی الگوریتم‌های سلسله‌مراتبی و ISOCLUS برای خوشه‌بندی تصاویر کلمات دست‌نوشته استفاده می‌شود.

در شکل 5-1 روند مراحل خوشه‌بندی تصاویر نشان داده شده است. همانطور که در شکل آمده است، این مراحل شامل پیش‌پردازش، استخراج ویژگی و خوشه‌بندی توسط الگوریتم سلسله‌مراتبی و ISOCLUS می‌باشد.

در مرحله پیش‌پردازش برای افزایش کیفیت تصاویر، کاهش تغییرات و اختلافات در بین تصاویر دست‌نوشته‌های هم‌خانواده یک سری عملیات بر روی تصاویر انجام می‌شود. این عملیات شامل باینری کردن، هموارسازی و حذف نویز، آنالیز مولفه‌های متصل، اصلاح کجی تصویر، نازک‌سازی و نرمال‌سازی اندازه تصاویر می‌باشد. روش‌های انجام هر یک از این عملیات در فصل چهارم آورده شده است.



شکل 5-1. مراحل خوشه‌بندی تصاویر کلمات دست‌نوشته

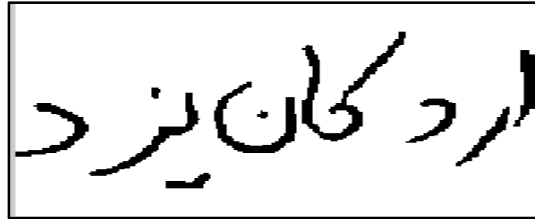
در مرحله کاهش فرهنگ لغت، از ویژگی‌هایی استفاده شده است که شکل ظاهری کلمه را به طور کلی توصیف کنند و سریع استخراج می‌شوند. برای این منظور از ویژگی‌های پروفایل بالایی و پایینی و پروفایل پروژکشن عمودی و پروفایل گذر سیاه به سفید استفاده شده است.

ویژگی‌های پروفایل بالا (و پایین)، ساختار بیرونی شکل کلمه را از بالا (و پایین) بیان می‌کنند. این ویژگی‌ها با اندازه‌گیری فاصله اولین پیکسل سیاه تا مرز بالایی (و پایینی) تصویر در هر ستون بدست می‌آیند. تعداد ویژگی‌ها برای هر کدام از بردارهای معرفی شده، برابر با تعداد ستون‌های تصویر است (شکل 2-5 الف و ب).

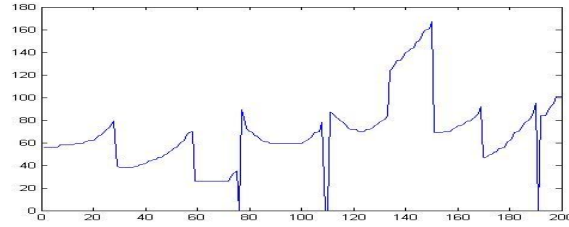
ویژگی پروفایل پروژکشن عمودی توزیع پیکسل‌های سیاه (پیکسل‌های قلم) را در راستای عمودی تصویر نشان می‌دهد و با شمردن تعداد پیکسل‌های سیاه در هر ستون بدست می‌آیند (شکل 2-5 ت).

سه ویژگی بیان شده ساختار بیرونی شکل کلمه را توصیف می‌کند. برای بیان ساختار داخلی شکل کلمه از ویژگی پروفایل گذر سیاه به سفید استفاده شده است. این ویژگی در هر ستون تعداد گذر ها از سیاه به سفید را نشان می‌دهد (شکل 2-5 ث).

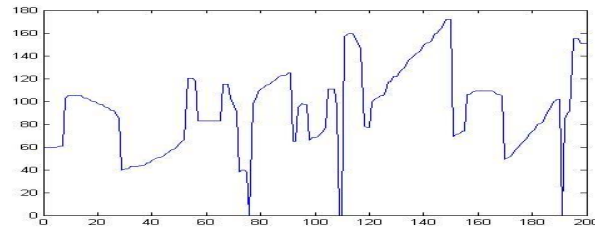
بنابراین از هر ستون تصویر کلمه دست نوشته چهار ویژگی استخراج می‌شود. تعداد کل ویژگی‌ها برابر با $4 \times m$ است که m تعداد ستون‌های تصویر کلمه است. در بخش پیش پردازش اندازه همه تصاویر، به ابعاد 200×200 نرمال شد. بنابراین تعداد کل ویژگی‌ها از تصویر نرمال شده $200 = 800$ $\times 4$ می‌باشد. با توجه به بعد بالای بردار ویژگی، تعداد ویژگی‌ها به کمک تبدیل موجک یک بعدی کاهش داده می‌شود. تبدیل موجک بر روی هر یک از چهار بردار ویژگی اشاره شده اعمال می‌شود. به کار بردن تبدیل موجک بر روی بردارهای ویژگی نه تنها ابعاد ویژگی‌ها را کاهش می‌دهد بلکه لرزش‌های موجود در لبه‌های تصویر را که در اثر نوشتن و یا اسکن کردن یا باینری کردن بوجود آمده‌اند را، بدون تغییر در ساختار کلی، کاهش می‌دهد. در شکل 2-5 نمونه‌ای از ویژگی‌های پروفایل بالا و پایین، پروژکشن عمودی و گذر سیاه به سفید استخراج شده از یک تصویر کلمه دست نوشته را نشان می‌دهد. شکل 3-5 پروفایل بالا را با به کار بردن تبدیل موجک بر روی آن نشان می‌دهد.



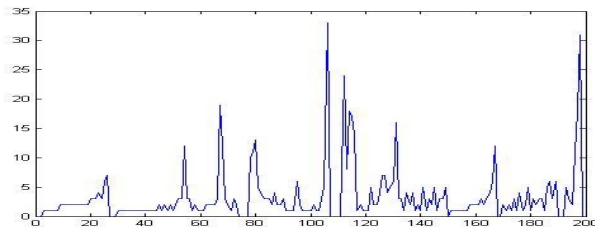
(الف)



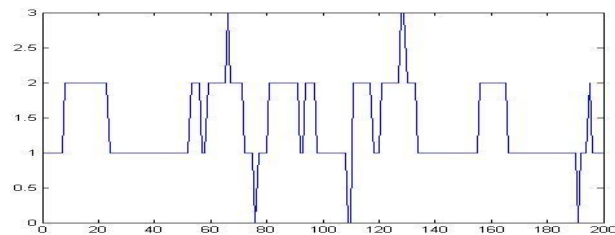
(ب)



(پ)

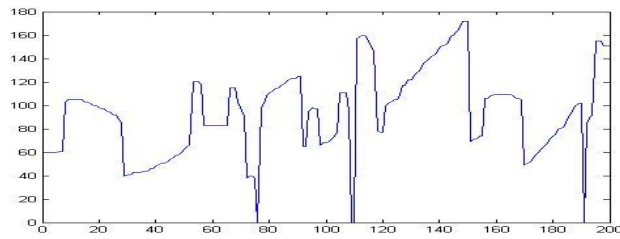


(ت)

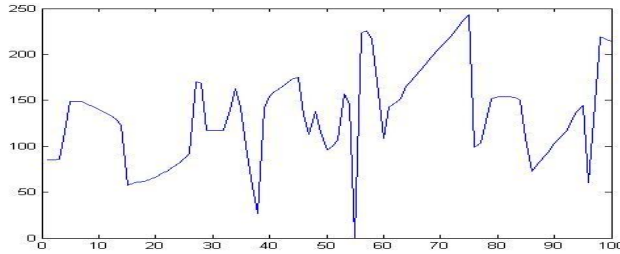


(ث)

شکل 5-2. (الف) یک نمونه از کلمه دست نوشته (ب) پروفایل پایین (پ) پروفایل بالا (ت) پروفایل پروژکشن عمودی (ث) پروفایل گذر سیاه به سفید



الف



ب

شکل 3-5. (الف) پروفایل بالا (ب) بردار ضرایب تقریب تبدیل موجک

بعد از استخراج ویژگی‌ها از تمام تصویر، مرحله بعدی، مرحله گروه‌بندی تصاویر بر مبنای شکل ظاهریشان که بوسیله ویژگی‌های بیان شده در بالا توصیف می‌شوند، است. در این مرحله از الگوریتم خوشه بندی سلسله مراتبی و ISOCLUS استفاده می‌شود.

1-2-5 الگوریتم ISOCLUS

یکی از الگوریتم‌های معروف خوشه بندی، الگوریتم ISOCLUS است که در بسیاری از کاربردها به طور گسترده از آن استفاده می‌شود [26]. این الگوریتم بر مبنای الگوریتم ISODATA با کمی اصلاحات می‌باشد و به طور کلی شبیه به الگوریتم k-میانگین است با این تفاوت که در الگوریتم k-میانگین تعداد خوشه‌ها ثابت است اما در الگوریتم ISOCLUS با حذف، تقسیم یا ادغام خوشه‌ها در هر مرحله، ممکن است تعداد خوشه‌های نهایی با تعداد خوشه‌های اولیه متفاوت باشند.

هدف الگوریتم ISOCLUS پیدا کردن بهترین مرکز خوشه‌ها با یک روش تکراری است. ورودی‌های این الگوریتم شامل: مجموعه داده آموزش، تعداد اولیه خوشه‌ها و یک سری پارامترهای از پیش تعیین شده برای حذف، تقسیم یا ادغام خوشه‌ها می‌باشند. این پارامترها توسط کاربر تعیین می‌شود و شامل:

- می نیمم تعداد نمونه‌ها در هر خوشه (SampRn)
- ماکزیمم تعداد تکرار (MaxIter)
- ماکزیمم انحراف استاندارد برای هر خوشه (StdDev)
- ماکزیمم تعداد جفت خوشه‌هایی که در هر تکرار می توانند ادغام شوند (MaxPair)
- ماکزیمم تعداد خوشه‌ها (MaxNum)
- می نیمم فاصله بین مرکز خوشه‌ها (MinDis)

مراحل اجرای الگوریتم ISOCLUS به صورت زیر است [26]:

مرحله اول: وارد کردن مقدار پارامترها به الگوریتم (مقدار پارامترها به صورت سعی و خطا بدست می آید)

مرحله دوم: فاصله هر داده تا مراکز خوشه‌ها محاسبه می‌شود و هر داده به نزدیکترین خوشه بر چسب می‌خورد.

