



دانشکده: برق و رباتیک

گروه: الکترونیک

# رفع اعوجاجات غیرخطی در تصاویر اسناد فارسی

دانشجو: سمیرا خسروی راد

استاد راهنما:

دکتر حسین خسروی

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

زمستان ۱۳۹۱

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

پاس بی کران پروردگار یکتا را که هستی مان بخشید، به طریق علم و دانش، نمونه‌مان شد، به هم‌نشینی رهروان

علم و دانش مصتخرمان نمود و خوشه‌چینی از علم و معرفت را روزیمان ساخت.

و تقدیر و شکر از استاد کرامت‌دور، جناب آقای دکتر حسین خسروی که بارها همای‌ها و کجک‌های ارزشمند خود،

مراد انجام این پایان‌نامه یاری کردند.

تقدیم به

پدر عزیزم که از نگاهش صلابت، از رفتارش محبت و از صبرش ایستادگی را آموختم

و

همسر مهربانم که سایه مهربانش سایه سازندگیم می باشد و با صبر و دلسوزی مشکلات مسیر را برایم تسهیل نمود

## چکیده

هدف از رفع اعوجاج غیر خطی اسناد، جبران اعوجاج صفحه و مسطح کردن اسناد تصویربرداری یا اسکن شده است به نحوی که این اسناد توسط سیستم OCR، قابل خواندن باشند. اغلب کارهای انجام شده در زمینه رفع اعوجاج غیر خطی تصاویر متن مربوط به اسناد و متونی است که به زبان‌های لاتین می‌باشند. در این پایان‌نامه الگوریتمی جدید جهت رفع چرخش سه بعدی در تصاویر اسناد فارسی پیشنهاد شده است. در الگوریتم پیشنهادی ابتدا پیش‌پردازش‌های لازم شامل رفع نویز، دوسطحی سازی، آغشته‌سازی و تشخیص خطوط زمینه بر روی تصویر انجام شده و سپس کلمات خطوط رفع اعوجاج شده بر روی خط زمینه قرار داده می‌شوند.

در الگوریتم پیشنهادی از الگوریتم آغشته‌سازی DLCCM برای تشخیص خطوط متن استفاده شده است. همچنین تصویر خروجی رفع اعوجاج شده در دو قالب دوسطحی و زمینه خاکستری ارائه می‌شود. جهت ارزیابی عملکرد الگوریتم پیشنهادی از نرم‌افزار نویسه‌خوان فارسی پرشیا نگار استفاده شده است. نتایج تجربی عملکرد مناسب الگوریتم پیشنهادی را در رفع اعوجاج تصاویر نشان می‌دهد.

کلمات کلیدی: نویسه‌خوان نوری، چرخش سه بعدی، مجموعه داده، آغشته سازی، بازسازی تصویر متن

## فهرست مطالب

### فصل اول - مفاهیم اولیه و مروری بر کارهای گذشته

|                                     |    |
|-------------------------------------|----|
| مقدمه.....                          | ۱  |
| ۱-۱ تعریف مسأله .....               | ۲  |
| ۲-۱ مروری بر کارهای گذشته.....      | ۳  |
| ۳-۱ نگاهی اجمالی به فصول آینده..... | ۱۳ |

### فصل دوم - پیش پردازش

|   |    |
|---|----|
| ۱-۲ روش‌های دو سطحی سازی .....                          | ۱۵ |
| مقدمه.....  | ۱۵ |
| ۱-۱-۲ طبقه‌بندی‌ها و مقدمات.....                        | ۱۷ |
| ۱-۱-۱-۲ روش‌های آستانه‌گذاری مبتنی بر هیستوگرام.....    | ۲۰ |
| ۱-۱-۱-۲-۱ آستانه‌گذاری convex hull.....                 | ۲۰ |
| ۱-۱-۱-۲-۲ آستانه‌گذاری قله و دره.....                   | ۲۲ |
| ۱-۱-۱-۲-۳ آستانه‌گذاری مدل کردن شکل shape modeling..... | ۲۳ |
| ۱-۱-۲-۲ روش‌های آستانه‌گذاری مبتنی بر خوشه بندی.....    | ۲۴ |
| ۱-۲-۱-۱-۲ آستانه‌گذاری مکرر.....                        | ۲۵ |
| ۲-۲-۱-۱-۲ آستانه‌گذاری به روش خوشه بندی.....            | ۲۵ |
| ۳-۲-۱-۱-۲ آستانه‌گذاری کمترین خطا.....                  | ۲۵ |

- ۲۶.....آستانه گذاری خوشه بندی فازی.....۴-۲-۱-۱-۲
- ۲۷.....روش های آستانه گذاری مبتنی بر انتروپی.....۳-۱-۱-۲
- ۲۹.....آستانه گذاری انتروپیک.....۱-۳-۱-۱-۲
- ۳۰.....cross-entropic آستانه گذاری.....۲-۳-۱-۱-۲
- ۳۱.....آستانه گذاری فازی انتروپی.....۳-۳-۱-۱-۲
- ۳۳.....آستانه گذاری مبتنی بر تشابه نسبی.....۴-۱-۱-۲
- ۳۴.....آستانه گذاری حفظ گشتاور.....۱-۴-۱-۱-۲
- ۳۴.....آستانه گذاری انطباق میدان لبه.....۲-۴-۱-۱-۲
- ۳۴.....آستانه گذاری تشابه فازی.....۳-۴-۱-۱-۲
- ۳۵.....آستانه گذاری حالت پایدار توپولوژیکی.....۴-۴-۱-۱-۲
- ۳۵.....آستانه گذاری بیشترین اطلاعات.....۵-۴-۱-۱-۲
- ۳۶.....Fuzzy آستانه گذاری بهبود یافته فشردگی.....۶-۴-۱-۱-۲
- ۳۷.....روش های آستانه گذاری فضایی.....۵-۱-۱-۲
- ۳۸.....روش های آستانه گذاری هم رخداد.....۱-۵-۱-۱-۲
- ۴۰.....آستانه گذاری انتروپی درجه بالاتر.....۲-۵-۱-۱-۲
- ۴۱.....آستانه گذاری وفقی محلی.....۶-۱-۱-۲
- ۴۲.....روش های واریانس محلی.....۱-۶-۱-۱-۲
- ۴۳.....روش های کنتراست محلی.....۲-۶-۱-۱-۲
- ۴۳.....طرح های مرکز- محیط.....۳-۶-۱-۱-۲

|         |                                   |
|---------|-----------------------------------|
| ۴۴..... | ۴-۱-۱-۲ آستانه گذاری منطبق بر سطح |
| ۴۴..... | ۵-۶-۱-۱-۲ روش کریگینگ             |
| ۴۵..... | ۲-۲ شناسایی خطوط متن              |
| ۴۵..... | مقدمه                             |
| ۴۵..... | ۱-۲-۲ روش های مبتنی بر افکنش      |
| ۴۷..... | ۲-۲-۲ روش های آغشته سازی          |
| ۴۸..... | ۳-۲-۲ روش های گروهی کردن          |
| ۵۲..... | ۴-۲-۲ روش های مبتنی بر تبدیل هاف  |
| ۵۴..... | ۵-۲-۲ روش شبکه‌ی دافعه-جاذبه :    |
| ۵۵..... | ۶-۲-۲ روش تصادفی                  |

## فصل سوم-الگوریتم پیشنهادی

|         |  |
|---------|--|
| ۵۷..... | مقدمه  |
| ۵۸..... | ۱-۳ فلوجارت الگوریتم پیشنهادی                |
| ۵۹..... | ۲-۳ دریافت داده‌ی اولیه                      |
| ۶۱..... | ۳-۳ رفع نویز                                 |
| ۶۲..... | ۴-۳ دوسطحی سازی                              |
| ۶۶..... | ۵-۳ تشخیص خطوط متن                           |
| ۶۶..... | ۱-۵-۳ آغشته سازی با استفاده از dilation      |
| ۶۶..... | ۲-۵-۳ آغشته سازی با استفاده از الگوریتم DLDM |

|  |    |
|--|----|
| ۶-۳ تشخیص زیر کلمات و تعیین شیب آن‌ها.....             | ۷۲ |
| ۷-۳ رفع چرخش کلمات و خطوط.....                         | ۷۴ |
| ۸-۳ ارائه تصویر خروجی.....                             | ۸۱ |
| ۱-۸-۳ رفع اعوجاج تصویر اولیه.....                      | ۸۱ |
| ۲-۸-۳ رفع تداخل ایجاد شده در هنگام چرخش مستطیل‌ها..... | ۸۲ |
| ۹-۳ نتیجه‌گیری.....                                    | ۸۴ |

## فصل چهارم - نتایج تجربی

|   |    |
|---|----|
| مقدمه.....  | ۸۵ |
| ۱-۴ نتایج تجربی الگوریتم پیشنهادی.....                            | ۸۵ |
| ۲-۴ بررسی عملکرد الگوریتم پیشنهادی توسط نرم افزار پرشیا نگار..... | ۹۱ |
| ۳-۴ نتیجه‌گیری.....   | ۹۷ |

## فصل پنجم - نتیجه‌گیری و پیشنهادات

|                                      |     |
|--------------------------------------|-----|
| ۱-۵ نتیجه‌گیری.....                  | ۹۹  |
| ۲-۵ پیشنهادات برای کارهای آینده..... | ۱۰۰ |

## منابع

|            |     |
|------------|-----|
| منابع..... | ۱۰۱ |
|------------|-----|

## فصل اول

مفاهیم اولیه و مروری بر کارهای گذشته

## مقدمه

OCR، کوته نوشت بازشناسی نوری نویسه‌ها<sup>۱</sup> یا نویسه خوان نوری<sup>۲</sup> است. OCR ابتدا تنها در مورد بازشناسی ارقام و حروف چاپی بکار گرفته شد. واژه نوری در مقابل عبارت مرکب مغناطیسی قرار داده شد تا این روش را از روش قدیمی‌تر بازشناسی نویسه‌ها با مرکب مغناطیسی<sup>۳</sup> (MICR) متمایز کند.

با گذشت زمان و پیشرفت قابل توجه در این زمینه، روش‌های بازشناسی دستنوشته و متون چاپی مطرح شدند که دامنه کاربرد را به کلمات و عبارات رساندند. با وجود عدم تطبیق OCR با این موارد این نام برای این روش -ها و تا حدی برای بازشناسی دستنوشته‌های بر خط<sup>۴</sup> هم استفاده شد و رواج پیدا کرد. این روزها OCR را بیشتر برای بازشناسی مستندات چاپی، مثل صفحات کتاب، مجله‌ها و نامه‌های چاپی به کار می‌برند. [۱]

یک سامانه OCR مثل یک نفر ماشین نویس، متن را می‌خواند و آن را به قالب مناسب برای ذخیره در کامپیوتر تبدیل می‌کند. سامانه OCR اشیاء موجود در تصویر را که ارقام، حروف، علائم و کلمات هستند بازشناسی می‌کند و در قالب مناسب ذخیره سازی می‌کند. معمولاً تصویر متن که به عنوان داده اولیه برای سیستم OCR است، توسط یک روبشگر یا دوربین دیجیتال فراهم می‌شود.

اسکنر وسیله‌ای است که از سندهای چاپی، پوسترها و صفحات مجله تصویربرداری می‌کند. اسکنرها در انواع مختلفی از جمله flatbed scanner, sheet-fed scanner, handheld scanner, drum scanner و... وجود دارند. آن‌ها می‌توانند تصویر را به صورت سیاه و سفید یا رنگی روبش کنند. اسکنرها کاربردهای زیادی در مصارف خانگی، اداری و تجاری برای در دسترس قرار دادن اطلاعات و کم کردن حجم اطلاعات کاغذی دارند. از مزایای

---

<sup>۱</sup> Optical Character Recognition

<sup>۲</sup> Optical Character Reader

<sup>۳</sup> Magnetic Ink Character Recognition

<sup>۴</sup> Online

اسکنر می‌توان به بالا بردن امنیت اطلاعات، آسان و سریع شدن دستیابی به اطلاعات و کم کردن حجم بایگانی ادارات نام برد.

علاوه بر اسکنر، دوربین های دیجیتال قابل اتصال به کامپیوتر، به علت در دسترس بودن، سادگی استفاده، حجم بالای ذخیره اطلاعات و... نقش قابل توجهی در تهیه و آنالیز تصویر متن دارند. استفاده از ادوات تصویربرداری دیجیتال همچون دوربین های موبایل و دوربین های کوچک دیجیتال به عنوان روشی ساده، کارآمد و کم هزینه برای جمع آوری اطلاعات به حساب می آیند. دوربین ها به عنوان ابزاری مفید در تصویر برداری از کتب تاریخی، کتاب های ضخیم و متون چاپی که دست زدن به آنها موجب تخریب آنها می شود، به شمار می روند. هر محیطی که محتوایی از نوع متن در خود جای داده باشد (مثل یک صفحه ی کاغذ، فریم ویدیویی حاوی متن، شماره ی پلاک وسیله ی نقلیه، کتاب باز شده) می تواند به عنوان داده ی اولیه برای سیستم OCR به حساب آید.