مرحله سوم: خوشه‌هایی که تعداد نمونه‌هایش کمتر از SampRn باشد حذف می‌شوند.

مرحله چهارم: مرحله دوم و سوم دوباره تکرار می‌شود تا اینکه هیچ خوشه‌ای در مرحله سوم حذف نشود.

مرحله پنجم: میانگین فاصله نقاط هر خوشه تا مرکز آن خوشه و میانگین کلی فاصله محاسبه می‌شود.

مرحله ششم: برای هر خوشه، انحراف معیار هر بعد و ماکزیمم انحراف معیار محاسبه می‌شود. با توجه به پارامترهای StdDev و MaxNum، خوشه‌های بزرگ از بعدی که بالاترین انحراف معیار را دارد نصف می‌شوند.

مرحله هفتم: اگر در مرحله ششم خوشه‌ای نصف شد به مرحله دوم می‌رود در غیر این صورت از پارامترهای MinDis, MaxPair برای ادغام خوشه‌های نزدیک استفاده می‌شود.

مرحله هشتم: با توجه به مقدار پارامترهای MaxIter, MaxNum تعیین می‌شود که به مرحله دو بازگردد یا الگوریتم تمام شود.

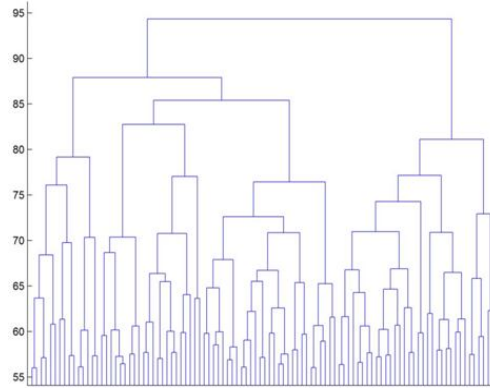
تعداد و مراکز اولیه خوشه‌ها در نتیجه پایانی تأثیر دارد و انتخاب اشتباه باعث ایجاد نتایج اشتباه و غیر معتبر خواهد شد. به همین دلیل از الگوریتم سلسله مراتبی که مستقل از پارامتر است، برای تعیین تعداد اولیه خوشه‌ها استفاده شده است. با این کار هم می‌توان تعداد خوشه‌های مناسب را تخمین زد و هم مراکز اولیه برای الگوریتم ISOCLUS آماده می‌شود.

2-2-5 الگوریتم سلسله مراتبی

در این الگوریتم پس از استخراج ویژگی از تمام نمونه‌ها، در مرحله اول فاصله‌ی بین تمام نمونه‌ها بر اساس معیار فاصله محاسبه می‌شود. در مرحله دوم از روی مقدار فاصله‌ها، نمونه‌های نزدیک به هم پیدا شده تا تشکیل یک خوشه بدهند. در دوره بعدی بردارهای میانگین خوشه‌های جدید به عنوان ورودی در نظر گرفته شده، مرحله اول و دوم تکرار می‌شود. باز هم خوشه‌های نزدیک به هم با یکدیگر ادغام می‌شوند. در حین اجرای این فرایند نمودار دندروگرام¹ تشکیل می‌شود که شاخه‌های آن بیانگر خوشه‌هایی است که ادغام شده‌اند و ارتفاع شاخه‌ها بیانگر میزان فاصله‌ی بین خوشه‌های ادغام شده است.

اگر خوشه‌یابی سلسله مراتبی تا جایی که هیچ خوشه‌ای باقی نماند، ادامه یابد، نمودار درختی شبیه نمودار شکل 4-5 حاصل می‌شود. در این نمودار خوشه‌های بسیار کوچک اولیه نشان داده نشده‌اند.

¹ Dendrogram



شکل 5-4. نمودار دندروگرام کامل مربوط به یکی از آزمایشات انجام شده

با نگاه به این نمودار و دقت در ارتفاع شاخه‌های حاصله می‌توان تشخیص داد که در کدام سطح فاصله اگر الگوریتم متوقف شود خوشه‌های مناسبی به دست خواهد آمد. در اینجا اگر الگوریتم بعد از سطح 70 روی محور فاصله متوقف شود، خوشه‌هایی بوجود خواهد آمد که فاصله بین نمونه‌های آنها به طور شهودی معقول بوده و اندازه‌ی خوشه‌ها نیز مناسب است.

برای اندازه‌گیری فاصله بین نمونه‌ها از الگوریتم DTW استفاده شده است. دلیل استفاده از این الگوریتم وجود طول‌های متفاوت حروف در کلمات مشابه در نوشته‌های افراد مختلف است (شکل 5-5). الگوریتم DTW سعی می‌کند تا حدودی با استفاده از فاصله نقاط مجاور این تغییرات را جبران کند.



شکل 5-5. تصاویر یک کلمه با دست نوشته‌های متفاوت

برای انجام این کار، همانطور که ذکر شد از هر تصویر چهار نوع ویژگی استخراج می‌شود و به صورت زیر با هم ترکیب می‌شوند تا یک مجموعه زمانی با نمونه‌های چند متغیره حاصل شود.

برای تصویر کلمه I مجموعه زمانی $X(I) = x_1, x_2, \dots, x_z$ به صورت زیر بدست می‌آید:

$$x_i = (f_1, f_2, f_3, f_4)^T$$

برای ویژگی‌های استخراج شده از تصویر است. بنابراین هر مجموعه زمانی استخراج شده از تصویر کلمه برابر با ماتریسی به ابعاد $4 \times z$ است که z برابر با ابعاد هر ویژگی که در آزمایشات ما برابر با 100 است.

برای اجرای الگوریتم DTW بر روی دو مجموعه زمانی $X(I)$ و $Y(J)$ استخراج شده از دو تصویر I و J ، تابع فاصله محلی مورد نیاز برای محاسبه فاصله DTW به صورت زیر تعریف می‌شود:

$$d(x_i, y_j) = \sum_{k=1}^4 (f_k(I, i) - f_k(J, j))^2 \quad (1.5)$$

3-2-5 معیار عملکرد الگوریتم کاهش فرهنگ لغت

اگر یک فرهنگ لغت شامل n کلمه $L = \{l_1, l_2, \dots, l_n\}$ باشد. یک سیستم کاهش دهنده فرهنگ لغت، یک زیر مجموعه ای $L' \subset L$ از کاندیدها برای بازشناسی کلمه آزمون ورودی تولید می‌کند. کلماتی که در این زیر مجموعه نیستند ($L - L'$)، در بازشناسی کلمه آزمون ورودی شرکت نمی‌کنند.

این سیستم بر خلاف یک طبقه بند، نیازی به رتبه‌بندی کلمات موجود در زیر مجموعه بدست آمده ندارد. برای اندازه‌گیری عملکرد سیستم کاهش دهنده فرهنگ لغت در مرجع [9] تعدادی معیار معرفی شده است.

اگر اتفاق درست قرار گرفتن کلمه‌ای (l_i) از فرهنگ لغت در زیر مجموعه L' را یک متغیر تصادفی A در نظر بگیریم داریم:

$$\begin{cases} A=1 & \text{if } l_i \in L' \\ A=0 & \text{if otherwise} \end{cases} \quad (2.5)$$

مقدار کاهش سیستم با متغیر R تعریف می‌شود:

$$R = \frac{(|L| - |L'|)}{|L|} \quad (3.5)$$

سه معیار اندازه‌گیری عملکرد سیستم کاهش دهنده فرهنگ لغت:

- دقت کاهش: $\alpha = E(A)$

- درجه کاهش: $\rho = E(R)$

- درجه تأثیر کاهش: $\eta = \alpha k \rho$

که $\alpha, \rho, \eta \in [0, 1]$ است. درجه و دقت کاهش معمولاً رابطه عکس نسبت به هم دارند. این دو معیار با هم ترکیب می‌شوند و یک معیار کلی η را تولید می‌کند. ضریب k تأکید دقت کاهش نسبت به درجه کاهش را بیان می‌کند. در گزارشات علمی مقدار k حدود 0/01 در نظر گرفته شده است [1]، [9].

در فرهنگ‌لغت‌های کوچک، دقت اهمیت بیشتری دارد اما در فرهنگ‌لغت‌های متوسط و بزرگ، علاوه بر دقت، سرعت نیز مهم است.

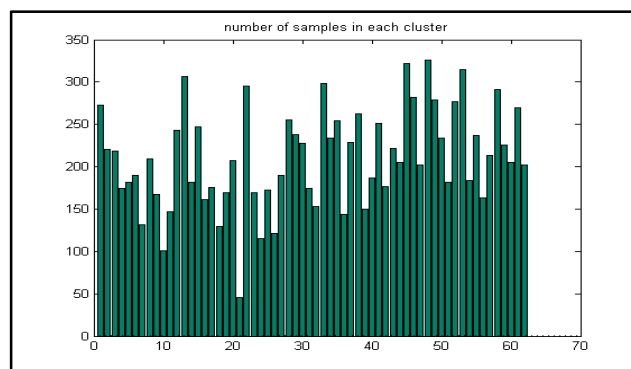
5-2-4 نتایج مرحله کاهش فرهنگ لغت

خوشه بندی تصاویر کلمات دست نوشته توسط روش ISOCLUS و الگوریتم DTW انجام می‌شود. بدین ترتیب کلمات با شکل کلی مشابه در خوشه‌های یکسان قرار گرفته و یک فرهنگ لغت تصویری تولید می‌کند. میانگین خوشه‌ها به عنوان مدخل اعضای آن خوشه در فرهنگ لغت تصویری بدست آمده، در نظر گرفته می‌شود.

به دلیل تنوع و تغییرات زیاد در بین دست‌نوشته‌ها، نمی‌توان برای هر کلمه (کلاس) نماینده‌ای پیدا کرد که بتواند تمام ویژگی‌های نمونه‌های آن کلاس را بیان کند. بنابراین در خوشه بندی کلمات دست نوشته برای افزایش دقت از تمام نمونه‌های آموزش هر کلمه استفاده شده است. در این حالت ممکن است یک شهر (یک کلمه در فرهنگ لغت)، متعلق به بیش از یک خوشه باشد.