## ۱-۱ تعریف مسأله

داده ی ایده آل برای یک سیستم OCR، یک تصویر متنی است که دارای تصویر و فرمول نباشد. متأسفانه در درصد بسیار کمی از تصاویر اسناد، داده ی اولیه ی ما ایده آل است. در نتیجه چالش های جدید برای نزدیک کردن داده ی اولیه ی OCR به داده ی ایده آل به وجود می آید. از عوامل مخرب تصویر متنی می توان به وضوح پایین، روشنایی متغیر صفحه، سطوح ناصاف، پس زمینه ی مختلط متن، زوم شدن دوربین و تمرکز روی بخش خاص، حرکت اشیاء حین تصویر برداری و چرخش سه بعدی<sup>۱</sup> اشاره کرد. وقتی یک صفحه از یک کتاب قطور را اسکن می کنیم یا از آن عکس می گیریم در تصویر متن یک بالآمدگی به وجود می آید. این بالا آمدگی در تمام صفحه یکسان نیست و غیر خطی می باشد که به آن پیچش سه بعدی یا warping می گوئیم. warping

---

<sup>۱</sup> warping

در لغت به معنای پیچ خوردن یا تابانده شدن است. تصویر متنی که در آن warping وجود داشته باشد، دقت OCR را پایین می آورد.

پیچش سه بعدی می تواند به علت تکنیک های خاصی که در اسکن تصویر توسط اسکنر های مختلف وجود دارد، ایجاد شود. پیچش سه بعدی در اسکن کاغذ های تاریخی مچاله شده یا تاخورده و چند بار استفاده شده نیز وجود دارد. از علت های دیگر در به وجود آمدن پیچش سه بعدی ، ماده پرینت شده، تنظیمات دوربین، رطوبت، شرایط محیطی و ... است. این اعوجاج ناخواسته علاوه بر کاهش خوانایی متن باعث می شود، دقت سامانه OCR کم شود. پس رفع پیچ خوردگی صفحه<sup>۱</sup> به عنوان چالشی برای بهبود دقت OCR می باشد و از مراحل پیش پردازش در فرآیند OCR است.

هدف ما در این پروژه جبران اعوجاج صفحه، مسطح کردن صفحه متن اسکن شده یا تصویر برداری شده پیش از مرحله OCR است بطوری که دقت OCR سند بهبود یابد. اعوجاج مورد نظر در کار ما، اعوجاجات حاصل از پدیده های خمیدگی کاغذ و پرسپکتیو (warping) است.

## ۱-۲ مروری بر کارهای گذشته

در سالهای اخیر روش های متعددی در زمینه رفع اعوجاج چرخش سه بعدی در تصاویر متنی لاتین پیشنهاد شده است. این تکنیک ها به دو دسته کلی تقسیم می شوند. دسته اول تکنیک های مبتنی بر بازسازی سه بعدی تصاویر متن هستند و دسته دوم تکنیک هایی که با استفاده از اطلاعات دو بعدی متن، آن را پردازش می کنند.

تکنیک های مبتنی بر بازسازی شکل سه بعدی تصاویر، خود به دو دسته ی سخت افزاری و نرم افزاری تقسیم می شوند. معمولاً در این روش یک شکل هندسی با مشخصات از پیش دانسته روی متن منعکس می شود.

---

<sup>۱</sup> dewarping

انعکاس آن در تصویر دیجیتال شناسایی می‌شود و با استفاده از تفاوت‌های شکل و انعکاس آن مدل اعوجاج بدست می‌آید. [۲]

روش‌های سخت افزاری برای بازسازی سه بعدی تصاویر متن مبتنی بر استفاده از تجهیزات خاص است، مثلاً در [۳] اسکنرهای لیزری خاصی طراحی کرده‌اند. استفاده از دوربین‌های استریو<sup>۱</sup> در [۴,۵] پیشنهاد شده است. بکارگیری ابزارهای فتومتریک مانند laser projector در [۶] و تنظیمات نوری خاص در دوربین در [۷] ارائه شده است. در [۸] هم از روش structured light ۳D acquisition بهره گرفته شده است.

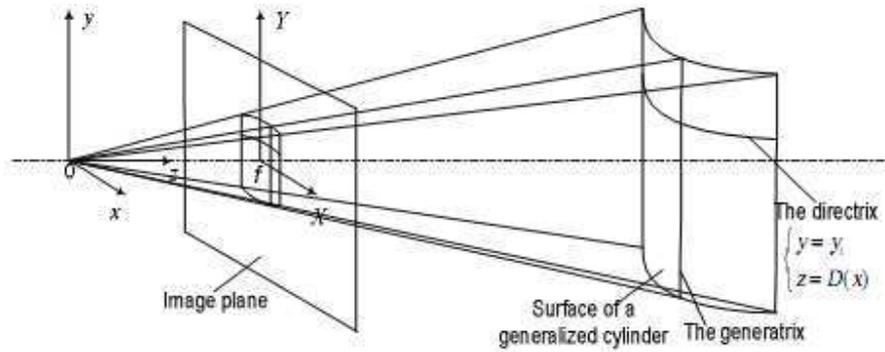
استفاده از سخت افزار سنگین و سخت است و برای حجم زیاد اطلاعات از نظر اقتصادی مقرون به صرفه نیست، و مشخصاً برای میلیون‌ها متن موجود دیجیتال شده که دارای اعوجاج هستند مناسب نیست و کاربردی ندارد. علاوه بر آن این روش‌ها در محیطی که تجهیزات در دسترس نیستند و به تصویربرداری یا اسکن کردن تصویر نیاز است بکار نمی‌آیند.

روش‌های نرم افزاری پردازش سه بعدی تصویر متن، تخمین شکل سه بعدی تصویر متن است که سعی می‌کند با استفاده از مدل‌هایی شکل سطح سه بعدی تصویر متن را برآورد کند.

در [۹] روشی برای یکسوسازی اعوجاج‌های ناخواسته بیان شده است که در آن با به وجود آوردن یک مدل استوانه‌ای سعی در تخمین شکل سه بعدی شده است. جدای از این که آیا واقعاً شکل استوانه‌ای است یا نه، محدودیت دیگر این روش وضعیت آن است، طوری که سطح تصویر موازی با مولد استوانه‌ای صفحه باشد. شکل ۱-۱ روند تشکیل تصویر سطح استوانه‌ای سند را نشان می‌دهد.

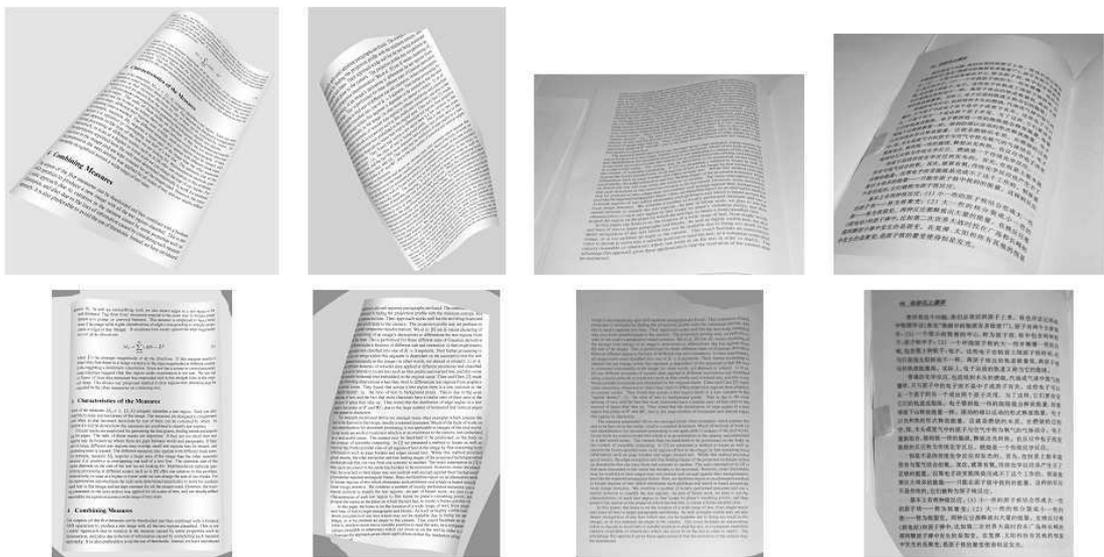
---

<sup>۱</sup> Stereo camera



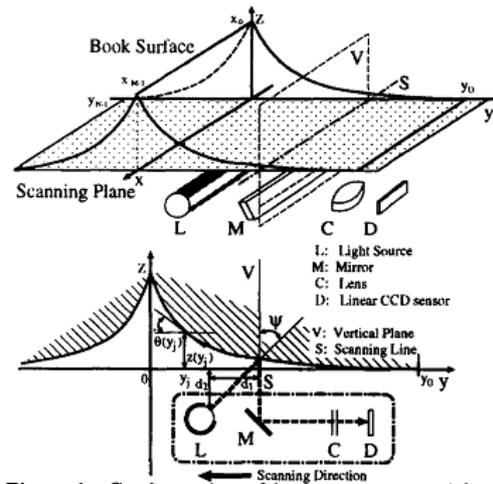
شکل ۱-۱: تشکیل سطح استوانه ای تصویر سند [۹]

در [۱۰] روشی پیشنهاد شده است که سطح صفحه را با سطوح پیچش دار قابل توسعه مدل می کند تا شکل سه بعدی صفحه را با استفاده از flow field های بافت تخمین بزند. فرض این روش نرم بودن پیچش صفحه است و این که دوربین یک دوربین استاندارد روزنه‌ای<sup>۱</sup> است. شکل ۲-۱ مقایسه ای از تصاویر ورودی و خروجی برای این روش را نشان می دهد.



شکل ۲-۱: مقایسه ای بین اسناد پیچش دار (ردیف بالا) با نتایج یکسوشده (ردیف پایین) [۱۰]

روش دیگری در [۱۱] از shape from shading استفاده می‌کند تا شکل سه بعدی سطح متن را بازسازی کند. این تکنیک نیاز به دانستنی‌هایی در مورد نور دارد که معمولاً ناشناخته است. شکل ۳-۱ پیکربندی اسکنر تصویر و سطح کتاب را نشان می‌دهد.

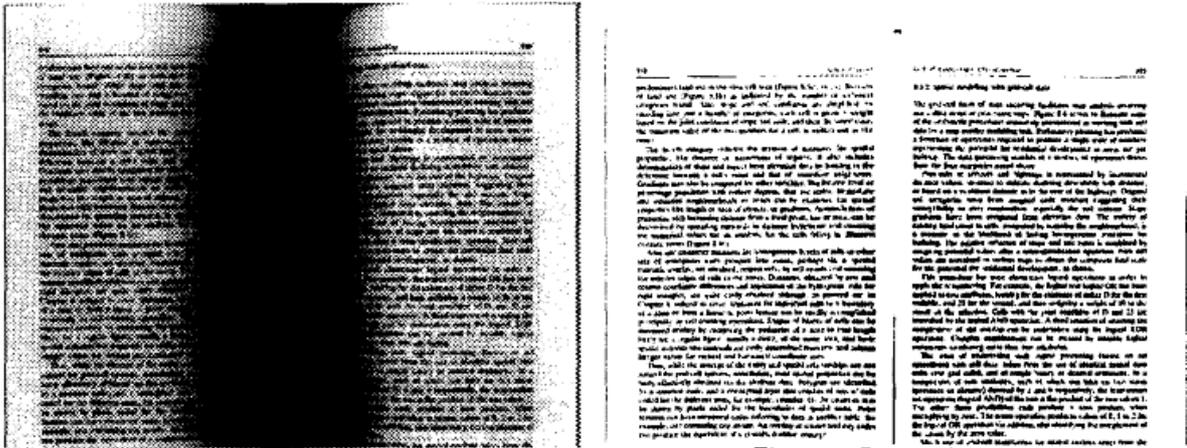


شکل ۳-۱: پیکربندی اسکنر تصویر و سطح کتاب [۱۱]

در [۱۲] یک روش deskewing عمومی برای تصاویر پیچانده شده دلخواه، بر مبنای شکل سه بعدی ارائه شده است. در این روش ابتدا عمق هر نقطه را در تصویر با روش‌های بینایی سلول بدست می‌آورند تا یک نگاشت عمق بسازند، سپس تصویر را با توجه به آن نگاشت رفع پیچش می‌کنند. مشکل این روش نگاشت نقاط روی سطوح سخت و نویزی به عمق نقاط روی سطوح صاف است.