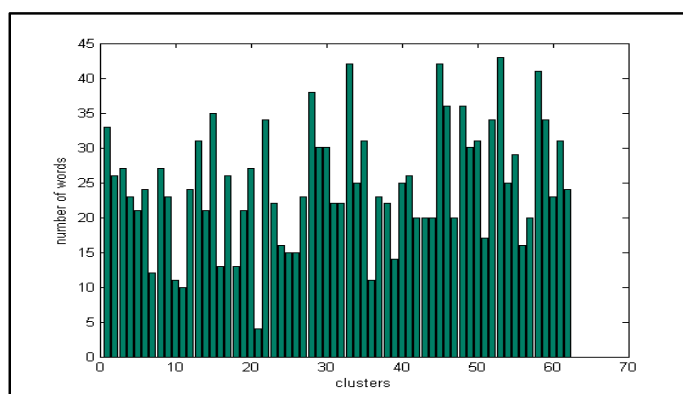
تعداد اولیه خوشه‌های بدست آمده توسط الگوریتم خوشه‌بندی سلسله مراتبی برابر با 45 و تعداد خوشه‌های نهایی برابر با 62 خوشه است. در پایان فصل تعدادی از نمونه‌های بعضی از خوشه‌ها آورده شده است.

نمودار تعداد نمونه‌های قرار گرفته در هر خوشه بر حسب خوشه‌ها در شکل 5-6 نشان داده شده است. به طور میانگین در هر خوشه 209 نمونه وجود دارد .



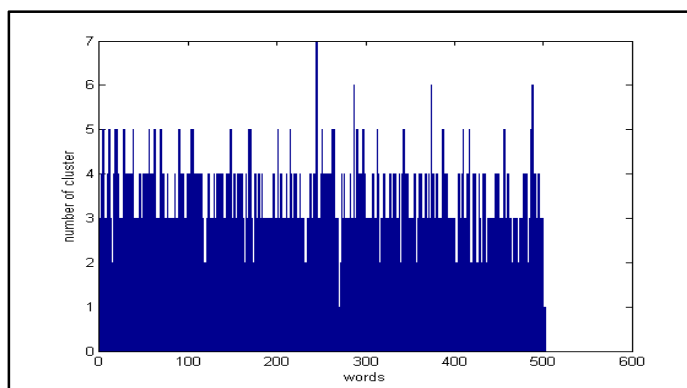
شکل 5-6. نمودار تعداد نمونه ها در هر خوشه

هر کلمه (شهر) که حداقل سه نمونه از آن در یک خوشه قرار بگیرد، متعلق به آن خوشه است. نمودار تعداد کلمه‌ها در هر خوشه در شکل 5-7 نشان داده است. به طور میانگین تعداد کلمه‌ها در هر خوشه، 21 کلمه است.



شکل 5-7. نمودار تعداد کلمه ها در هر خوشه

همانطور که گفته شده هر کلمه ممکن است در چند خوشه قرار بگیرد. شکل 5-8 نمودار تعداد خوشه‌هایی که هر کلمه در آن قرار گرفته است را نشان می دهد. (تعداد کلاس‌ها برابر با تعداد کلمه هاست. در پایگاه داده ایران شهر تعداد کلاس‌ها برابر با 503 می باشد). به طور میانگین هر کلمه در 3 خوشه قرار دارد.



شکل 5-8. نمودار تعداد خوشه‌هایی که هر کلمه در آن قرار گرفته است

برای ارزیابی عملکرد الگوریتم کاهش دهنده فرهنگ لغت از دو معیار دقت و درجه کاهش استفاده می‌شود که این دو معیار با هم ترکیب می‌شود و معیار درجه تأثیر را تولید می‌کند. بنابراین معیار درجه تأثیر برای مقایسه عملکرد الگوریتم‌های کاهش دهنده فرهنگ لغت مناسب است.

در ارزیابی عملکرد الگوریتم کاهش فرهنگ لغت پیشنهادی، فاصله کلمه آزمون ورودی با میانگین خوشه‌ها توسط الگوریتم DTW محاسبه می‌شود. تعدادی از نزدیکترین خوشه‌ها به کلمه آزمون ورودی انتخاب می‌شود (n). دقت و درجه کاهش الگوریتم وابسته به این تعداد است. اگر n زیاد باشد، دقت افزایش ولی درجه کاهش، کم می‌شود و اگر n کم باشد، دقت کاهش می‌یابد اما درجه کاهش فرهنگ لغت، زیاد می‌شود. برای تعیین مقدار n مناسب، استفاده از معیار درجه تأثیر مناسب‌تر است. جدول 5-1 میزان دقت، درجه کاهش و درجه تأثیر را برای مقادیر مختلفی از n برای 2500 تصویر مورد آزمون نشان می‌دهد.

با توجه به جدول، با انتخاب 5 خوشه نزدیک به تصویر کلمه مورد آزمون ورودی، احتمال وجود کلمه آزمون در این 5 خوشه 94/14% و میزان کاهش کلمات کاندید مورد بررسی برای بازشناسی تصویر ورودی 76/87% و درجه تأثیر الگوریتم 73/35% خواهد شد. یعنی در این مرحله به طور میانگین تعداد کلمات مورد بررسی برای بازشناسی تصاویر مورد آزمون ورودی از 503 به 105 کلمه کاهش می‌یابد.

جدول 5-1. دقت، درجه کاهش و درجه تأثیر برای مقادیر مختلفی از n

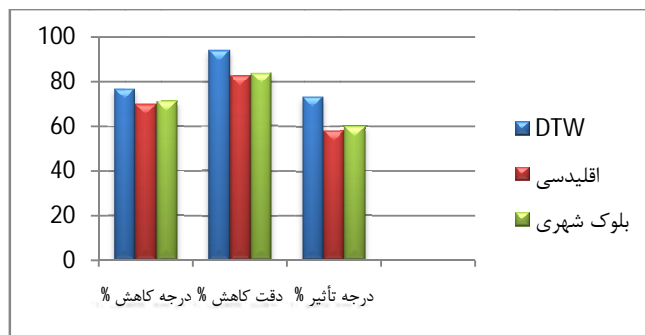
| درجه تأثیر % | دقت کاهش % | درجه کاهش % | خوشه های نزدیک (n) |
|--------------|--------------|--------------|--------------------|
| 59/86 | 75/28 | 91/7 | 1 |
| 68/43 | 84/27 | 81/21 | 2 |
| 73/35 | 94/14 | 76/87 | 5 |
| 65/64 | 94/34 | 69/58 | 7 |
| 57/51 | 95/01 | 60/54 | 10 |
| 50/01 | 95/9 | 52/15 | 12 |
| 46/22 | 97/07 | 47/62 | 15 |
| 39/08 | 98/37 | 39/73 | 20 |

جدول 5-2 عملکرد الگوریتم کاهش دهنده فرهنگ لغت پیشنهادی را برای ترکیب‌های مختلفی از ویژگی‌ها نشان می‌دهد.

جدول 5-2. عملکرد الگوریتم کاهش دهنده فرهنگ لغت پیشنهادی برای ترکیب‌های مختلفی از ویژگی‌ها

| درجه تأثیر % | دقت کاهش % | درجه کاهش % | پروفایل گذر | پروفایل پایینی | پروفایل بالایی | پروفایل پروژکشن عمودی |
|--------------|--------------|--------------|-------------|----------------|----------------|-----------------------|
| 40/89 | 55/36 | 73/88 | | | | * |
| 66/34 | 86/64 | 78/39 | | | * | * |
| 66/95 | 86/52 | 77/32 | * | | * | * |
| 68/53 | 86/64 | 79/1 | | * | * | * |
| 62/76 | 80/12 | 78/34 | * | * | * | |
| 73/35 | 94/12 | 79/88 | * | * | * | * |

از الگوریتم DTW برای اندازه‌گیری فاصله بین نمونه‌ها و فاصله بین کلمه آزمون ورودی و میانگین خوشه‌ها استفاده شده است. نمودار شکل 5-9 نتایج بدست آمده با الگوریتم DTW را با معیارهای فاصله دیگر مانند فاصله اقلیدسی و بلوک شهری مقایسه می‌کند.



شکل 5-9. مقایسه عملکرد الگوریتم DTW با معیار فاصله اقلیدسی و بلوک شهری

همانطور که نمودار نشان می‌دهد نتایج بدست آمده توسط الگوریتم DTW بهتر از معیار بلوک شهری و نتایج بدست آمده از معیار بلوک شهری بهتر از فاصله اقلیدسی می‌باشد.

3-5 مرحله دوم: بازشناسی کلمه آزمون ورودی

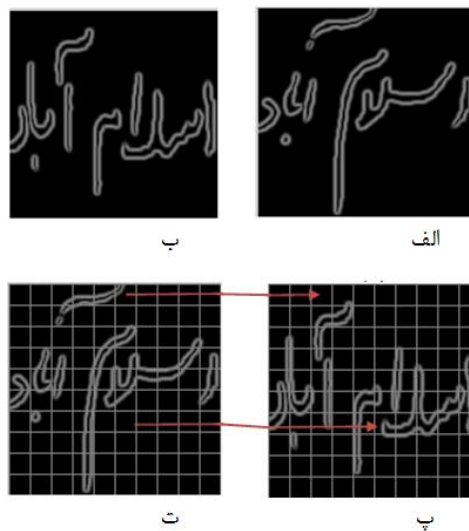
در مرحله اول روش پیشنهادی با استفاده از خوشه بندی تصاویر کلمات دست نوشته با ویژگی‌هایی کلی یک لیستی از کلمات کاندید برای بازشناسی کلمه آزمون ورودی تولید می‌شود. این مرحله به عنوان مرحله‌ای برای کاهش دامنه جستجو استفاده می‌شود. در مرحله دوم با استفاده از ویژگی‌های جزئی‌تر بهترین کاندید از لیست بدست آمده از مرحله اول برای شناسایی کلمه ورودی انتخاب می‌شود.

در مرحله دوم از طبقه بند k نزدیکترین همسایه¹ و ویژگی هیستوگرام گرادیان روشنایی محلی استفاده شده است. نحوه استخراج این ویژگی در فصل سوم ارائه شده است.