در [۱۳] هم بازیابی سطح سه بعدی کتاب با ترکیب بازتاب‌های داخلی انجام شده است که فرض‌های این پیشنهاد موازی بودن تیر کتاب با نور اسکن کننده و نیز استوانه‌ای بودن سطح کتاب است. عیب این روش حجم بالای محاسبات است که حتی پس از استفاده از روش موزائیک سازی<sup>۱</sup> نیز کاملاً برطرف نشد. شکل ۴-۱ تصاویر ورودی و خروجی این الگوریتم را نشان می‌دهد.

<sup>۱</sup> tessellation



شکل ۱-۴: تصویر مشاهده شده (سمت چپ)، تصویر خروجی (سمت راست) [۱۳]

دسته دوم که تکنیک‌های مبتنی بر اطلاعات دو بعدی هستند، سعی در شناسایی خطوط متن دارای اعوجاج دارند. بعضی از این روش‌ها یک الگوریتم را پیشنهاد می‌کنند تا بوسیله تقریب زدن یک مدل به هر خط، متن دارای اعوجاج را صاف کنند.

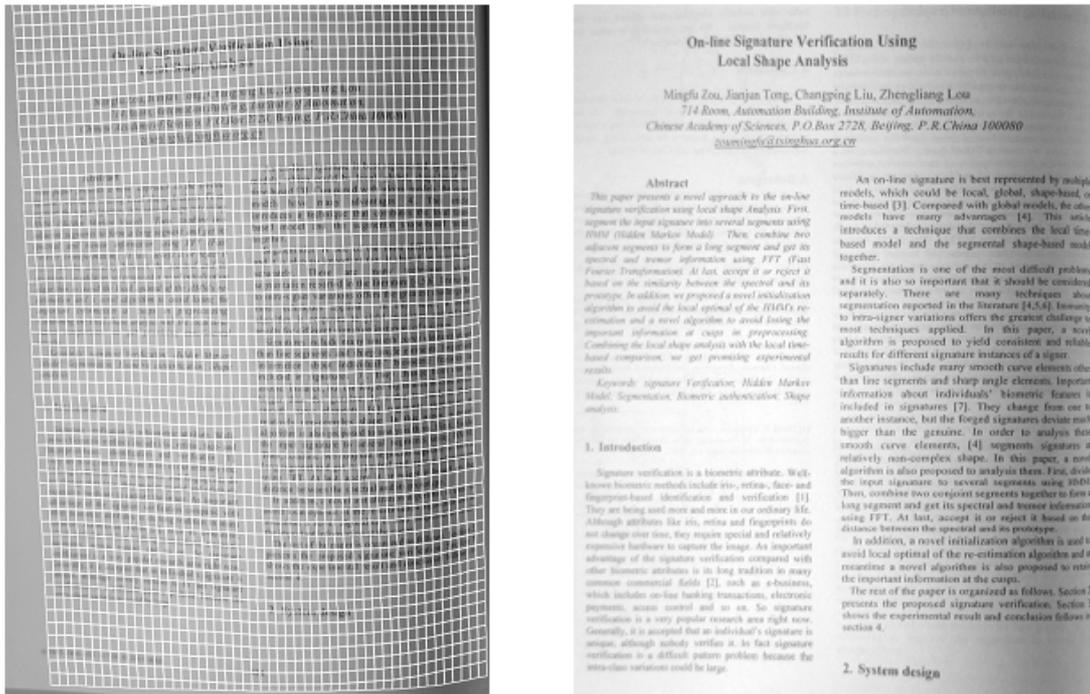
در [۱۴] با استفاده از یک مدل تحلیلی، B-spline های Cubic برای شناسایی خط استفاده شده است. اما این روش در صورت ناهمگون بودن فاصله بین خطوط ناکارآمد است. شکل ۱-۵ نتایج حاصل از این الگوریتم را نشان می‌دهد.

people are lesse Philofophers and more foolifih, then *Platoes* Phylodoxes, or lovers of their owne opinions. We must knowe whether fire be hot, whether fnowe be white, whether in our knowledge there be any thing hard or soft. And touching the answers, wherefore they telld tales, as to him who made a doubt of heate, to whom one replied, that to try he should cast himselfe into the fire; to him that denied the yfe to be cold, that he should put some in his bofome; they are most vnworthie the profession of a Philofopher. If they had leaft vs in our owne naturall estate, admitting of strange apperances, as they present themselves vnto vs by our senses, and had suffred vs to follow our naturall appetites, directed by the condition of our birth, they should then have reason to speake so. But from them it is, that we have learnt to become judges of the world; it is from them we hold this conceit, that mans reason is the generall controuler of all that is, both without and within heavens-vaults, which embraceth al, and can do all, by meanes whereof all things are knowne and discerned. This answer were good among the Caniballs, who without any of *Aristotles* precepts, or so much as knowing the name of Phisike, enjoy most happily, a long, a quiet, and a peaceable life. This answer might happily avaiile more, and be of more force, then all those they can borrow

people are lesse Philofophers and more foolifih, then *Platoes* Phylodoxes, or lovers of their owne opinions. We must knowe whether fire be hot, whether fnowe be white, whether in our knowledge there be any thing hard or soft. And touching the answers, wherefore they telld tales, as to him who made a doubt of heate, to whom one replied, that to try he should cast himselfe into the fire; to him that denied the yfe to be cold, that he should put some in his bofome; they are most vnworthie the profession of a Philofopher. If they had leaft vs in our owne naturall estate, admitting of strange apperances, as they present themselves vnto vs by our senses, and had suffred vs to follow our naturall appetites, directed by the condition of our birth, they should then have reason to speake so. But from them it is, that we have learnt to become judges of the world; it is from them we hold this conceit, that mans reason is the generall controuler of all that is, both without and within heavens-vaults, which embraceth al, and can do all, by meanes whereof all things are knowne and discerned. This answer were good among the Caniballs, who without any of *Aristotles* precepts, or so much as knowing the name of Phisike, enjoy most happily, a long, a quiet, and a peaceable life. This answer might happily avaiile more, and be of more force, then all those they can borrow

شکل ۱-۵: تصویر ورودی دارای اعوجاج (سمت چپ)، تصویر خروجی (سمت راست) [۱۴]

در [۱۵] از یک منحنی غیر خطی برای هر خط از مش استفاده می‌کند که عیب آن این است که از عهده ستون‌های غیر یکسان<sup>۱</sup> در مش هدف بر نمی‌آید. شکل ۱-۶ مش پیچش دار برای یک قالب چند ستونه را نشان می‌دهد و تصویری را که توسط این مش تصحیح شده است، نمایش داده شده است.

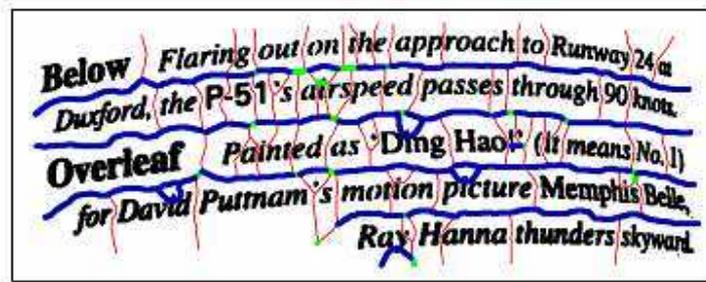


شکل ۱-۶: مش پیچش دار برای یک قالب چند ستونه (سمت چپ) و متن تصحیح شده توسط مش (سمت راست) [۱۵]

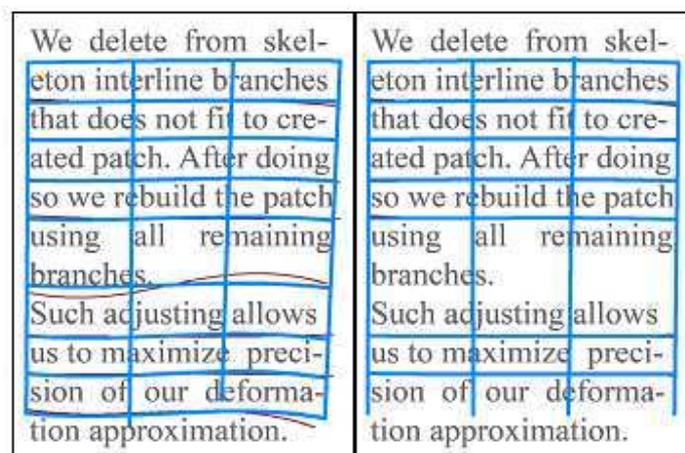
در [۱۶] از نمایش اسکلتی پیوسته تصویر سند استفاده شده است تا فضاهای بین خطی تصویر را که بین خطوط متن قرار گرفته‌اند، تخمین بزند. در ضمن روشی برای تخمین بد فرمی کل تصویر با استفاده از فضاهای بین خطی ارائه شده است. شکل ۱-۷ اسکلت تصویر سند پس از خوشه بندی شاخه‌ها و شکل ۱-۸ تخمین بد فرمی<sup>۲</sup> تصویر را نشان می‌دهد.

<sup>۱</sup> nonuniform

<sup>۲</sup> deformation



شکل ۷-۱ اسکلت تصویر سند پس از خوشه بندی شاخه ها [۱۶]



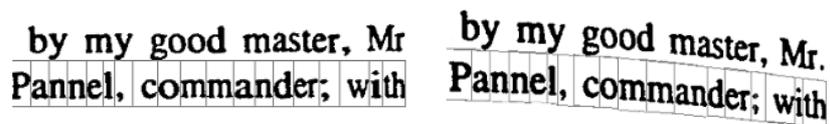
شکل ۸-۱ تخمین بدفرمی تصویر [۱۶]

در [۱۷] از spline های Cubic برای مدل کردن خطوط دارای اعوجاج و فضای بین آنها استفاده می کند. در این روش فرض بر این است که تیر<sup>۱</sup> کتاب در طول خطهای ایزوپارامتریک است. در این روش از محاسبات پیچیده‌ای برای مدل کردن پیچش استفاده می کند و بازسازی را به یک سری پیچش‌های خاص محدود می کند.

در [۱۸] از تخمین یک سلول quadrilateral برای هر حرف، بر پایه‌ی خط زمینه‌ی محلی استفاده شده که آنرا به یک مستطیل در سایز تصحیح شده با موقعیتی در تصویر رفع پیچش شده نگاشت می دهد.

<sup>۱</sup> Book spine

در [۱۹] الگوریتمی برای حذف اعوجاج پرسپکتیو و پیچش صفحه<sup>۱</sup> ارائه داده است. در این الگوریتم فرض بر این است که متن حاوی خطوط کاملاً موازی با فاصله خط یکسان است. در شکل ۱-۹ زنجیره سلول ها در تصویر ورودی و خروجی نشان داده شده است. جدول ۱-۱ نرخ خطا را قبل و بعد از پیاده سازی الگوریتم در نرم افزار FineReader نشان می‌دهد.



شکل ۱-۹: زنجیره سلول ها در تصاویر ورودی (سمت چپ) و خروجی (سمت راست) [۱۹]

جدول ۱-۱: نرخ های خطای OCR در نرم افزار FineReader قبل و بعد از پیاده سازی الگوریتم [۱۹]

| input font     | error rate on input | error rate on output |
|----------------|---------------------|----------------------|
| times 9pt      | 17.9%               | 1.6%                 |
| times 12pt     | 7.2%                | 1.5%                 |
| arial 9pt      | 14.5%               | 0.8%                 |
| arial 12pt     | 10.7%               | 0.0%                 |
| <b>average</b> | <b>12.6%</b>        | <b>1.0%</b>          |

در [۲۰] از تبدیلات فضایی<sup>۲</sup> بین تصویر پیچش دار و رفع پیچش شده استفاده شده است. فرض بر این است که سطح عکس، استوانه است، بنابراین از یک تبدیل استفاده می‌شود تا تصویر متن مسطح شود. این روش به محاسبات پیچیده نیاز دارد، زمان بر است و فرض این که یک تک استوانه بر یک صفحه دارای اعوجاج برازش<sup>۳</sup> شود فرض خوبی نیست.