همانطور که در فصل سوم توضیح داده شد برای استخراج ویژگی هیستوگرام گرادیان محلی، تصویر گرادیان به بلوک‌هایی تقسیم می‌شود و از هر بلوک یک بردار 32 تایی استخراج می‌شود. در استخراج ویژگی هیستوگرام گرادیان محلی با اندازه بلوک‌های ثابت در تصاویر کلمات دست نوشته، بر اساس آنالیز خطاهای ایجاد شده در بازشناسی کلمات دست نوشته، به دو دلیل نتایج بازشناسی کاهش یافته است:

¹ K nearest neighbor(Knn)

1- امکان وجود جابه‌جایی و تغییر مکان در کلمات دست نوشته، باعث می‌شود که پیکسل‌های سیاه (پیکسل‌های قلم) موجود در بلوک‌های متناظر در تصاویر از یک جنس مشابه نباشد و این امر باعث کاهش دقت بازشناسی می‌شود. (شکل 5-10)



شکل 5-10. (الف) و (ب) تصاویر اندازه‌گردیدان نمونه های یک کلاس و (پ) و (ت) تصاویر بلوک بندی شده با سایز بلوک های ثابت

همانطور که در شکل دیده می‌شود کلمه "اسلام شهر" در تصویر شکل (الف) نسبت به تصویر شکل (ب) بالاتر نوشته شده است و این باعث می‌شود که بلوک‌های متناظر در دو تصویر باهم همخوانی نداشته باشند. نمونه‌هایی از این ناهمخوانی‌ها با فلش نشان داده شده است.

2- ثابت بودن اندازه بلوک‌ها در کل یک تصویر کلمه می‌تواند نشان‌دهنده این مطلب باشد که به کل تصویر با یک وزن نگاه شده در صورتی که بیشتر اطلاعات در تصویر کلمه دست نوشته را می‌توان در اطراف خط اصلی¹ کلمه دست نوشته پیدا کرد و بهتر است که در این ناحیه بلوک‌های بیشتر و با سایز کوچکتر قرار بگیرد تا اطلاعات در این ناحیه دقیق‌تر بیان شود.

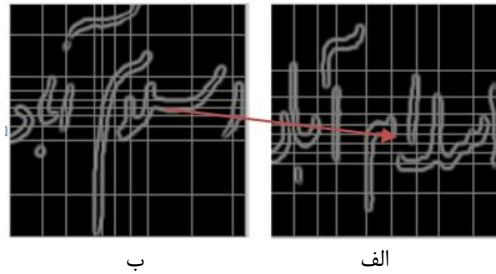
در این تحقیق برای کاهش این مشکلات، از بلوک‌های تطبیقی استفاده شده است. برای این منظور دو روش پیشنهاد شده است.

¹ baseline

5-3-1 روش پیشنهادی اول

اندازه بلوک‌ها بر اساس توزیع پیکسل‌های سیاه تغییر می‌کند. این کار باعث می‌شود که اندازه بلوک‌ها بر اساس پیکسل‌های قلم تنظیم شود و یک همخوانی نسبی بین بلوک‌های متناظر تصاویر نمونه‌های یک کلمه ایجاد شود و همچنین بلوک‌های کوچکتر و بیشتری در اطراف خط اصلی کلمه دست نوشته تولید شود. برای بلوک‌بندی تطبیقی تصویر به $m \times n$ بلوک، ابتدا تصویر به n نوار افقی تقسیم می‌شود به طوری که تعداد پیکسل‌های سیاه در همه نوارهای افقی با هم برابر باشند. سپس تصویر به m نوار عمودی تقسیم می‌شود به طوری که تعداد پیکسل‌های سیاه در همه نوارهای عمودی با هم برابر باشند. اندازه بلوک‌ها در هر تصویر کلمه با توجه به توزیع پیکسل‌های سیاه (پیکسل‌های قلم) آن تغییر می‌کند. ابعاد بردار ویژگی برابر با $m \times n \times 16$ است. (شکل 5-11)

همانطور که شکل 5-11 نشان می‌دهد، اندازه بلوک‌ها طبق مکان و نحوه نوشتن کلمه دست نوشته تغییر کرده است و بلوک‌های بیشتر و کوچکتری در اطراف خط اصلی کلمه دست نوشته ایجاد شده است.



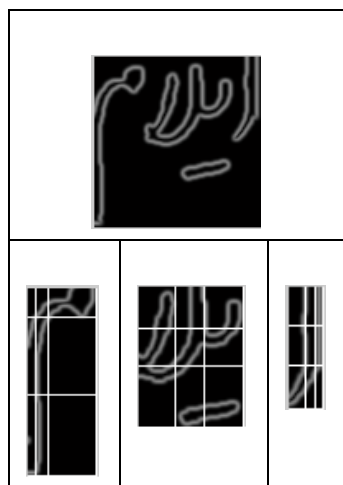
شکل 5-11. تصویر گرادین بلوک بندی شده با بلوک های تطبیقی برای نمونه هایی از یک کلمه

5-3-2 روش پیشنهادی دوم

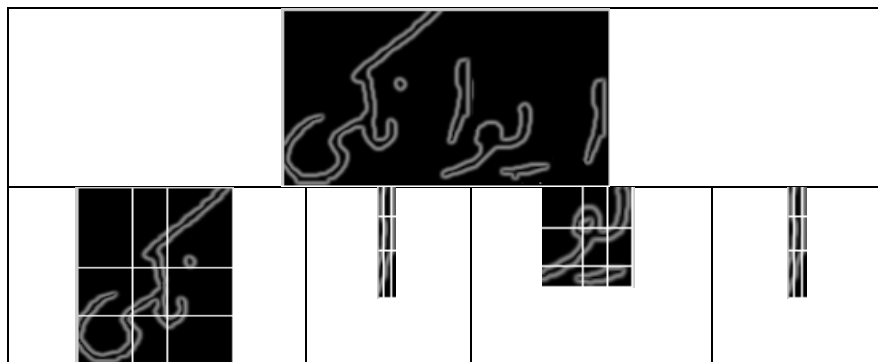
در این روش به جای اینکه کل تصویر بلوک بندی شود، اجزاء اصلی کلمه دست نوشته (زیر کلمات) آن به طور مجزا بلوک بندی می‌شود و سپس با هم ترکیب می‌شوند. برای استخراج اجزاء اصلی یک کلمه دست‌نوشته در مرحله اول خط اصلی کلمه بر اساس پروفایل پروژکشن افقی پیدا می‌شود سپس

مولفه‌های متصلی¹ که از خط اصلی یا خطوط نزدیک به خط اصلی عبور می‌کنند و تعداد پیکسل-هایشان از یک آستانه از پیش تعیین شده‌ای بیشتر است، به عنوان اجزاء اصلی کلمه انتخاب می‌شوند. بقیه مولفه‌های متصل به نزدیکترین جزء اصلی نسبت داده می‌شوند. سپس هر جزء اصلی به صورت تطبیقی (روش اول) بلوک بندی می‌شود و از هر بلوک هیستوگرام جهت‌گردان استخراج می‌شود (شکل 5-12).

در پایگاه داده نام شهرهای ایران (ایران شهر) استفاده شده در این تحقیق، ماکزیمم تعداد اجزاء اصلی در کلمات برابر با 7 است. بنابراین تعداد اجزاء اصلی برای تمام تصاویر 7 در نظر گرفته می‌شود. تعداد ابعاد ویژگی استخراج شده از هر تصویر $7 \times n \times m \times 16$ بدست می‌آید و در صورتی که تعداد زیر کلمات کمتر از هفت باشد، ویژگی‌های مربوط به جزء غایب صفر گذاشته می‌شود.



¹ Connected component



شکل 5-12. تصویر گرادیان اجزاء اصلی بلوک بندی شده کلمه دست نوشته

اهداف استفاده از اجزاء اصلی به جای کل تصویر برای استخراج هیستوگرام گرادیان را می توان به صورت زیر بیان کرد:

- 1- تأثیر جابه جایی و تغییر مکانها در نمونه های یک کلمه دست نوشته حذف می شود.
- 2- فضاهای خالی موجود در تصاویر که هیچ اطلاعاتی ندارد در استخراج هیستوگرام گرادیان در نظر گرفته نمی شوند.

3- هر جزء اصلی در نمونه های کلمه دست نوشته با هم مقایسه می شود.

4- تعداد اجزاء اصلی نیز می تواند به عنوان یک ویژگی در نظر گرفته شود.

برای ارزیابی و مقایسه عملکرد ویژگی محلی گرادیان و روش های پیشنهاد شده گرادیان با بلوک های تطبیقی (روش اول و دوم) از ترکیب دو معیار جداپذیری، پراکندگی درون کلاسی¹ و پراکندگی بین کلاس ها² استفاده شده است.

پراکندگی درون کلاس میزان پراکندگی بردارهای ویژگی در نمونه های یک کلاس و پراکندگی بین کلاس میزان پراکندگی ویژگی های نمونه های کلاس ها نسبت به یکدیگر را نشان می دهد. ویژگی هایایی برای کلاس بندی مناسب است که میزان پراکندگی بین نمونه های یک کلاس کم و میزان پراکندگی بین کلاس ها زیاد باشد. به طور کلی نسبت پراکندگی درون کلاسی به پراکندگی بین

¹ Between-class scatter

² Within-class scatter

کلاس ها می تواند یک معیار برای ارزیابی مناسب بودن یک ویژگی باشد، هر چه این نسبت بزرگتر باشد، ویژگی استخراج شده برای جدا کردن کلاس ها مناسب تر است .