<sup>۱</sup> Page curl

<sup>۲</sup> Spatial

<sup>۳</sup> fit

از میان روش‌های گفته شده در دسته دوم سه روش عمده جبران اعوجاج تصویر با نام‌های ارائه‌ی تصویر اسکلتی پیوسته برای جبران اعوجاج تصویر متن (SKEL)<sup>۱</sup>، بخش بندی برای جبران اعوجاج تصویر متن (SEG)<sup>۲</sup>، مدل انتقال مختصات و یکسوسازی متن برای جبران اعوجاج کتاب (CTM)<sup>۳</sup>، بیشتر مورد توجه هستند. [۲۱]

ایده‌ی اصلی روش SKEL بر اساس استخراج اسکلت بیرونی تصاویر متنی می‌باشد که در آن شاخه‌های به هم پیوسته که توسط منحنی‌هایی تقریب زده می‌شوند، فضای درون خطوط متن را تعیین می‌کنند. روش CTM با استفاده از یک مدل انتقال مختصات استوانه‌ای به کارتزین و در ادامه با یکسوسازی<sup>۴</sup> متن، اعوجاج چرخش سه-بعدی در کتاب را برطرف می‌کند. همچنین روش SEG بر اساس بخش بندی کلمات موجود در متن و تصحیح شیب این کلمات با استفاده از خطوط کرسی تعیین شده برای آن‌ها و در نهایت انتقال عمودی این کلمات چرخیده شده است. این سه روش، در کاهش اعوجاج چرخش سه‌بعدی متون لاتین کاربرد دارند، اما هیچ‌کدام از این روش‌ها نمی‌توانند به طور کامل، اعوجاجات مربوط به این متون را برطرف کنند. روش SKEL اعوجاجات انحنای صفحه را می‌تواند برطرف کند ولی اعوجاجات پرسپکتیو را برطرف نمی‌کند. همچنین اعوجاجات مربوط به فرمول‌ها و مناطق متنی را که از گوشه‌های صفحات مجاور پدیدار می‌شوند را نیز برطرف نمی‌کند. روش SEG، مناطق متنی را که از صفحات مجاور پدیدار می‌شوند و فرمول‌ها را تا حدی تصحیح می‌کند، اما اعوجاجات پرسپکتیو و اعوجاجات مربوط به انحنای صفحه را نمی‌تواند برطرف کند. روش SEG، مناطق متنی را که از صفحات مجاور پدیدار می‌شوند، اعوجاج مربوط به فرمول‌ها و انحنای صفحه را برطرف می‌کند، اما در از بین بردن اعوجاج پرسپکتیو ناتوان است.

---

<sup>۱</sup> Skeletonization

<sup>۲</sup> Segmentation

<sup>۳</sup> Cao Transform Model

<sup>۴</sup> Rectification

در تمامی مقالات بررسی شده تاکنون الگوریتم‌های ارائه شده بر روی تصاویر متنی لاتین است و کارهای انجام شده در زمینه رفع اعوجاجات هندسی تصاویر متن فارسی محدود بوده است.

در [۲۱] الگوریتمی برای رفع اعوجاج در تصاویر شامل متن‌های فارسی که با یک دوربین دیجیتال گرفته شده- اند، پیشنهاد شده است. روش پیشنهادی روی تصاویر سطح خاکستری اعمال می‌شود. مراحل این روش شامل دوسطحی کردن تصویر ورودی، آشکارسازی خطوط متن و نوار زمینه می‌باشد. سپس اعوجاج هندسی ناشی از تصویربرداری رفع می‌گردد. تصویر نمونه ورودی و خروجی برای این الگوریتم در شکل ۱-۱۰ و ۱-۱۱ آمده است.

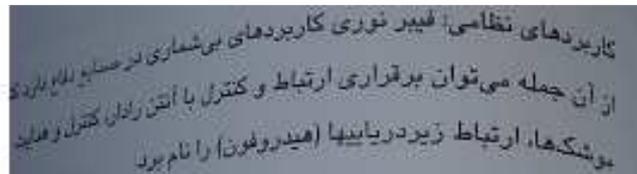
شکل ۱۴-۸ فرآیند رمزگذاری دوبعدی را برای پیمایش یک خط نشان می‌دهد. توجه کنید که مراحل اولیه این روش، به یافتن چندین عنصر متغیر<sup>۳</sup> مهم هدایت شده‌اند:  $a_0, a_1, a_2, b_1$  و  $b_2$ . عنصر متغیر، توسط استاندارد، به عنوان پیکسلی تعریف می‌شود که مقدارش متفاوت از پیکسل قبلی در همان خط است. مهمترین عنصر متغیر  $a_0$  (عنصر مرجع) است، که یا برابر با مکان عنصر متغیر سفید فرضی در سمت چپ اولین پیکسل هر خط رمزگذاری می‌شود یا از حالت رمزگذاری قبلی تعیین می‌شود. حالت‌های رمزگذاری در ادامه بحث می‌شود. پس از تعیین مکان  $a_0, a_1$  به عنوان مکان عنصر متغیر بعدی در سمت راست  $a_0$  روی خط رمزگذاری فعلی،  $a_2$  به عنوان عنصر متغیر بعدی در سمت راست  $a_1$  روی خط رمزگذاری فعلی،  $b_1$  به عنوان عنصر متغیر با مقدار مخالف (با  $a_0$ ) و در سمت راست  $a_0$  در خط مرجع (با خط قبلی)، و  $b_2$  به عنوان عنصر متغیر بعدی در سمت راست  $b_1$  روی خط مرجع مشخص می‌شود. اگر هر کدام از این عناصر متغیر تشخیص داده نشوند، مکان آن، نسبت به یک فرضی در سمت راست آخرین پیکسل، روی خط مناسب تعیین می‌شود. شکل ۱۵-۸ دو نمونه از روابط کلی بین عناصر متغیر مختلف را نشان می‌دهد.

شکل ۱-۱۰: تصویر نمونه ورودی [۲۱]

شکل ۱۴ ۸ فرآیند رمزگذاری دوبعدی را برای پیمایش یک خط نشان می‌دهد. توجه کنید که مراحل اولیه این روش، به یافتن چندین عنصر متغیر<sup>۳</sup> مهم هدایت شده‌اند:  $a_0, a_1, a_2, b_1$  و  $b_2$ . عنصر متغیر، توسط استاندارد، به عنوان پیکسلی تعریف می‌شود که مقدارش متفاوت از پیکسل قبلی در همان خط است. مهمترین عنصر متغیر  $a_0$  (عنصر مرجع) است، که یا برابر با مکان عنصر متغیر سفید فرضی در سمت چپ اولین پیکسل هر خط رمزگذاری می‌شود یا از حالت رمزگذاری قبلی تعیین می‌شود. حالت‌های رمزگذاری در ادامه بحث می‌شود. پس از تعیین مکان  $a_0, a_1$  به عنوان مکان عنصر متغیر بعدی در سمت راست  $a_0$  روی خط رمزگذاری فعلی،  $a_2$  به عنوان عنصر متغیر بعدی در سمت راست  $a_1$  روی خط رمزگذاری فعلی،  $b_1$  به عنوان عنصر متغیر با مقدار مخالف (با  $a_0$ ) و در سمت راست  $a_0$  در خط مرجع (با خط قبلی)، و  $b_2$  به عنوان عنصر متغیر بعدی در سمت راست  $b_1$  روی خط مرجع مشخص می‌شود. اگر هر کدام از این عناصر متغیر تشخیص داده نشوند، مکان آن، نسبت به یک فرضی در سمت راست آخرین پیکسل، روی خط مناسب تعیین می‌شود. شکل ۱۵ ۸ دو نمونه از روابط کلی بین عناصر متغیر مختلف را نشان می‌دهد.

شکل ۱-۱۱: تصویر نمونه خروجی [۲۱]

در [۲۲] الگوریتمی برای کاهش اعوجاج چرخش سه بعدی روی تصاویر متنی حاصل شده از دوربین دیجیتال پیاده‌سازی شده است. ایده اصلی الگوریتم، بر اساس بخش‌بندی عناصر تشکیل دهنده متن می‌باشد. در این الگوریتم، خطوط کرسی بالا و پایین کلمات مشخص می‌شود و از آن برای تعیین شیب کلمات استفاده می‌گردد. همچنین انتقال کلمات چرخیده شده با توجه به برازش یک منحنی درجه دوم به مراکز کلمات یک خط نسبت به خط افق صورت می‌پذیرد. تصویر ورودی و تصویر حاصل از پیاده‌سازی این الگوریتم در شکل ۱-۱۲ نشان داده شده است.



شکل ۱-۱۲: تصویر ورودی (بالا) و تصویر خروجی (پایین) [۲۲]

### ۳-۱ نگاهی اجمالی به فصول آینده

متن حاضر ۵ فصل را شامل می‌شود. در فصل دوم که عنوان آن پیش پردازش می‌باشد، به دو موضوع اصلی در پیش‌پردازش تصاویر اسناد فارسی قبل از ورود به مرحله OCR پرداخته است. در بخش اول، دو سطحی سازی به عنوان یک مرحله مهم در پیش‌پردازش مورد بررسی قرار گرفته و روش‌های مختلف دو سطحی سازی متون در مراجع گوناگون بیان شده است و در بخش دوم انواع روش‌های شناسایی خطوط تصویر سند بحث شده است.

فصل سوم را می‌توان اصلی‌ترین فصل متن حاضر دانست که در آن به بررسی جزئیات الگوریتم پیشنهادی پرداخته شده است. تهیه مجموعه داده، رفع نویز، دوسطحی‌سازی، تشخیص خطوط متن، تشخیص زیرکلمات و تعیین شیب آن‌ها، رفع چرخش کلمات و خطوط عناوین اصلی الگوریتم پیشنهادی هستند.

در فصل چهارم با عنوان نتایج تجربی، بررسی عملکرد الگوریتم پیشنهادی روی چند نمونه و تحلیل آن‌ها انجام شده است.

در نهایت، فصل پنجم شامل نتیجه‌گیری و پیشنهادات است. در بخش نخست نتیجه‌گیری در مورد روش پیشنهادی بیان می‌گردد. بخش دوم نیز به ارائه پیشنهادات جهت ادامه پروژه اختصاص یافته است.

فصل دوم

پیش پردازش

## ۱-۲ روش‌های دو سطحی سازی

### مقدمه

در بسیاری از کاربردهای پردازش تصویر، سطوح خاکستری پیکسل‌های متعلق به شیء اساساً با سطوح خاکستری پیکسل‌های پس زمینه متفاوتند. آستانه‌گذاری در این مورد ابزاری ساده اما کارآمد برای جداکردن اشیاء از پس زمینه است. یکی از مثالهای کاربردهای آستانه‌گذاری، تحلیل تصویر سند به منظور استخراج نویسه‌های چاپ شده، لوگوها و محتوای گرافیکی یا نشان‌های موسیقی است. از دیگر کاربردها می‌توان به پردازش نقشه با هدف کشف خطوط، نقوش یا نویسه‌ها، پردازش صحنه به منظور شناسایی یک هدف و بازرسی کیفیت مواد با هدف مشخص کردن قسمت‌های معیوب نام برد. [۲۳] سایر کاربردها هم به شرح زیر است: تصاویر سلولی، نمایش یا ارائه دانش، بخش‌بندی کیفیت‌های متفاوت تصویر برداری برای کاربردهای آزمایشات غیر مخرب (NDT)<sup>۱</sup> مانند تصاویر آلتراسونیک، تصاویر جریان مخالف<sup>۲</sup>، تصاویر گرمایی<sup>۳</sup>، توموگرافی محاسبه شده اشعه X<sup>۴</sup>، تصاویر آندوسکوپی، میکروسکوپی تصویربرداری لیزری هم کانون، استخراج لبه صفحه، بخش‌بندی عمومی تصویر، بخش‌بندی سه بعدی گرمایی تصاویر ویدیویی و ...

خروجی عملیات آستانه‌گذاری یک تصویر دودویی است که می‌توان یک بخش آن را اشیاء و مواد پیش‌نما<sup>۵</sup> دانست که شامل یک متن چاپ شده، علائم و اختصارات، یک هدف، بخش معیوب یک جنس و ... باشد و بخش مکمل دیگر را مرتبط به پس زمینه در نظر گرفت. بر اساس کاربرد، پیش‌نما می‌تواند با سطح خاکستری که برای متن، به رنگ سیاه است و پس زمینه را بیشترین روشنایی صفحه‌ی سند قرار داد که در تصاویر ۸ بیتی ۲۵۵ است و یا بطور کاملاً بر عکس سفید برای پیش‌نما و سیاه برای پس زمینه.

<sup>۱</sup> Nondestructive Testing

<sup>۲</sup> eddy current

<sup>۳</sup> thermal images

<sup>۴</sup> X-ray computed tomography ( CAT)

<sup>۵</sup> foreground

فاکتورهای متعددی همچون نویز غیر ایستا و هم بسته، روشنایی محدود، کنتراست ناکافی و اندازه نامتناسب شیء با صفحه، اشتغال سطوح خاکستری هم در مش و هم در پس زمینه، باعث پیچیدگی عملیات آستانه گذاری می شود. در نهایت نبود اندازه های هدف برای ارزیابی عملکرد الگوریتم های آستانه گذاری مختلف و سخت بودن آزمودن گسترده در یک محیط وظیفه گرا از دیگر موانع کار هستند.

در این بخش یک طبقه بندی از الگوریتم های آستانه گذاری ارائه می شود. شش دسته اصلی برای الگوریتم های آستانه گذاری در نظر گرفته شده که عبارتند از الگوریتم هایی بر مبنای استفاده از: ۱. اطلاعات شکل هیستوگرام ۲. خوشه بندی فضای اندازه گیری ۳. اطلاعات انتروپی هیستوگرام ۴. اطلاعات نسبی تصویر ۵. اطلاعات فضایی ۶. مشخصه های محلی

یک سیستم تحلیل تصویر سند شامل قسمتهای مختلف پردازش تصویر می شود که با دیجیتال کردن سند شروع شده و به تشخیص نویسه و پردازش زبان طبیعی ختم می شود. مرحله آستانه گذاری به میزان بسیار زیادی بر روی عملکرد گام های بعدی مانند طبقه بندی سند در اشیاء متن و نیز OCR تاثیر می گذارد.

آستانه گذاری نامناسب باعث ایجاد لکه، رگه، محوشدگی روی بخش بندی سند و وظائف شناسایی می شود. فرورفتگی ها، شکستگی ها و دیگر تغییر شکل ها در شکل نویسه ها که از عواقب آستانه گذاری اشتباه است، دلایل اصلی خراب شدن عملکرد OCR است.

مقالات تحقیقی بسیاری در زمینه آستانه گذاری ارائه شده است. در [۲۴] تحلیلی مقایسه ای از ۵ روش عمومی آستانه گذاری و معیاری بهبود یافته برای ارزیابی این روشها ارائه شده است. نویسندگان [۲۵] هم چندین معیار اندازه گذاری مطرح کردند در [۲۶] سه روش آستانه گذاری در دودویی کردن سند نشان داده شده است.

[۲۷] گسترده ترین تعداد الگوریتم ها یعنی ۱۹ تا از آنها را مورد بررسی قرار داده که همه در مورد جداسازی نویسه از پس زمینه های پیچیده است. [۲۸] هم به ارتباطات و اختلافات عملکردهای میان ۱۱ الگوریتم مینا بر

پایه‌ی مطالعه آماری اشاره کرده است. در ادامه این بخش به فرموله کردن گستره‌ی وسیعی از الگوریتم‌ها تحت یک نگارش واحد می‌پردازیم.

## ۲-۱-۱ طبقه‌بندی‌ها و مقدمات

روش‌های آستانه‌گذاری با توجه به اطلاعات استخراج شده از آنها، در ۶ گروه دسته‌بندی می‌شوند که این گروه‌ها به شرح زیر می‌باشند:

۱. روش‌های مبتنی بر شکل هیستوگرام که در آن‌ها قله‌ها، دره‌ها و انحنای هیستوگرام نرم شده مورد تحلیل قرار می‌گیرند.

۲. روش‌های مبتنی بر خوشه‌بندی که در آنها نمونه‌های سطوح خاکستری در دو دسته پس زمینه و پیش‌نما (شیء) خوشه‌بندی می‌شوند یا به عنوان تلفیقی از دو گوسین مدل می‌شوند.

۳. روش‌های مبتنی بر انتروپی که منجر به الگوریتم‌هایی می‌شود که انتروپی نواحی پس زمینه و پیش‌نما، اختلاف انتروپی میان تصویر اصلی دوسطحی شده و ... را استفاده می‌کنند.

۴. روش‌های مبتنی بر شیء‌گرایی که معیاری از تشابه میان تصاویر خاکستری و دوسطحی شده مانند تشابه شکل فازی، انطباق لبه و ... را دنبال می‌کند.

۵. روش‌های فضایی که از توزیع‌های احتمال مرتبه بالاتر و یا همبستگی پیکسل‌ها استفاده می‌کنند.

۶. روش‌های محلی که مقدار آستانه روی هر پیکسل را به مشخصه‌های محلی تصویر منطبق می‌کنند. در

ادامه ما از نگارشی که توضیح خواهیم داد تبعیت می‌کنیم. هیستوگرام و تابع جرم احتمال (PMF)<sup>۱</sup>

یک تصویر به ترتیب با  $g=0, \dots, G$  و  $h(g)$  و  $p(g)$  نشان داده شده‌اند که  $G$ ، بیشترین مقدار روشنایی در

یک تصویر است که اگر کوانتیزه‌سازی ۸ بیتی باشد همان ۲۵۵ است. اگر گستره‌ی مقادیر خاکستری به

<sup>۱</sup> Probablity Mass Function

وضوح نشان داده نشده مثلاً  $[g_{min} \ g_{max}]$  فرض می‌شود که به بازه  $0$  تا  $G$  گسترده است. تابع احتمال تجمعی به شکل زیر تعریف شده است.

$$P(g) = \sum_{i=0}^g p(i) \quad (۱ - ۲)$$

فرض بر این است که PMF بوسیله نرمالیزه کردن هیستوگرام تصویر به کل تعداد نمونه‌های آن تخمین زده می‌شود. در موضوع پردازش سند، پیش‌نما(شیء) مجموعه‌ای از پیکسل‌ها می‌شود با مقدار روشنایی کمتر از  $T$ ، در حالی که پیکسل‌های پس‌زمینه مقدار روشنایی بیشتر از این آستانه دارند. در مفاهیم بعدی هنگامی که شیء از پس‌زمینه روشن‌تر باشد، مشخصاً مجموعه پیکسل‌هایی با روشنایی کمتر از  $T$  به عنوان پیش‌نما شناخته می‌شوند. PMF‌های پیش‌نما (شیء) و پس‌زمینه به ترتیب  $0 \leq g \leq T$  و  $T+1 \leq g \leq G$ ،  $P_f(g)$  و  $P_b(g)$  هستند که در آنها  $T$  که مقدار آستانه است. احتمالات پس‌زمینه و پیش‌نما به شرح زیر محاسبه می‌شوند.

$$P_f(T) = P_f = \sum_{g=0}^T p(g) \quad (۲ - ۲)$$

$$P_b(T) = P_b = \sum_{g=T+1}^G p(g) \quad (۳ - ۲)$$

انترپیی Shannon که به طور پارامتریک وابسته به مقدار آستانه  $T$  برای پیش‌نما و پس‌زمینه است به صورت زیر فرموله می‌شود:

$$H_f(T) = - \sum_{g=0}^T p_f(g) \log p_f(g) \quad (۴ - ۲)$$

$$H_b(T) = - \sum_{g=T+1}^G p_b(g) \log p_b(g) \quad (۵ - ۲)$$

مجموع این دو عبارت به عنوان  $H(T) = H_f(T) + H_b(T)$  بیان می‌شود. هنگامی که انترپیی روی توزیع تصویر ورودی  $P(g)$  (و نه روی توزیع‌های کلاس) محاسبه می‌شود، بنابراین بطور مشخص به آستانه‌ی  $T$

وابسته نیست و نتیجتاً با  $H$  معرفی می‌شود. برای تعریف‌های دیگر انترپی در مفهوم آستانه‌گذاری با مقداری تغییر در علائم نگارشی ما از نمادهای مشابه  $H_b(T)$ ،  $H_f(T)$  استفاده می‌کنیم.

معیارهای فازی متناسب با وقایع پیش‌نما و پس‌زمینه که همان درجه‌ی سطح خاکستری است که متعلق به پس‌زمینه و شیء می‌باشد به ترتیب بوسیله‌ی  $\mu_b(g)$ ،  $\mu_f(g)$  نمادگذاری می‌شود. میانگین و واریانس پیش‌نما و پس‌زمینه به عنوان تابعی از سطح آستانه  $T$  به شکل زیر تعریف می‌شوند.

$$m_f(T) = \sum_{g=0}^T gp(g) \quad (۲ - ۶)$$

$$\sigma_f^2(T) = \sum_{g=0}^T [g - m_f(T)]^2 p(g) \quad (۲ - ۷)$$

$$m_b(T) = \sum_{g=T+1}^G gp(g) \quad (۲ - ۸)$$

$$\sigma_b^2(T) = \sum_{g=T+1}^G [g - m_b(T)]^2 p(g) \quad (۲ - ۹)$$

در اینجا به یک روش آستانه‌گذاری خاص اشاره شد که در تحلیل شبیه‌سازی برنامه نویسی شده و فرمول آن در جدول نشان داده شده است، مثلاً shape-sezan و cluster-otsu به ترتیب به معنی روش آستانه-گذاری مبتنی بر شکل که به وسیله sezan ارائه شده است و نیز روش آستانه‌گذاری مبتنی بر خوشه بندی پیشنهاد شده توسط otsu می‌باشد.

جدول ۱-۲: توابع آستانه گذاری برای الگوریتم‌های مبتنی بر شکل

|                               |  |
|-------------------------------|--|
| Shape_Rosenfeld <sup>24</sup> | $T_{opt} = \arg \max\{[p(g) - \text{Hull}(g)]\}$ by considering object attributes, such as busyness.   |
| Shape_Sezan <sup>32</sup>     | $T_{opt} = \gamma\{\text{first terminating zero of } \bar{p}(g)\} + (1 - \gamma) \times \{\text{first initiating zero of } \bar{p}(g)\}$<br>$0 \leq \gamma \leq 1$ , $\bar{p}(g) = d/dg[p(g) * \text{smoothing\_kernel}(g)]$ ,<br>where $\gamma = 1$ and kernel size is 55 |
| Shape_Olivio                  | $T_{opt} = \langle T^0, \text{valley at the highest resolution}   T^1, \dots, T^k, \text{given valleys at } k \text{ lower resolutions} \rangle$ , $k = 3$   |
| Shape_Ramesh <sup>30</sup>    | $T_{opt} = \min \left[ \sum_{g=0}^T [b_1(T) - g]^2 + \sum_{g=T+1}^G [b_2(T) - g]^2 \right]$<br>where $b_1(T) = m_f(T)/P(T)$ , $b_2(T) = m_b(T)[1 - P(T)]$  |
| Shape_Guo <sup>28</sup>       | $T_{opt} = \min_g \frac{1}{ 1 - \sum_{i=1}^p a_i \exp(-j2\pi g/256) ^2}$ ,<br>where $\{a_i\}_{i=1}^p$ the $n$ 'th order AR coefficients  |

### ۱-۱-۱-۲ روش‌های آستانه‌گذاری مبتنی بر هیستوگرام

این دسته از روش‌ها آستانه‌گذاری را بر مبنای خواص شکل هیستوگرام انجام می‌دهند که در جدول ۱-۲ به آن‌ها اشاره شده است. خواص شکل به اقسام گوناگونی نقش ایفا می‌کند. فاصله از convex hull هیستوگرام، هیستوگرام نرم شده در نمایش دو قله‌ای با مدل کردن autoregressive از این قبیل هستند. بعضی از الگوریتم‌ها بطور مشخص دنبال قله‌ها و دره‌ها و بطور ضمنی به دنبال قله‌های همپوشانی دار با تحلیل انحنا هستند.

### ۱-۱-۱-۲-۲ آستانه‌گذاری convex hull

در [۲۹] یک روش شناخته شده برای بخش بندی یک تصویر به زیرمجموعه‌های خاکستری انتخاب آستانه‌ها، در انتهای دره‌های هیستوگرام تصویر، بیان شده است. هنگامی که این زیرمجموعه‌ها همپوشانی

می‌کنند، دره‌ها ممکن است وجود نداشته باشند، اما معمولاً این امکان وجود دارد که آستانه‌های خوبی در شانه‌های قله‌های هیستوگرام تعریف نمود.

هم شانه‌ها و هم قله‌ها با تقعرهای روی هیستوگرام مرتبط هستند و بنابراین این امکان وجود دارد که بوسیله‌ی تحلیل ساختار تقعر هیستوگرام آستانه‌های کاندید مناسب پیدا شوند. در مرجع [۲۹] تحلیل تقعر هیستوگرام به عنوان رویکردی در انتخاب آستانه مورد بررسی قرار گرفته است و عملکرد آن روی مجموعه‌ای از هیستوگرام‌های تصاویر مادون قرمز تانک‌ها به نمایش گذاشته شده است.

در مرجع [۳۰] روشهای متعددی برای ایجاد یک هیستوگرام انتقال یافته مطرح شده است که در آنها دره، عمیق‌تر است یا به یک قله تبدیل شده است و بنابراین راحت‌تر قابل شناسایی است.

هیستوگرام‌های منتقل شده<sup>۱</sup> مورد استفاده در این روش‌ها، همه به‌وسیله‌ی ایجاد (gray level, edge value) نمودارهای پراکنده و محاسبه افکنش<sup>۲</sup> های وزن‌گذاری شده این نمودارها روی محور سطح خاکستری بدست می‌آیند. با استفاده از این رویکرد یکپارچه فهم این که این روش‌ها چگونه کار می‌کنند و پیش‌بینی این که کدام روش موثرتر خواهد بود را آسان‌تر می‌سازد.

در [۳۱] مفهوم بدنه‌نمایی<sup>۳</sup>، به عنوان یک تغییر در convex hull بالایی برای یک هیستوگرام تعریف شده است. خواص این مفهوم به ما اجازه می‌دهد که از آن به عنوان معیاری جهت انتخاب تعداد و موقعیت-های آستانه‌ها برای جداسازی سطح خاکستری تصویر، از روی هیستوگرام شدت روشنایی تصویر استفاده کنیم.