پراکندگی درون کلاسی S_w و پراکندگی بین کلاس ها S_b به صورت زیر بدست می آید:

$$S_w = \sum_{i=1}^M P_i \Sigma_i \quad (4.5)$$

M تعداد کلاس ها و Σ_i ماتریس کواریانس ، P_i احتمال پیشین کلاس i ام است.

$$S_b = \sum_{i=1}^M P_i (\mu_i - \mu_0)(\mu_i - \mu_0)^T \quad (5.5)$$

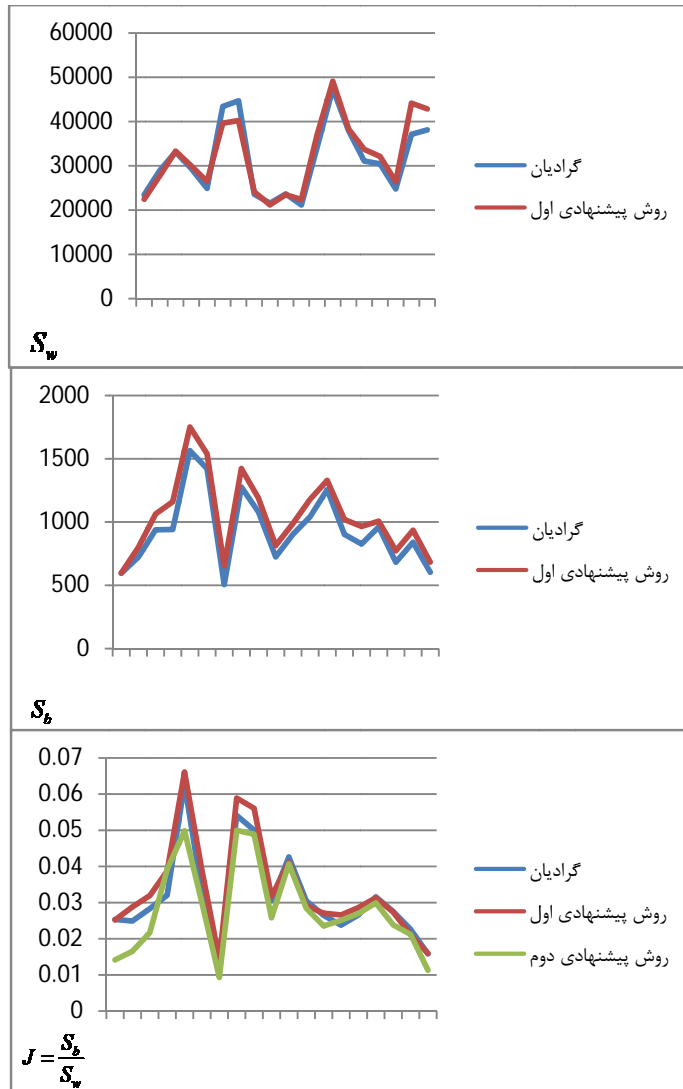
$$\mu_0 = \sum_{i=1}^M P_i \mu_i$$

μ_i میانگین کلاس i ام و μ_0 میانگین تمام کلاس هاست .

$$J = \frac{S_b}{S_w} \quad (6.5)$$

برای نشان دادن میزان پراکندگی درون کلاس و بین کلاس ویژگی های محلی گرادیان و گرادیان های اصلاح شد ، از 400 تصویر کلمه دست نوشته که شامل بیست کلمه و از هر کلمه بیست نمونه، از پایگاه داده ایران شهر استفاده شده است. تعداد بلوک ها در ویژگی گرادیان و گرادیان اصلاح شده به روش یک، 5×5 در نظر گرفته شده است. بنابراین ابعاد بردار ویژگی در هر دو روش برابر با 400 است. تعداد بلوک ها در ویژگی گرایان به روش دوم 3×3 است. بنابراین ابعاد ویژگی برابر با 1008 می باشد.

نمودارهای شکل 5-13 میزان پراکندگی درون کلاس S_w و بین کلاس ها S_b و نسبت بین آنها J ، را برای ویژگی های مورد نظر نشان می دهد. همانطور که نمودارها نشان می دهد میزان پراکندگی درون کلاس ها برای ویژگی گرادیان و گرادیان اصلاح شده به روش یک که دارای ابعاد ویژگی مساوی هستند، تقریباً یکسان است ولی میزان پراکندگی بین کلاس ها برای ویژگی گرادیان اصلاح شده به روش یک به مراتب بیشتر از ویژگی گرادیان است و این نشان دهنده مناسب بودن ویژگی گرادیان اصلاح شده به روش یک نسبت به گرادیان برای جدا کردن کلاس ها است.



شکل 5-13. الف- میزان پراکندگی درون کلاس S_w و ب- بین کلاس ها S_b و پ- نسبت S_b به S_w از آنجایی که ابعاد ویژگی گراדיان تطبیقی به روش دوم بیشتر از دو ویژگی دیگر است، مقادیر S_w, S_b آن از دو ویژگی دیگر بیشتر است. اما با توجه به نمودار معیار J ، مقادیر J ویژگی گراדיان اصلاح شده به روش دوم نسبت به دو ویژگی دیگر برای بیشتر کلمه ها پایین تر است. دلیل مناسب نبودن ویژگی گراדיان اصلاح شده به روش دوم را می توان خطاهای ایجاد شده در جداسازی اجزاء اصلی کلمه دانست. دلایل خطاهای ایجاد شده در جداسازی اجزاء اصلی را می توان به طور کلی به چهار دسته تقسیم کرد:

1- چسبیدن اجزاء اصلی به یکدیگر

2- بعضی از اجزاء اصلی به دلیل فاصله از خط اصلی، به عنوان مولفه اصلی در نظر گرفته نمی شوند.

3- جابه جایی مکمل هایی مثل سرکش "ک و گ" و کلاه "آ" از جزء اصلی اش

4- بزرگ نوشتن بعضی از نقاط و علائم که باعث می شود به عنوان یک جزء اصلی در نظر گرفته شوند.

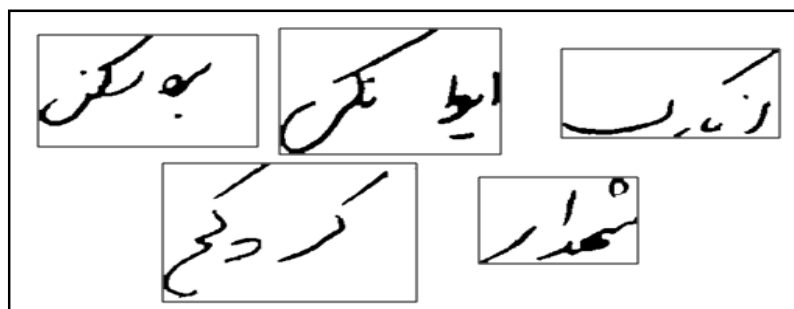
در شکل 5-14 برخی از کلماتی که در جداسازی اجزاء اصلی خطا داشته اند، را نشان می دهد.

در کلمه دست نوشته "انارک" جابه جایی سرکش "ک" باعث شده است که سرکش در جزء اصلی "نا" قرار بگیرد و همچنین کوچک نوشتن حرف "ر" باعث شده که به عنوان جزء اصلی در نظر گرفته نشود.

در کلمات دست نوشته "ایوانکی" و "بردسکن" چسبیدن اجزاء اصلی به یکدیگر باعث ایجاد خطا در جداسازی اجزاء اصلی شده است.

بالتر نوشتن جزء اصلی "ا" در کلمه دست نوشته "شهدار" باعث شده که به عنوان یک جزء اصلی در نظر گرفته نشود.

بزرگ نوشتن و همچنین نزدیک بودن به خط اصلی نقاط حرف "چ" در کلمه "کرکچ" باعث شده که این نقاط به عنوان یک جزء اصلی در نظر گرفته شود.



شکل 5-14. برخی از کلماتی که در جدا سازی مولفه های اصلی خطا داشته اند

3-3-5 نتایج مرحله بازشناسی

در ادامه ابتدا نتایج بازشناسی کلمات دست‌نوشته به کمک ویژگی هیستوگرام گرادیان محلی بدون مرحله کاهش فرهنگ‌لغت و سپس تأثیر اضافه کردن مرحله کاهش فرهنگ لغت را بررسی می‌کنیم. برای طبقه‌بند KNN، پس از ارزیابی مقادیر مختلف k و معیارهای مختلف فاصله (فاصله اقلیدسی، بلوک شهری و زاویه ای) طی آزمایشات متعدد، معیار فاصله بلوک شهری و مقدار $k=5$ انتخاب شد. جدول 3-5 نتایج بازشناسی کلمات دست‌نوشته (بدون مرحله کاهش فرهنگ لغت) برای تعداد بلوک های مختلف ویژگی هیستوگرام گرادیان محلی را نشان می‌دهد.

در ویژگی هیستوگرام اصلاح شده به روش دوم، اجزاء اصلی از هم جدا و سپس هر جزء به طور مجزا بلوک بندی می‌شوند. محدوده مرزی¹ بعضی از اجزا آنقدر کوچک است که نمی‌توان آنها را به تعداد بلوک های بیش از 3×3 تقسیم کرد. بنابراین فقط تعداد بلوک های 3×3 در این روش بررسی می‌شوند.

با توجه به جدول، ویژگی هیستوگرام اصلاح شده به روش اول نسبت به دو روش دیگر دقت بازشناسی بالاتری دارد. البته هر چه تعداد بلوک ها افزایش می‌یابد، دقت افزایش ولی به همان نسبت ابعاد ویژگی‌ها نیز زیاد می‌شود. با در نظر گرفتن ابعاد ویژگی و دقت بازشناسی، ویژگی هیستوگرام گرادیان محلی اصلاح شده به روش اول با تعداد بلوک های 5×5 مناسب‌ترین ویژگی برای بازشناسی کلمه دست‌نوشته نسبت به دو ویژگی دیگر است.

جدول 3-5. دقت بازشناسی برای ویژگی های مختلف

| تعداد بلوک ها ویژگی | 3×3 | 5×5 | 7×7 | 10×10 |
|--|--------------|--------------|--------------|----------------|
| هیستوگرام گرادیان محلی | 49/78 | 65/32 | 67/14 | 69/44 |
| هیستوگرام گرادیان محلی اصلاح شده به روش اول | 57/32 | 76/62 | 78/12 | 79/70 |
| هیستوگرام گرادیان محلی اصلاح شده به روش دوم | 69/10 | - | - | - |

¹ Bounding box

دلیل مناسب بودن ویژگی هیستوگرام گرادیان محلی اصلاح شده به روش اول نسبت به دو روش دیگر را می توان تطبیق نسبی بلوک ها با توزیع پیکسل های سیاه (بیکسل های قلم) دانست. خطاهای ایجاد شده در جداسازی اجزاء اصلی در کلمات دست نوشته و ابعاد بالا در ویژگی هیستوگرام گرادیان محلی اصلاح شده به روش دوم باعث ضعیف عمل کردن این ویژگی نسبت به ویژگی هیستوگرام گرادیان محلی اصلاح شده به روش اول شده است.

جدول 4-5 دقت و سرعت بازشناسی را برای 2500 تصویر مجموعه آزمون در هنگام استفاده از مرحله کاهش فرهنگ لغت نشان می دهد (ویژگی هیستوگرام گرادیان اصلاح شده به روش اول با تعداد بلوک های 5×5 در مرحله بازشناسی کلمه ورودی استفاده شده است) همانطور که جدول نشان می دهد، استفاده از مرحله کاهش فرهنگ لغت باعث افزایش دقت بازشناسی شد اما در سرعت بازشناسی تأثیر زیادی نداشت و دلیل آن را می توان، به کار بردن الگوریتم وقت گیر DTW دانست.