---

<sup>۱</sup> transformed  
<sup>۲</sup> projection  
<sup>۳</sup> exponential hull

## ۲-۱-۱-۱-۲ آستانه گذاری قله و دره :

در [۳۲] تحلیل قله را با کانوالو کردن تابع هیستوگرام با یک کرنل نرم و مشتقی انجام می‌دهد. به وسیله تنظیم کردن دهانه نرم کننده کرنل و ذخیره کردن برای ترکیب قله‌ها، هیستوگرام به یک تابع دو لب<sup>۱</sup> تبدیل می‌شود. عملیات مشتق‌گیری، سه خروجی گذر از صفرهای ابتدایی، ماکزیمم و انتهایی را در هر لب هیستوگرام بدست می‌دهد.  $S = [(e_i, m_i, s_i), i = 1, \dots, 2]$  . Sought آستانه باید جایی میان عبور از صفر ابتدایی و انتهایی باشد، یعنی  $0 \leq \gamma \leq 1$  که در این تحقیق نتیجه با  $\gamma = 1$  خوب بوده است.

تغییرات در این تم در [۳۳] آورده شده است، که در آن توزیع تجمعی تصویر ابتدا بر حسب توابع چبی شف گسترده شده‌اند که بوسیله‌ی تحلیل منحنی دنبال می‌شوند. [۳۴] یک هیستوگرام نرم شده با گوسین‌ها بدست آورده و هیستوگرام منتهجه برای قله‌ها و نقاط تیز منحنی مورد بررسی قرار می‌گیرد. دریافتیم که تحلیل منحنی کارآمد است اگر هیستوگرام bimodality خود را بخاطر همپوشانی بیش از حد کلاس‌های هیستوگرام از دست داده باشد. در مقاله‌ای مشابه نویسندگان تحلیل چند مقیاسه PMF را انجام داده‌اند و نتایج‌های آن را تفسیر کردند که همان گذر از صفرها و اکسترمم‌ها در مقیاس‌ها می‌باشند.

در [۳۵] با استفاده از یک تبدیل موجک گسسته می‌توان تحلیل چند وضوحی هیستوگرام را به دست آورد.  $p^s(g) = p(g) * \psi_s(g)$  که در آن  $p'(g) = p(g)$  هیستوگرام اصلی نرمالیزه شده است، آستانه به عنوان دره (نقطه‌ی مینیمم) که بعد از اولین قله در هیستوگرام نرم شده می‌آید تعریف می‌شود. این مکان آستانه متناوباً روی مقیاس‌های مختلف که از وضوح خشن‌تر شروع می‌شوند، پیدا می‌شود.

---

<sup>۱</sup> two-lobe

بنابراین با شروع با نقطه دره  $T^{(k)}$  در  $k$  امین سطح خشن، موقعیت با توجه به اکستریم مربوطه در هیستوگرام‌هایی با وضوح بالاتر  $p^{(k-1)}(g), \dots, p^{(1)}(g)$  رد گم می‌کند<sup>۱</sup> که به وسیله‌ی تصحیح کردن دنباله  $T^{(1)}, \dots, T^{(k)}$  تخمین زده می‌شود. (در این تحقیق  $k=3$  استفاده شده است)

### ۳-۱-۱-۱-۲ آستانه‌گذاری مدل کردن شکل shape modeling :

[۳۶] از یک تخمین تابعی ساده با PMF شامل یک تابع دو پله‌ای استفاده کرده، بنابراین جمع مربعات میان یک تابع دو سطحی و هیستوگرام مینیمم می‌شود. و راه حل برای  $T_{opt}$  بوسیله‌ی جستجوی مکرر بدست می‌آید. [۳۷] ایده‌ی تخمین شکل را تعمیم داده است.

در [۳۸] نویسندگان، طیف را به عنوان طیف سیگنال‌های نمایی چند مختلطه تخمین زده‌اند و این کار را با روش تحلیل طیفی Prony انجام دادند. یک روش با مدل تمام قطب مشابه در [۳۹] بحث شده است. در منبع [۲۳] از یک رویکرد بهبود یافته بهره گرفته و یک مدل autoregressive (AR) استفاده شده تا هیستوگرام نرم شود. در اینجا با تفسیر PMF و قرینه‌ی آن حول  $g=0$ ،  $p(-g)$  به عنوان یک چگالی طیف توان نویزی شروع می‌کنیم. و داریم  $\tilde{p}(g) = p(g)$  برای  $g \geq 0$ ،  $p(-g)$  برای  $g \leq 0$

سپس ضرایب auto correlation را در  $K=0, \dots, G$  به وسیله تبدیل معکوس سری فوریه گرفتن از هیستوگرام اصلی به دست می‌آید. که در آن  $r(k) = IDFT[\tilde{p}(g)]$ . پس ضرایب auto correlation  $\{r(k)\}$  برای حل  $n$  امین ضرایب AR  $\{a_i\}$  استفاده می‌شود. در نتیجه، هیستوگرام نرم می‌شود و به سمت یک نمایش دو قله ای یا bimodal با  $n$  امین درجه مدل AR رهنمون می‌شود.  $(n=1, \dots, b)$  آستانه به عنوان نقطه مینیمم در میان موقعیت دو قطب آن قرار گرفته و حاصل از نرم شده‌ی طیف AR است، قرار دارد.

---

<sup>۱</sup> backtrack

## ۲-۱-۱-۲ روش های آستانه گذاری مبتنی بر خوشه بندی

در این دسته از الگوریتم ها، داده سطح خاکستری یک تحلیل خوشه بندی با تعداد خوشه ها که همیشه روی ۲ تنظیم شده است را متحمل می شود. جدول ۲-۲ به بعضی از این روش ها اشاره می کند. به خاطر این که دو خوشه به دو لب یک هیستوگرام مرتبط هستند، بعضی نویسندگان به دنبال نقاط میانی قله ها می گردند. در بعضی منابع الگوریتم، مبتنی بر منطبق کردن مخلوط گوسین ها می باشد. خوشه بندی mean square و خوشه بندی فازی در سایر مقالات پیشنهاد شده است.

جدول ۲-۲: توابع آستانه گذاری برای الگوریتم های مبتنی بر خوشه بندی

|                                 |   |
|---------------------------------|---|
| Cluster_Riddler <sup>34</sup>   | $T_{opt} = \lim_{n \rightarrow \infty} \frac{m_A(T_n) + m_B(T_n)}{2}$ <p>where</p> $m_A(T_n) = \sum_{g=0}^{T_n} g p(g) \quad m_B(T_n) = \sum_{g=T_n+1}^G g p(g)$  |
| Cluster_Yanni <sup>41</sup>     | $T_{opt} = (g_{max} - g_{min}) \sum_{g=g_{min}}^{g_{mid}} p(g)$   |
| Cluster_Otsu <sup>46</sup>      | $T_{opt} = \arg \max \left[ \frac{P(T)[1-P(T)][m_A(T) - m_B(T)]^2}{P(T)\sigma_A^2(T) + [1-P(T)]\sigma_B^2(T)} \right]$  |
| Cluster_Lloyd <sup>24</sup>     | $T_{opt} = \arg \min \left[ \frac{m_A(T) + m_B(T)}{2} + \frac{\sigma^2}{m_A(T) - m_B(T)} \log \frac{1-P(T)}{P(T)} \right]$ <p><math>\sigma^2</math> is the variance of the whole image.</p>                                       |
| Cluster_Kittler <sup>32</sup>   | $T_{opt} = \arg \min \{ P(T) \log \sigma_A(T) + [1-P(T)] \log \sigma_B(T) - P(T) \log P(T) - [1-P(T)] \log [1-P(T)] \}$ <p>where <math>\{\sigma_A(T), \sigma_B(T)\}</math> are foreground and background standard deviations.</p> |
| Cluster_Jawahar_1 <sup>47</sup> | $T_{opt} = \arg \min_g [\mu_A^-(g) - \mu_B^-(g)]$ <p>where</p> $d(g, m_k) = \sum_{g=0}^T (g - m_k)^2, \quad k = b, f$   |
| Cluster_Jawahar_2 <sup>47</sup> | $T_{opt} = \arg \min_g [\mu_A^-(g) - \mu_B^-(g)]$ <p>where</p> $d(g, m_k) = \frac{1}{2} \left( \frac{g - m_k}{\sigma_k} \right)^2 + \log \sigma_k - \log \beta_k, \quad k = b, f$   |

### ۱-۲-۱-۱-۲ آستانه گذاری مکرر

منبع [۴۰] یکی از اولین طرح های مکرر مبتنی بر مدل های مخلوط گوسی دو کلاسه را بهبود داد. در تکرار  $n$ ، یک آستانه جدید  $T_n$  با استفاده از میانگین گرفتن از متوسط پیش نما و پس زمینه پایه ریزی می شود. در عمل، تکرار ها هنگامی که  $[T_n - T_{n+1}]$  به اندازه کافی کوچک شود خاتمه می یابند.

[۴۱]، اولین نقطه میانی بین دو قله ی فرضی هیستوگرام را به عنوان  $g_{mid} = (g_{max} + g_{min})/2$  انتخاب کرد که در آن  $g_{max}$  بالاترین سطح خاکستری غیر صفر و  $g_{min}$  پایین ترین سطح است. بنابراین محدوده ی مقادیر خاکستری غیر صفر در هیستوگرام می شود. نقطه ی میانی با استفاده از میانگین دو قله ی سمت راستی و سمت چپی به روز می شود که همان  $g_{mid}^* = (g_{peak1} + g_{peak2})/2$  است.

### ۱-۲-۱-۲ آستانه گذاری به روش خوشه بندی

[۴۲]، مینیمم کردن مجموع وزن گذاری شده ی واریانس های بین کلاسی پیکسل های پس زمینه و پیش نما را پیشنهاد می کند تا آستانه ی بهینه پایه ریزی شود. یادآور می شویم که مینیمم گیری از واریانس های کلاس، معادل با ماکزیمم گیری از scatter بین کلاسی است. این روش هنگامی که تعداد پیکسل ها در هر کلاس نزدیک به هم هستند، نتایج رضایت بخشی می دهد. روش Otsu همچنان یکی از پرکاربردترین روش های مورد استفاده در مقالات آستانه گذاری است.

### ۱-۲-۱-۳ آستانه گذاری کمترین خطا

در این روش ها فرض بر این است که یک تصویر می تواند با مخلوطی از توزیع های پیکسل های پس زمینه و پیش نما توصیف شود:

$$p(g) = P(T) \cdot p_f(g) + [1 - P(T)] \cdot p_b(g) \quad (10 - 2)$$

[۴۳] توابع چگالی گوسی هم واریانس را مورد توجه قرار می دهد و با یک جستجوی مکرر خطای اشتباه در طبقه بندی را مینیمم می کند. برعکس آن در [۴۴]، فرض واریانس های برابر را برداشته و در اصل یک مساله انطباق کمترین خطای چگالی گوسی را مطرح می کند. [۴۵] هم پیشرفتی در روش آستانه گذاری را با مشاهدهی متوسط ها و واریانس های تخمین زده شده از توزیع های کوتاه شده ناشی از بایاس ارائه داده است. البته بایاس، تنها هنگامی مورد توجه قرار می گیرد که دو حالت هیستوگرام قابل تشخیص نباشند.

## ۲-۱-۱-۲-۴ آستانه گذاری خوشه بندی فازی

[۴۶]، عضویت های خوشه بندی فازی را به پیکسل ها با توجه به اختلافاتشان نسبت به میانگین دو کلاس اختصاص داده اند. میانگین های خوشه و توابع عضویت به شکل زیر محاسبه می شوند

$$m_k = \frac{\sum_{g=1}^G g.p(g)\mu_k^T(g)}{\sum_{g=1}^G p(g)\mu_k^T(g)} \quad K = f, b \quad (11-2)$$

$$\mu_f^T(g) = \frac{1}{1 + \left(\frac{d(g,m_f)}{d(g,m_b)}\right)^{\frac{1}{\tau-1}}} \quad (12-2)$$

$$\mu_b^T(g) = 1 - \mu_f^T(g) \quad (13-2)$$

در این عبارات  $d(\dots)$  تابع فاصله ی اقلیدسی میان سطح خاکستری  $g$  و میانگین کلاس است، در حالی که  $\tau$  اندیس فازی بودن است. به یاد داشته باشید که در  $\tau=1$  خوشه بندی  $k$ -mean به دست می آید. در منبع [۴۷]، تابع فاصله و تابع عضویت به شکل زیر تعریف شدند.

$$d(g, m_k) = \frac{1}{\tau} \left(\frac{g-m_k}{\sigma_k}\right)^\tau + \log \sigma_k - \log \beta_k \quad (14-2)$$

$$\beta_k = \frac{\sum_{g=1}^G g.p(g)\mu_k^T(g)}{\sum_{g=1}^G g.p(g)(\mu_f^T(g) + \mu_b^T(g))} \quad (15-2)$$

$$\sigma_k^2 = \frac{\sum_{g=0}^G g \cdot p(g) \mu_k^T(g) (g - m_k)^2}{\sum_{g=0}^G p(g) \mu_k^T(g)} \quad (۱۶ - ۲)$$

در هر کدام از این روش ها، آستانه به عنوان یک نقطه متقاطع در توابع عضویت است. یعنی

$$T_{opt} = \arg_g \text{equal} \{ \mu_f^T(g) = \mu_b^T(g) \} \quad (۱۷ - ۲)$$

### ۲-۱-۱-۳ روش های آستانه گذاری مبتنی بر انتروپی

این دسته از الگوریتم ها، انتروپی توزیع سطوح خاکستری را در یک صفحه استخراج می کند. دستهای از آن ها در جدول ۲-۳ آمده است. ماکزیمم کردن انتروپی تصویر آستانه گذاری شده به عنوان نشان دهنده ی ماکزیمم اطلاعات انتقال، تفسیر می شود. بعضی نویسندگان سعی در مینیمم کردن انتروپی میانه بین تصویر سطح خاکستری ورودی و تصویر دودویی خروجی به عنوان محافظت کننده اطلاعات دارند.