جدول 4-5 دقت و سرعت بازشناسی برای روش های مختلف

| روش | دقت بازشناسی % | سرعت بازشناسی (ثانیه بر نمونه) |
|------------------------------------|----------------|--------------------------------|
| بازشناسی بدون مرحله کاهش فرهنگ لغت | 76/62 | 0/66 |
| بازشناسی با مرحله کاهش فرهنگ لغت | 81/19 | 0/61 |

در الگوریتم دو مرحله ای پیشنهاد شده برای بازشناسی کلمات دست نوشته، در مرحله اول تعداد کلمات مورد بررسی کاهش می یابد. بنابراین ممکن است کلمات مورد بررسی برای هر کلمه آزمون ورودی متفاوت باشد در این صورت باید در مرحله دوم از طبقه بندی استفاده شود که مرحله آموزش سریعی داشته باشد. به همین دلیل از طبقه بند k نزدیکترین همسایه به دلیل سریع بودن این طبقه بند هم در مرحله آموزش و هم در مرحله آزمون، استفاده شد.

برای مقایسه عملکرد الگوریتم دو مرحله ای پیشنهاد شده برای بازشناسی کلمات دست نوشته، از طبقه بند SVM چند کلاسه با تابع کرنل RBF و ویژگی هیستوگرام گرادیان محلی اصلاح شده به روش اول استفاده شده است. مقدار پارامترهای لازم برای SVM چند کلاسه از روش Cross Validation بدست آمده است. نتایج بدست آمده در جدول 5-5 آورده شده است.

همانطور که جدول نشان می‌دهد دقت بازشناسی طبقه بند SVM چند کلاسه بهتر از الگوریتم دو مرحله‌ای پیشنهادی است. اما زمان صرف شده برای آموزش و آزمون الگوریتم پیشنهادی بسیار کمتر از طبقه بند SVM چند کلاسه است.

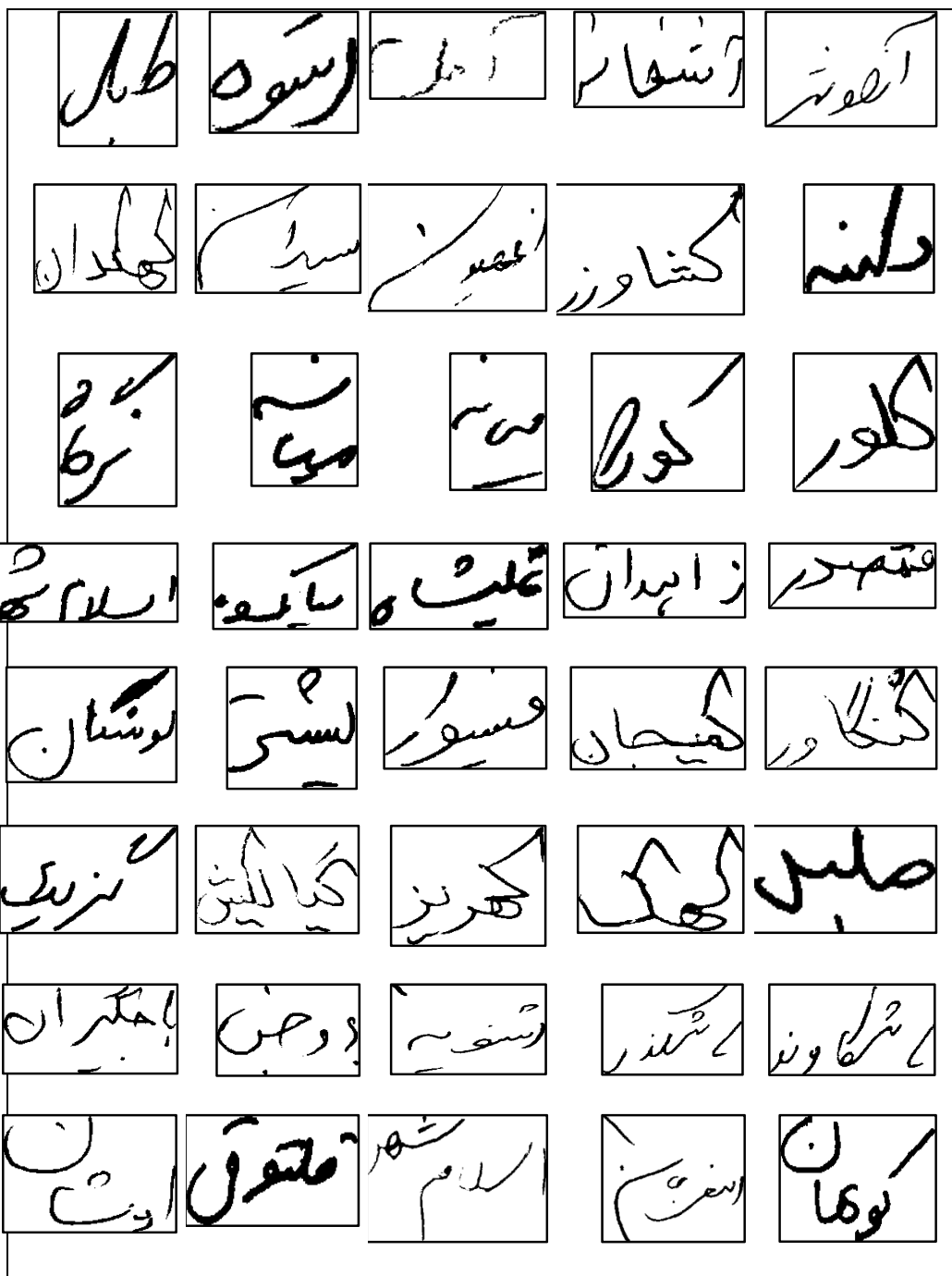
جدول 5-5. مقایسه دقت و زمان بازشناسی الگوریتم دو مرحله‌ای پیشنهادی با SVM چند کلاسه

| روش | دقت بازشناسی | زمان آموزش (ثانیه بر نمونه) | زمان آزمون (ثانیه بر نمونه) |
|-------------------------------|--------------|-----------------------------|-----------------------------|
| الگوریتم دو مرحله‌ای پیشنهادی | 81/19 | 1/6 | 0/61 |
| چند کلاسه SVM | 83/34 | 6/65 | 1/35 |

در ادامه تعدادی از تصاویر کلمات دست‌نوشته‌ای که در الگوریتم دو مرحله‌ای پیشنهادی به اشتباه بازشناسی شده اند آمده است.

با توجه به تصاویر کلمات دست‌نوشته‌ای که باعث ایجاد خطا در بازشناسی شده اند از جمله دلایل خطاهای بازشناسی روش پیشنهادی را می‌توان به صورت زیر بیان کرد:

- 1- چسبندگی زیر کلمات به یکدیگر (مثل تصاویر کلمات دست‌نوشته "آزادشهر"، "دامنه"، "کورا")
- 2- خط خوردگی و غلط‌های املائی (مثل تصاویر کلمات دست‌نوشته "علیشاه"، "گالیکش"، "اسلام شهر" یا "کنگاور")
- 3- جابه‌جایی‌ها یا بالاتر یا پایین‌تر نوشتن بعضی از زیرکلمات نسبت به بقیه زیر کلمات دیگر (مثل تصاویر کلمه دست‌نوشته "میانه" و یا "نرگه")
- 4- تغییر در نحوه نوشتن بعضی حروف (مثل تصاویر کلمه دست‌نوشته "کهکدان" و یا "کهک" یا "سیدان")
- 5-



شکل 5-15. تعدادی از تصاویر کلمات دست نوشته که باعث ایجاد خطا شده اند

| | | | | | | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| ورنگش | ورنگش | ورنگش | ورنگش | ورنگش | ورنگش | ورنگش | ورنگش | ورنگش | ورنگش | ورنگش | ورنگش | ورنگش |
| ورنگ | ورنگ | ورنگ | ورنگ | ورنگ | ورنگ | ورنگ | ورنگ | ورنگ | ورنگ | ورنگ | ورنگ | ورنگ |
| ارنگ | ارنگ | ارنگ | ارنگ | ارنگ | ارنگ | ارنگ | ارنگ | ارنگ | ارنگ | ارنگ | ارنگ | ارنگ |
| اشنگ | اشنگ | اشنگ | اشنگ | اشنگ | اشنگ | اشنگ | اشنگ | اشنگ | اشنگ | اشنگ | اشنگ | اشنگ |
| اک | اک | اک | اک | اک | اک | اک | اک | اک | اک | اک | اک | اک |
| انارک | انارک | انارک | انارک | انارک | انارک | انارک | انارک | انارک | انارک | انارک | انارک | انارک |
| باغچران | باغچران | باغچران | باغچران | باغچران | باغچران | باغچران | باغچران | باغچران | باغچران | باغچران | باغچران | باغچران |
| بنجار | بنجار | بنجار | بنجار | بنجار | بنجار | بنجار | بنجار | بنجار | بنجار | بنجار | بنجار | بنجار |
| دوستان | دوستان | دوستان | دوستان | دوستان | دوستان | دوستان | دوستان | دوستان | دوستان | دوستان | دوستان | دوستان |

شکل 5-17. تعدادی از تصاویر کلمات دست نوشته خوشه 10

| | | | | | | | | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| بسطام | بسطام | بسطا | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام |
| بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام |
| بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام |
| بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام |
| بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام |
| بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام |
| بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام |
| بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام | بسطام |

شکل 5-18. تعدادی از تصاویر کلمات دست نوشته خوشه 15

| | | | | | | | | | | | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک |
| اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار |
| آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک | آبِک |
| اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار |
| اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار |
| اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار |
| اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار |
| اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار |
| اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار | اکار |

شکل 5-19. تعدادی از تصاویر کلمات دست نوشته خوشه 43

نتیجه‌گیری و پیشنهادات

فصل هفتم: نتیجه گیری و پیشنهادات

در این پایان نامه سیستمی برای بازشناسی کلمات دست‌نوشته فارسی مبتنی بر یک روش دو مرحله-ای ارائه شد. این سیستم بر روی پایگاه داده "ایران شهر" که شامل 16000 تصویر کلمه دست‌نوشته از 503 نام شهرهای ایران است، مورد ارزیابی قرار گرفت.

در این سیستم برای افزایش کیفیت تصاویر و کاهش اختلافات بین تصاویر کلمات دست‌نوشته در یک کلاس یک سری عملیات پیش‌پردازش انجام می‌شود. این عملیات شامل باینری کردن، هموار سازی، حذف نویز، آشکار سازی خط اصلی، اصلاح شیب خط اصلی، اصلاح کجی دست‌نوشته، آنالیز مولفه‌های متصل، نرمال سازی ابعاد تصاویر و نازک سازی می‌باشد. برای هر کدام از این عملیات، تکنیک‌های مختلفی پیاده سازی و مقایسه شد. در هر مورد مناسب‌ترین تکنیک بر روی تصاویر پایگاه داده "ایران شهر" انتخاب شد.

در مرحله اول الگوریتم پیشنهادی برای افزایش دقت و سرعت، تعداد کلمات مورد بررسی برای بازشناسی کلمه دست‌نوشته ناشناخته ورودی کاهش داده شد که به این مرحله کاهش فرهنگ لغت گفته می‌شود. برای این کار از خوشه بندی تصاویر کلمات دست‌نوشته برای ایجاد یک فرهنگ لغت تصویری و کاهش دامنه جستجو برای بازشناسی کلمه آزمون ورودی استفاده شد. برای خوشه بندی تصاویر کلمات دست‌نوشته، الگوریتم خوشه بندی ISOCLUS و برای تعیین تعداد و مقدار مراکز اولیه خوشه‌ها، الگوریتم خوشه‌بندی سلسله مراتبی اعمال شد. در این مرحله از هر تصویر چهار نوع ویژگی استخراج شد: پروفایل بالا، پایین و پروژکشن عمودی که شکل بیرونی کلمه دست‌نوشته و پروفایل گذر سیاه به سفید که ساختار داخلی کلمه را توصیف می‌کند و برای کاهش ابعاد ویژگی در این مرحله، از هر بردار ویژگی یک بار تبدیل موجک یک بعدی گرفته می‌شود.

برای اندازه‌گیری فاصله بین بردارهای ویژگی استخراج شده از تصاویر از الگوریتم DTW به خاطر کاهش اختلاف کشیدگی بین حروف استفاده شد و نتایج نشان داد که استفاده از الگوریتم DTW باعث بهبود عملکرد مرحله کاهش فرهنگ لغت شد.

تصاویر کلمات دست نوشته در 62 خوشه قرار گرفتند. میانگین هر خوشه به عنوان مدخل اعضای آن خوشه در فرهنگ لغت تصویری در نظر گرفته شد. 5 خوشه نزدیک به هر کلمه دست نوشته ناشناخته ورودی انتخاب و کلمات موجود در این خوشه‌ها در لیست کلمات کاندید قرار می‌گیرند. کلمه نامعلوم ورودی با دقت 94% در 5 خوشه نزدیک قرار دارد و تعداد کلمات مورد بررسی موجود در فرهنگ لغت برای بازشناسی کلمه نامعلوم ورودی، 77% کاهش می‌یابد.

مرحله دوم سیستم پیشنهادی، بازشناسی کلمه نامعلوم ورودی از کلمات کاندید موجود در لیست بدست آمده از مرحله اول است. در این مرحله از ویژگی هیستوگرام گرادیان محلی استفاده شد که برای استخراج آن تصویر گرادیان بلوک‌بندی می‌شود سپس از هر بلوک یک بردار 32 تایی استخراج می‌شود. برای کاهش ابعاد از فیلتر گوسی و عملیات کاهش نرخ نمونه‌برداری استفاده شده است. در این تحقیق برای تطبیق بلوک‌ها به نحوه نوشتن کلمه دست‌نوشته دو روش پیشنهاد شد. در روش اول اندازه بلوک‌ها در تصویر گرادیان بر اساس توزیع پیکسل‌های قلم تغییر می‌کنند. با این کار اندازه بلوک‌ها در مکان‌هایی که تعداد پیکسل‌های قلم بیشتر باشند کوچکتر و تعداد آنها زیاده‌تر خواهد شد و در مکان‌هایی که تعداد پیکسل‌های قلم کمتر و اطلاعاتی زیادی وجود نداشته باشد، اندازه بلوک‌ها بزرگتر می‌شود.

در روش دوم برای تطبیق بهتر اجزاء متناظر نمونه‌های دست نوشته کلمات، ابتدا اجزاء اصلی را از هم جدا کرده سپس هر جزء به طور مجزا بر اساس توزیع پیکسل‌های سیاه‌اش بلوک بندی می‌شود و سپس بردارهای ویژگی استخراج شده از اجزاء با هم ترکیب می‌شوند.

نتایج بازشناسی 2500 کلمه دست نوشته در مجموعه آزمون توسط طبقه بند Knn نشان می‌دهد که نرخ بازشناسی با استفاده از ویژگی هیستوگرام گرادیان محلی با تعداد بلوک‌های هم اندازه 5×5 ,

65% و نرخ بازشناسی با استفاده از ویژگی هیستوگرام گرادیان محلی با تعداد بلوک های تطبیقی بر اساس پیکسل های سیاه 5×5 (روش اول پیشنهادی)، 76% می باشد. نرخ بازشناسی برای ویژگی هیستوگرام گرادیان با جداسازی اجزاء اصلی و بلوک بندی هر یک از اجزا به تعداد 3×3 برابر با 69% شد. دلایل مناسب نبودن ویژگی گرادیان پیشنهادی به روش دوم نسبت به روش اول را می توان خطاهای ناشی از جدا سازی و ابعاد بالای آن بیان کرد.

با استفاده از مرحله کاهش فرهنگ لغت به همراه ویژگی هیستوگرام گرادیان تطبیقی به روش اول در مرحله بازشناسی نرخ دقت به دلیل کاهش کلمات نامتشابه به 81% افزایش یافت ولی سرعت بازشناسی نسبت به زمانیکه که از مرحله کاهش فرهنگ لغت استفاده نشد، افزایش چندانی نیافت که مهمترین دلیل آن را می توان استفاده از الگوریتم زمانبر DTW دانست.

در پایان الگوریتم دو مرحله ای پیشنهادی با طبقه بند چند کلاسه SVM همراه با تابع کرنل RBF مقایسه شد، از نظر دقت بازشناسی تقریباً مشابه هستند اما از نظر زمان آموزش و زمان بازشناسی، الگوریتم پیشنهادی سریعتر است.

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

1- اضافه کردن ویژگی های ساختاری مثل تعداد نقاط، زیر کلمات و... در مرحله کاهش تعداد کلمات مورد بررسی

2- استفاده از طبقه بندهایی با کارایی بهتر از طبقه بند Knn در مرحله بازشناسی

3- اضافه کردن معیارهای دیگر برای تطبیق بهتر بلوک ها مثل واریانس پیکسل های قلم در ویژگی هیستوگرام گرادیان محلی اصلاح شده به روش اول

4- ارائه روشی دقیق تر برای جداسازی اجزاء اصلی کلمات دست نوشته به منظور کاهش خطاهای ناشی از جداسازی در ویژگی هیستوگرام گرادیان محلی اصلاح شده به روش دوم

مراجع

- [1] A. L. Koerich, R. Sabourin, C. Y. Suen, "Large vocabulary off-line handwriting recognition: A survey", *Pattern Anal Applic* (2003) 6: 97–121
- [2] Rui Xu, Donald Wunsch II, "Survey of Clustering Algorithms" *IEEE Transactions on Neural Networks*, VOL. 16, NO. 3, MAY 2005
- [3] Zaher Al Aghbari, Salama Brook, "HAH manuscripts: A holistic paradigm for classifying and retrieving historical Arabic handwritten documents", *Expert Systems with Applications* 36 (2009) 10942–10951
- [4] M. M. Haji, H. J. Eghbali, S. D. Katebi, "Farsi Handwritten Word Recognition Using Continuous Hidden Markov Models and Structural Features" MS These, JANUARY 2005
- [5] Malik Waqas Sagheer, Chun Lei He, Nicola Nobile, Ching Y. Suen, "Holistic Urdu Handwritten Word Recognition Using Support Vector Machine", 1051-4651/10 \$26.00 © 2010 IEEE
- [6] A. Broumandnia, J. Shanbehzadeh, M. RezaKhah Varnoosfaderani, "Persian/arabic handwritten word recognition using M-band packet wavelet transform", *Image and Vision Computing* 26 (2008) 829–842
- [7] Ramy Al-Hajj Mohamad, Laurence Likforman-Sulem, and Chafic Mokbel, "Combining Slanted-Frame Classifiers for Improved HMM-Based Arabic Handwriting Recognition" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, VOL. 31, NO. 7, JULY 2009
- [8] Jawad H. AlKhateeb, Jianmin Jiang, Jinchang Ren, Fouad Khelifi and Stan S. Ipson, "Multiclass Classification of Unconstrained Handwritten Arabic Words Using Machine Learning Approaches" *The Open Signal Processing Journal*, 2009, 2, 21-28
- [9] Saeed Mozaffari, Karim Faez, Volker Ma'rgner, Haikal El-Abed, "Lexicon reduction using dots for off-line Farsi/Arabic handwritten word recognition" *Pattern Recognition Letters* 29 (2008) 724–734

- [10] Zahra bahmani, Fatemh Alamdar, Reza Azmi, Saman Haratizadeh ,” Off-line Arabic/Farsi Handwritten Word Recognition Using RBF Neural Network and Genetic algorithm “ , 978-1-4244-6585-9/10/26.00 ©2010 IEEE
- [11] Powalka R. K., Sherkat N. and Whitrow R. J., “The Use of Word Shape Information for Cursive Script Recognition”, Fourth International Workshop on Frontiers of Handwriting Recognition, pp .67-76, Taiwan. 1994.
- [12] Hull J. J., “Word shape analysis in a knowledge-based system for reading text” , The Second IEEE Conf. on Artificial Intelligence Applications, pp. 114-119, 1985.
- [13] Ho T. K. and Hull J. J. and Srihari S. N., “A Computational Model for Recognition of Multifont Word Images”, Machine Vision and Applications, Vol. 5, 3, pp. 157-168, 1992.
- [14] Bunke H, Roth M, Schukat-Talamazzini EG. “ Off-line cursive handwriting recognition using hidden markov models” . Pattern Recognition1995; 28(9):1399–1413
- [15] Dong Xiang, Huahua Yan, Xianqiao Chen, Yanfen Cheng , ” Offline Arabic Handwriting Recognition System based on HMM” , 978-1-4244-5540-9/10/26.00 ©2010 IEEE
- [16] El-Yacoubi A, Gilloux M, Sabourin R, Suen CY, ”Unconstrained handwritten word recognition using hidden markov models”, IEEE Trans Pattern Analysis and Machine Intelligence 1999; 21(8):752–760
- [17] Toni M. Rath and R. Manmatha, ” Word Image Matching Using Dynamic Time Warping “ Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’03)
- [18] J. Weston and C. Watkins, “ Multi-class support vector machines” , Technical report CSD-TR-98-04 (1998).
- [19] Kavallieratou, E., Fakotakis, N. and Kokkinakis, G. “A Slant Removal Algorithm”, Pattern Recognition (2000)., pp. 1261-1262.
- [20] Dehghan M., Faez K., Ahmadi M. and Shridhar M., “Handwritten Farsi (Arabic) word recognition: a holistic approach using discrete HMM” , Pattern Recognition, vol. 34, no. 5, pp. 1057-1065, 2001.
- [21] Dehghan M., Faez K., Ahmadi M. and Shridhar M., “Unconstrained Farsi handwritten word recognition using fuzzy vector quantization and hidden Markov models”, Pattern Recognition Letters, vol. 22, no. 2, pp. 209-214, 2001.