جدول ۲-۳: توابع آستانه گذاری برای الگوریتم‌های مبتنی بر انتروپی

|                               |  |
|-------------------------------|--|
| Entropy_Kapur <sup>43</sup>   | $T_{opt} = \arg \max (H_X(T) + H_Y(T)) \text{ with}$ $H_X(T) = - \sum_{g=0}^T \frac{p(g)}{P(T)} \log \frac{p(g)}{P(T)} \text{ and } H_Y(T) = - \sum_{g=T+1}^G \frac{p(g)}{P(T)} \log \frac{p(g)}{P(T)}$  |
| Entropy_Yen <sup>54</sup>     | $T_{opt}(T) = \arg \max (C_X(T) - C_Y(T)) \text{ with}$ $C_X(T) = - \log \left[ \sum_{g=0}^T \left( \frac{p(g)}{P(T)} \right)^2 \right] \text{ and } C_Y(T) = - \log \left[ \sum_{g=T+1}^G \left( \frac{p(g)}{1-P(T)} \right)^2 \right]$   |
| Entropy_Sahoo <sup>55</sup>   | $T_{opt} = T_{(1)} [P_{T_{(1)}} + \frac{1}{2} \cdot w \cdot B_1] + \frac{1}{2} \cdot T_{(2)} \cdot w \cdot B_2 + T_{(3)} \cdot [1 - P_{T_{(2)}} + \frac{1}{2} \cdot w \cdot B_3]$ <p>where <math>P_{T_{(k)}} = \sum_{g=0}^{T_{(k)}} p(g)</math>, <math>k=1,2,3</math>, <math>w = P_{T_{(2)}} - P_{T_{(1)}}</math> and</p> $B_1, B_2, B_3 = \begin{cases} (1,2,1) & \text{if }  T_{(1)} - T_{(2)}  \leq 5 \text{ and }  T_{(2)} - T_{(3)}  \leq 5 \text{ or }  T_{(1)} - T_{(2)}  > 5 \text{ and }  T_{(2)} - T_{(3)}  > 5 \\ (0,1,3) & \text{if }  T_{(1)} - T_{(2)}  \leq 5 \text{ and }  T_{(2)} - T_{(3)}  > 5 \\ (3,1,0) & \text{if }  T_{(1)} - T_{(2)}  > 5 \text{ and }  T_{(2)} - T_{(3)}  \leq 5 \end{cases}$ |
| Entropy_Pun_a <sup>51</sup>   | $T_{opt} = \arg \max \left[ \frac{\alpha \log P(T)}{\log(\max\{p(1), \dots, p(T)\})} + (1-\alpha) \frac{\log(1-P(T))}{\log(\max\{p(T+1), \dots, p(G)\})} \right]$  |
| Entropy_Pun_b <sup>52</sup>   | $T_{opt} = \arg \left\{ \sum_{g=0}^T p(g) = (0.5 +  0.5 - \alpha ) \right\}$ <p>optimizing histogram symmetry by tuning <math>\alpha</math>.</p>   |
| Entropy_LI <sup>56,57</sup>   | $T_{opt} = \arg \min \left[ \sum_{g=0}^T g p(g) \log \frac{g}{m_X(T)} + \sum_{g=T+1}^G g p(g) \log \frac{g}{m_Y(T)} \right]$ <p>where <math>\sum_{g=0}^T g = \sum_{g=0}^T m_X(T)</math> and <math>\sum_{g=T+1}^G g = \sum_{g=T+1}^G m_Y(T)</math>.</p>   |
| Entropy_Brink <sup>58</sup>   | $T_{opt} = \arg \min (H(T)) \text{ where } H(T) \text{ is}$ $\sum_{g=0}^T p(g) \left[ m_X(T) \log \frac{m_X(T)}{g} + g \log \frac{g}{m_X(T)} \right] + \sum_{g=T+1}^G p(g) \left[ m_Y(T) \log \frac{m_Y(T)}{g} - g \log \frac{g}{m_Y(T)} \right]$  |
| Entropy_Pa <sup>59</sup>      | $T_{opt} = \arg \max (H_X(T) + H_Y(T))$ <p>where <math>H_X(T) = \sum_{g=0}^T \left[ p_X(g) \log \frac{p_X(g)}{q_X(g)} - q_X(g) \log \frac{q_X(g)}{p_X(g)} \right]</math></p> $H_Y(T) = \sum_{g=T+1}^G \left[ p_Y(g) \log \frac{p_Y(g)}{q_Y(g)} - q_Y(g) \log \frac{q_Y(g)}{p_Y(g)} \right]$ <p>and <math>q_X(g) = \exp[-m_X(T)] \frac{m_X(T)^g}{g!}</math>, <math>g=0, \dots, T</math>, <math>q_Y(g) = \exp[-m_Y(T)] \frac{m_Y(T)^g}{g!}</math>, <math>g=T+1, \dots, G</math>.</p>   |
| Entropy_Shanbag <sup>60</sup> | $T_{opt} = \arg \min (H_X(T) - H_Y(T)) \text{ where}$ $H_X(T) = - \sum_{g=0}^T \frac{p(g)}{P(T)} \log \mu_X(g), \quad H_Y(T) = - \sum_{g=T+1}^G \frac{p(g)}{1-P(T)} \log \mu_Y(g)$   |
| Entropy_Cheng <sup>61</sup>   | $\max_{a,b,c} \left[ H(A, \mu_A) = - \frac{1}{\log 2} (Q(A_1) \log Q(A_1) - Q(A_2) \log Q(A_2)) \right]$ <p>where <math>\mu_A</math> is Zadeh's membership with parameters <math>a, b, c</math>, and</p> $Q(A_1) = \sum_{\mu_g \in A_1} p(g), \quad Q(A_2) = \sum_{\mu_g \in A_2} p(g)$  |

### ۲-۱-۱-۳-۱ آستانه گذاری انتروپیک

در [۴۸]، پیش زمینه و پس نمای تصویر، به عنوان دو منبع سیگنال متفاوت در نظر گرفته می شوند. بنابراین هنگامی که جمع انتروپی های دو کلاس به ماکزیمم خود می رسد، گفته می شود تصویر، به طور بهینه آستانه گذاری شده است. در نتیجه entropic correlation به صورت زیر تعریف می شود:

$$TC(T) = C_b(T) + C_f(T)$$

$$= -\log \left\{ \sum_{g=0}^T \left[ \frac{p(g)}{P(T)} \right]^{\gamma} \right\} - \log \left\{ \sum_{g=T+1}^G \left[ \frac{p(g)}{P(T)} \right]^{\gamma} \right\} \quad (۱۸ - ۲)$$

و آستانه ای که آن را ماکزیمم می کند به دست می آورد. منبع [۴۹]، انتروپی منابع پیش نما و پس زمینه را برای بعضی پارامترهای  $\rho$  به صورت زیر تعریف می کند.

$$H_f^{\rho} = \frac{1}{1-\rho} \ln \left\{ \sum_{g=0}^T \left[ \frac{p(g)}{P(T)} \right]^{\rho} \right\} \quad (۱۹ - ۲)$$

$$H_b^{\rho} = \frac{1}{1-\rho} \ln \left\{ \sum_{g=T+1}^G \left[ \frac{p(g)}{1-P(T)} \right]^{\rho} \right\} \quad (۲۰ - ۲)$$

سپس آنها سه مقدار متفاوت آستانه  $T_1, T_2, T_3$  را با ماکزیمم کردن انتروپی پس زمینه و پیش نما برای سه محدوده  $\rho < 1$ ،  $\rho = 1$  پیدا می کنند.  $T$  بهینه با مخلوط وزن دار شده آنها پیدا می شود.

در [۵۰]، فرض می شود هیستوگرام سطح خاکستری یک منبع نماد  $G$  است که همه ی نمادها به صورت آماری، مستقل هستند.

$$H(T) = \sum_{g=0}^T p(g) \log[p(g)] - \sum_{g=T+1}^G p(g) \log[p(g)] \quad (۲۱ - ۲)$$

## ۲-۱-۱-۳-۲ آستانه گذاری cross-entropic

[۵۱] و [۵۲] آستانه گذاری را به عنوان مینیمم کردن یک فاصله‌ی تئوریک اطلاعات فرموله کردند. این معیار فاصله‌ی kullback\_leibler است.

$$D(q, p) = \sum q(g) \log \frac{q(g)}{p(g)} \quad (۲ - ۲۲)$$

توزیع های تصویر مشاهده شده  $p(g)$  و تصویر بازسازی شده  $q(g)$  در معیار kullback\_leibler تحت شرطی مینیمم می شود. شرط این است که تصاویر مشاهده شده و بازسازی شده، میانگین روشنایی مشابه در پیش نما و پس زمینه شان داشته باشند. یعنی

$$\sum_{g \leq T} g = \sum_{g \leq T} m_f(T) \quad \text{و} \quad \sum_{g \geq T} g = \sum_{g \geq T} m_b(T) \quad (۲۳ - ۲)$$

[۵۳] پیشنهاد داد که آستانه‌ی انتخاب شده برای مینیمم کردن cross entropy به صورت زیر تعریف می شود.

$$H(T) = \sum_{g=0}^T q(g) \log \frac{q(g)}{p(g)} + \sum_{g=T+1}^G p(g) \log \frac{p(g)}{q(g)} \quad (۲۴ - ۲)$$

Cross entropy به عنوان معیاری از پایداری میان تصاویر اصلی و دوسطحی شده تفسیر می شود. آنها نشان می دهند که این آستانه بهینه می تواند به وسیله‌ی عبارتی بر حسب میانگین های کلاس ها پیدا شود. یک تغییر در این رویکرد cross entropy ، به وسیله‌ی مدل کردن یک PMF postreri از نواحی پس نما و پیش زمینه داده شده است. با استفاده از اصل انتروپی ماکزیمم در منبع [۵۴]، PMF های مرتبط بر حسب میانگین کلاس ها به شکل زیر تعریف می شوند.

$$q_f(g) = \exp[-m_f(T)] \frac{m_f(T)^g}{g!} \quad g = 0, \dots, T \quad (۲۵ - ۲)$$

$$q_b(g) = \exp[-m_b(T)] \frac{m_b(T)^g}{g!} \quad g = T + 1, \dots, G \quad (۲۶ - ۲)$$

### ۲-۱-۱-۳-۳ آستانه گذاری فازی انتروپی

[۵۵] فرض می کند، عضویت های فازی نمایشی از این هستند که یک ارزش خاکستری با چه قدرتی به پیش نما یا پس زمینه متعلق است. در واقع هرچه یک ارزش خاکستری از آستانه‌ی فرض شده دورتر باشد (در ناحیه اش عمیق تر باشد) پتانسیلش برای اختصاص داشتن به یک کلاس خاص، بزرگتر است. بنابراین برای هر پیکسل پیش نما و پس زمینه که سطوح  $i$  زیر یا بالای آستانه داده شده‌ی  $T$  هستند، مقادیر عضویت به وسیله‌ی رابطه‌ی زیر تعریف می شوند:

$$\mu_f(T - i) = 0.5 + \frac{p(T) + \dots + p(T-1-i) + p(T-i)}{2 P(T)} \quad (27 - 2)$$

که معیار متعلق به پیش نما است و نیز به وسیله‌ی رابطه‌ی زیر که متعلق به پس زمینه است.

$$\mu_b(T + i) = 0.5 + \frac{p(T+1) + \dots + p(T-1+i) + p(T+i)}{2 (1-P(T))} \quad (28 - 2)$$

مشخصاً روی ارزش خاکستری مرتبط با آستانه، باید بیشترین خطای تخمینی را داشته باشیم طوری که  $\mu_f(T) = \mu_b(T) = 0.5$ . آستانه بهینه به عنوان  $T$  که مجموع انتروپی های فازی را مینیمم می کند، پیدا می شود.

$$T_{opt} = \arg_T \min\{|H_f(T) - H_b(T)|\} \quad (29 - 2)$$

$$H_f(T) = -\sum_{g=0}^T \frac{p(g)}{P(T)} \log[\mu_f(g)] \quad (30 - 2)$$

$$H_b(T) = -\sum_{g=T+1}^G \frac{p(g)}{1-P(T)} \log[\mu_b(g)] \quad (31 - 2)$$

[۵۶] بر مبنای ماکزیمم گیری انتروپی های رخداد فازی، یعنی زیررخدادهای پیش نمای  $A_f$  و پس زمینه- $A_b$  می باشند. تابع عضویت با استفاده از تابع Zadeh's S در منبع [۵۷] به دست آمده است. احتمال

زیررخداد پیش‌نما  $Q(A_f)$  با جمع کردن احتمالات سطح خاکستری که در زیررخداد  $A_f$  نگاشت می‌یابند، پیدا می‌شود.

$$Q(A_f) = \sum_{\mu(g) \in A_f} p(g) \quad (۲ - ۳۲)$$

$$Q(A_b) = \sum_{\mu(g) \in A_b} p(g) \text{ as } H(A, \mu_A)$$

$$= -\frac{1}{\log 2} [Q(A_f) \log Q(A_f) + Q(A_b) \log Q(A_b)] \quad (۲ - ۳۳)$$

به عبارت دیگر  $Q(A_b)$  و  $Q(A_f)$  به احتمالاتی که در دامنه، برای همه‌ی سطوح خاکستری نگاشت یافته در زیررخداد  $A_i$  مرتبط هستند، می‌باشد. این انتروپی رخداد فازی حول پارامترهای  $(a, b, c)$  از تابع  $s$  ماکزیمم می‌شود. آستانه  $T$  مقدار  $g$  است که بخشی برای  $\mu_A(g) = 0.5$  را ارضا می‌کند.