[22] Mohamed El-Adawi, Hesham A. Keshk, Manal A. Ismail, and Essam M. Zaki, "Fast skew and slant correction for Arabic writtenword or line " ,Helwan University, Faculty of Engineering.

[23] Zhang, D., and Lu, G., "Review of shape representation and description techniques", Pattern Recognition, vol. 37, pp. 1-19, 2004.

[24] Hiroaki Sakoa , Seibi Chiba "Dynamic Programming Algorithm Optimization for Spoken Word Recognition "IEEE Transactions on Acoustics, Speech, and SIGNAL Processing, VOL. ASSP-26, NO. 1, FEBRUARY 1978

[25] PCI Geomatics Corp , "ISOCLUS–Isodata clustering program" <http://www.pcigeojmatics.com/cgibin/pcihlp/ISOCLUS>

[26] Nargess Memarsadeghi, David M. Mount, Nathan S. Netanyahu, Jacqueline Le Moigne, " A Fast Implementation of the ISOCLUS Algorithm",0 – 7803 – 7929 - 2/03/\$17.00 © 2003IEEE

[27] Hossein Khosravi, Ehsanollah Kabir , " Introducing a very large dataset of Handwritten Farsi digits and a study on their varieties" , Pattern Recognition Letters 28 (2007) 1133–1141

[28] Afshin Ebrahimi, Ehsanollah Kabir , " A pictorial dictionary for printed Farsi subwords " Pattern Recognition Letters 29 (2008) 656–663

[29] Kianoosh Bagheri Noaparast and Ali Broumania," Persian Handwritten Word Recognition Using Zernike and Fourier–Mellin Moments" 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications March 22-26, 2009 – TUNISIA

[30] Mohamed Cheriet,Nawwaf Kharma ,Cheng –lin liu ,ching Y.suen ," character recognition system, A Guide for Students and Practioners" ,wiley 2007

[31] Guillevic D, Suen CY. "HMM-KNN word recognition engine for bank cheque processing". Proc International Conference on Pattern Recognition, Brisbane, Australia, 1998; 1526–1529

[32] Liana M. Lorigo and Venu Govindaraju," Off-line Arabic Handwriting Recognition: A Servey" , To Appear In IEEE Transactions on Pattern Analysis And Machine Intelligence (2008).

[33] Amin, A. and Mansoon, W., "Recognition of Arabic printed text using neural networks" , Proc. 4th Int. Conf. on Document Analysis Recognition, Ulm, Germany, August 1997.

[34] Sameh M. Awaidah, Sabri A. Mahmoud ,” A multiple feature/resolution scheme to Arabic (Indian) numerals recognition using hidden Markov models” , Signal Processing 89 (2009) 1176–1184

[35] Kim JH, Kim KK, Suen CY. “Hybrid schemes of homogeneous and heterogeneous classifiers for cursive word recognition” , Proc 7th International Workshop on Frontiers in Handwriting Recognition, Amsterdam. Netherlands, 2000; 433–442

[36] Malik Waqas Sagheer, Chun Lei He, Nicola Nobile, Ching Y. Suen,” Holistic Urdu Handwritten Word Recognition Using Support Vector Machine” 1051-4651/10 /26.00 © 2010 IEEE . DOI 10.1109/ICPR.2010.468

[37] Alireza Alaei, P. Nagabhushan. ” A Baseline Dependent Approach for Persian Handwritten Character Segmentation “ ,1051-4651/10 \$26.00 © 2010 IEEE

[38] Otsu, N., “ A Threshold Selection Method from Gray Level Histograms ”, IEEE Trans. on Systems, Man and Cybernetics (Jan. 1979), vol. 9, pp. 62-66.

[39] Hennig A, Sherkat N. “Cursive script recognition using wildcards and multiple experts”. Pattern Analysis and Applications 2001; 4(1):51–60

[40] Chen MY, Kundu A, Srihari SN. “Variable duration hidden markov model and morphological segmentation for handwritten word recognition”. IEEE Trans Image Processing 1995; 4(12):1675–1688

[41] Blumenstein M, Verma B .“Neural-based solutions for the segmentation and recognition of difficult handwritten words from a benchmark database” , Proc 5th International Conference on Document Analysis and Recognition, Bangalore, India, 1999; 281–284

[42] Favata JT. “Offline general handwritten word recognition using an approximate beam matching algorithm “ . IEEE Trans Pattern Analysis and Machine Intelligence 2001; 23(9):1009–1021

[43] Powalka RK, Sherkat N, Whitrow RJ. “Word shape analysis for a hybrid recognition system” . Pattern Recognition 1997; 30(3):412–445

[44] مسروری، ک.، کبیر، ا.، " شناسایی برون خط کلمات دستنویس فارسی در یک مجموعه محدود"، رساله دکتری مهندسی برق -الکترونیک، دانشگاه تربیت مدرس،

[45] شیرعلی شهرضا، م.، فائز، ک.، " تشخیص کلمات و ارقام دستنویس فارسی بوسیله شبکه های عصبی (خط نسخ)" رساله دکترای مهندسی برق -کامپیوتر، دانشگاه صنعتی امیر کبیر، 1374

[46] ابراهیمی، ا.، کبیر، ا.، "استفاده از شکل کلی زیر-کلمات چاپی در بازیابی تصویر مستندات و بازشناسی متون فارسی"، رساله دکترای مهندسی برق – الکترونیک، دانشگاه تربیت مدرس، تابستان

84

[47] خسروی، ح.، کبیر، ا.، "ارزیابی روشهای بازشناسی متون فارسی بر مبنای شکل کلی زیر-کلمات" نشریه مهندسی برق و کامپیوتر

[48] خسروی، ح.، کبیر، ا.، "معرفی دو ویژگی سریع و کارآمد برای بازشناسی ارقام دستنویس فارسی"

[49] sergios theodoridis, konstantinos koutroumbas "Pattern recognition", -fourth-edition.9781597492720.52057

[50] Harish Kasiviswanathan, Gregory R. Ball and Sargur N. Srihari, "Top Down Analysis of Line Structure in Handwritten Documents", 1051-4651/10 \$26.00 © 2010 IEEE

[51] Faisal Farooq, Venu Govindaraju, Michael Perrone, "Pre-processing Methods for Handwritten Arabic Documents", 1520-5263/05 \$20.00 © 2005 IEEE

[52] Postl, W. "Detection of linear oblique structure and skew scan in digitized documents", Proceedings of International Conference on Pattern Recognition, (1986) pp. 687- 689.

[53] Yue Lu and Chew Lim Tan. "Improved Nearest Neighbor Based Approach to Accurate Document Skew Estimation", IEEE International Conference on Document Analysis and Recognition (August 2003), pp. 503-507.

[54] Avanindra and Subhasis Chaudhuri (February 1997). "Robust Detection of Skew in Document Images", IEEE Trans. Image Processing, vol. 6(2), pp. 344-349.

[55] Najman, L. "Using Mathematical Morphology for Document Skew Estimation", Proceedings of SPIE conference on Document Recognition and Retrieval (2004), pp. 182-191.

[56] Rondel, M. and Burel, G. "Cooperation of Multi-Layer Perceptrons for the Estimation of Skew Angle in Text Document Images", Proceedings of the International Conference on Document Analysis and Recognition (August 1995), pp. 1141-1144.

Abstract

In this these has been studied an off line Farsi handwritten words recognition system in limited lexicon. The proposed system is based on a two-step method. In the first step, the words in the lexicon are clustered using ISOCLUS and Hierarchical clustering algorithms based on similarity. The similarity features used in this step are upper, lower, vertical projection and background/Ink transitions profile vectors extracted from the each column of image. To reduce the feature vectors dimension and eliminate the diagram shaking, one dimensional Discrete Wavelet Transform is used. DTW algorithm is applied for measuring similarity between the feature vectors extracted from words. The mean of each cluster is used as its representative cluster and its entry in the pictorial dictionary. Total number of handwritten word images in the studied dataset is 16,000 images included 502 city names of Iran. This dataset is called IRANSHAR. In this stage, the handwritten word images are clustered in 62 clusters. In the recognition of input word, first five closest clusters to the input word are selected. This makes lexicon reduction of 77% with 94% recognition rate.

The second stage is recognition of the input unknown word image from the list of candidate words from the first stage. In this stage is used histogram of local gradient feature. For the feature extraction, gradient image is divided into a number of blocks. In this these, these blocks are modified in two methods for improved recognition performance. In first method, the size of the blocks is set based on the distribution of black (pen) pixels. In second method, each main component of handwritten word is divided separately and then they are combined. Finally, the gradient feature vectors of input words are compared with gradient feature vectors of candidate words using KNN and multi-class SVM classifications.

The recognition result on handwritten words of IRANSHAR dataset showed that The lexicon reduction stage increased accuracy and speed relatively due to the elimination of dissimilar word in recognition of handwritten word and the first method modification of the local gradient histogram achieved 13% improvement in proposed system recognition rate.

Keywords: word handwritten recognition, Lexicon reduction, ISOCLUS clustering algorithm, Hierarchical clustering algorithm, DTW algorithm, local gradient histogram, k nearest neighbor classification, multi class SVM classification



Shahrood University of Technology

Thesis Submitted of Msc Degree of Electrical Engineering

OFFLINE FARSI HANDWRITTEN WORD RECOGNITION IN A SET OF WORDS

By:

Elham Bayesteh Tashk

Supervisor:

Dr. Alireza Ahmadyfard

Adviser:

Dr. hossein khosravi

Winter 2012