### ۲-۱-۱-۴ آستانه گذاری مبتنی بر تشابه نسبی

این الگوریتم ها مقدار آستانه را بر مبنای بعضی کمیت های نسبی یا معیارهای تشابه میان تصویر اصلی و تصویر دوسطحی شده را انتخاب می کنند. بعضی از آن ها در جدول ۲-۴ آمده است. این نسبت ها می توانند شکل انطباق لبه، فشردگی شکل، گشتاورهای سطح خاکستری، اتصال، بافت یا پایداری اشیاء مجزا را به خود بگیرند. بعضی از دیگر الگوریتم ها به طور مستقیم، شباهت تصویر اصلی سطح خاکستری را با شباهت تصویر دودویی شده با استفاده از معیار فازی، ارزیابی می کند. یا مثلاً شباهت توزیع های احتمالی تجمعی یا بر حسب کمیت اطلاعات استخراج شده حاصل از جداسازی.

جدول ۲-۴: الگوریتم های آستانه گذاری مبتنی بر تشابه نسبی

|                               |  |
|-------------------------------|--|
| Attribute_Tsai <sup>70</sup>  | $T_{opt} = \arg \text{equal}[m_1 = b_1(T), m_2 = b_2(T), m_1 = b_3(T)]$ <p>where <math>m_k = \sum_{g=0}^G p(g)g^k</math> and <math>b_k = P_f m_f^k + P_b m_b^k</math></p>  |
| Attribute_Hertz <sup>67</sup> | $T_{opt} = \arg \max[E_{\text{gray}} \cap E_{\text{binary}}(T)],$ <p>where <math>E_{\text{gray}}</math>: gray-level edge field, <math>E_{\text{binary}}(T)</math> binary edge field</p>  |
| Attribute_Huang <sup>82</sup> | $T_{opt} = \arg \min \left\{ -\frac{1}{N^2 \log 2} \sum_{g=0}^G [\mu_f(g, T) \log(\mu_f(g, T)) \right. \\ \left. + [1 - \mu_f(g, T)] \log[1 - \mu_f(g, T)] p(g) \right\}$ <p>where <math>\mu_f[l(i, j), T] = \frac{G}{G +  l(i, j) - m_f(T) }</math></p>   |
| Attribute_Pikaz <sup>76</sup> | $T_{opt} = \text{Arg}[\text{most stable point of } s\text{-sized object function } N_s(T)]$  |
| Attribute_Leung <sup>81</sup> | $T_{opt} = \arg \max \{ -p_f \log p_f - (1 - p_f) \log(1 - p_f) \\ + p_f (\pi_{ff} \log \pi_{ff} + \pi_{bf} \log \pi_{bf}) + (1 - p_f) (\pi_{fb} \log \pi_{fb} \\ + \pi_{bb} \log \pi_{bb}) \}$ <p>where <math>\pi_{bf} = \text{Prob}(\text{segmented as background}   \text{belongs to the foreground})</math> etc.</p> |
| Attribute_Pai <sup>68</sup>   | $T_{opt} = \arg \max \{ \text{Compactness}[\mu(T)] \} = \frac{\text{Area}[\mu(T)]}{\text{Perim}[\mu(T)]^2},$ <p>over all foreground regions.</p>   |

### ۲-۱-۱-۴-۱ آستانه گذاری حفظ گشتاور

منبع [۵۸] تصویر سطح خاکستری را به عنوان نسخه مات شده از تصویر دودویی ایده آل فرض می کند. آستانه گذاری طوری پایه ریزی می شود که سه گشتاور سطح خاکستری اول با سه گشتاور اولی تصویر دودویی انطباق می یابند. گشتاورهای سطح خاکستری  $m_k$  و گشتاورهای تصویر دودویی  $b_k$  به شکل زیر تعریف می شوند:

$$m_k = \sum_{g=0}^G p(g)g^k, \quad b_k = P_f m_f^k + P_b m_b^k \quad (۲-۳۴)$$

### ۲-۱-۱-۴-۲ آستانه گذاری انطباق میدان لبه

در [۵۹] یک تکنیک چند آستانه ای را در نظر گرفته اند که در آن یک میدان لبه باریک شده که از تصویر سطح خاکستری  $E_{gray}$  به دست آمده است را با میدان لبه حاصل از تصویر دوسطحی  $E_{binary}(T)$  مقایسه می کند. آستانه ای عمومی، به وسیله ای ارزشی که هم‌رخداد دو میدان لبه‌ی مبتنی بر شمارش لبه‌های منطبق‌شونده را ماکزیمم می کند، تعیین می شود و لبه‌های اصلی اضافی و لبه‌های تصویر آستانه‌گذاری شده را جریمه می کند. هم میدان لبه ی تصویر خاکستری و هم میدان لبه تصویر دودویی به وسیله عمگر  $sobel$  به دست می آیند.

### ۲-۱-۱-۴-۳ آستانه گذاری تشابه فازی

[۶۰] لیستی از Fuzziness ها را با اندازه گیری فاصله‌ی میان تصویر سطح خاکستری و نسخه‌ی دودویی آن پیشنهاد داده است. در این طرح‌ها، یک مجموعه تصویر به عنوان آرایه دابل  $F = \{I(i, j), \mu_f[I(i, j)]\}$  که در آن  $0 \leq \mu_f[I(i, j)] \leq 1$  برقرار است، پیشنهاد شده که برای هر پیکسل در مکان  $(i, j)$  اندازه فازی آن را برای تعلق به پیش نما نشان می دهد. با داشتن ارزش عضویت فازی برای هر پیکسل، لیستی از Fuzziness برای کل تصویر برای آنتروپی Shannon می تواند به دست آید. آستانه بهینه به وسیله مینیمم کردن اندیس

Fuzziness های تعریف شده، برحسب میانگین‌ها، میانه‌ها و توابع عضویتی کلاس (پس زمینه یا پیش نما) پیدا می‌شود. میانگین‌ها  $m_f(T)$  و  $m_b(T)$  و توابع عضویتی  $\mu_f[I(i, j), T]$ ،  $\mu_b[I(i, j), T]$  نشان داده می‌شوند.

[۶۱] معیارهای مختلفی برای انتخاب آستانه را ارزیابی می‌کند که عبارتند از اندیس خطی از Fuzziness‌ها، اندیس ترتیبی از Fuzziness‌ها، معیار لگاریتمی آنتروپی و معیار نمایی آنتروپی که باید گفت اندیس خطی از همه بهتر کار می‌کند.

#### ۲-۱-۱-۴-۴ آستانه گذاری حالت پایدار توپولوژیکی

[۶۲] یادآور شده است که expert ها در میکروسکوپی، سطح آستانه گذاری را در نقطه ای که لبه ها و شکل شیء پایدار می‌شود به طور هدفمند تنظیم می‌کنند. در بعضی منابع هم یک مقدار آستانه را دنبال می‌کنند که هنگامی پایدار می‌شود که اشیاء پیش نما به اندازه‌ی صحیح خود برسند. این کار با تابع اندازه‌ی آستانه  $N_s(T)$  اندازه گیری می‌شود که به عنوان تعداد اشیائی تعریف می‌شوند که حداقل  $s$  تعداد از پیکسل‌ها را دارا می‌باشند. آستانه در وسیع ترین فلات گراف تابع  $N_s(T)$  پایه ریزی می‌شود. به خاطر اینکه اشیاء نویری به سرعت با تغییر آستانه ناپدید می‌شوند، فلات، گستره‌ی آستانه را آزاد می‌کند که از آن اشیاء پیش نما به راحتی از پس زمینه قابل تشخیص هستند. در [۲۳] نقطه‌ی میانی بزرگترین فلات را به عنوان آستانه‌ی بهینه انتخاب کرده‌اند.

#### ۲-۱-۱-۴-۵ آستانه گذاری بیشترین اطلاعات

[۶۳] مساله‌ی آستانه گذاری را به عنوان تغییری در خطای تخمینی یک مشاهده روی مشخصات کلاس‌های پس زمینه و پیش نما تعریف کرده است. نمایش اطلاعات پیش نما/ پس زمینه خطای تخمینی یک پیکسل را کاهش می‌دهد و این بهره اطلاعات با  $H(p) - \alpha H(p_f) - (1 - \alpha)\alpha H(p_b)$  اندازه گیری می‌شود که  $H(p)$  خطای تخمین اولیه پیکسل و  $\alpha$  احتمال پیکسل برای تعلق به کلاس پیش نما است. آستانه‌ی بهینه به عنوان

تولیدکننده یک نگاشت جداسازی، به نوبه خود میانگین باقی مانده خطای تخمینی را در کلاسی که پیکسل متعلق به آن است مینیمم می کند؛ به عبارت دیگر  $\pi_{fb}$  (پیکسلی که به نظر می رسد پیش نماست در حالی که در حقیقت به پس زمینه متعلق است).  $\pi_{ff}$  و  $\pi_{bb}$  معرف شرایط صحیح طبقه بندی هستند. آستانه‌ی بهینه‌ی مرتبط با ماکزیمم، خطای تخمینی را کاهش می دهد که مفهومی این است که تصویر جداسازی شده همان قدر اطلاعات را از نظر کمیت حمل می کند که در اطلاعات اصلی است.

### ۲-۱-۱-۴-۶ آستانه گذاری بهبود یافته فشرده‌گی Fuzzy

[۶۴] مفهوم هندسه فازی را تعمیم داد. برای مثال سطح یک مجموعه فازی به شکل زیر تعریف می شود و محیطش طبق رابطه بعد:

$$\text{Area}(\mu) = \sum_{i,j=1}^N \mu [I(i,j)] \quad (2-35)$$

که در آن سیگما روی هر ناحیه عضویت غیرصفر گرفته شده است و  $N$  تعداد نواحی در یک تصویر جداسازی شده است. [۶۵] نتیجه جداسازی را طوری ارزیابی کردند که هم محیط و هم مساحت توابعی از آستانه  $T$  هستند. آستانه‌ی بهینه برای ماکزیمم کردن فشرده‌گی مجموعه های پیش نمای جدا شده تعیین شده است. در عمل می توان از تابع  $S$  استاندارد برای تخصیص تابع عضویت در پیکسل استفاده کرد.

$I(i,j): \mu[I(i,j)] = S[I(i,j); a, b, c]$  که در منبع [۶۶] با نقطه تقاطع  $b=(a+c)/2$  و پهنای باند  $\Delta b = b - a = c - b$  وجود دارد. آستانه بهینه  $T$  به وسیله جستجو روی جفت های  $(b, \Delta b)$  برای مینیمم کردن فشرده‌گی تصویر پیدا می شود. مشخصا مزیت معیار فشرده‌گی روی اندیس های دیگر Fuzziness این است که هندسه اشیاء یا Fuzziness در فضای مکانی مورد ملاحظه قرار گرفته است.

## ۲-۱-۱-۵ روش های آستانه گذاری فضایی

این دسته از الگوریتم ها که بعضی از آن ها در جدول ۲-۵ آمده اند، نه تنها توزیع هایی با ارزش سطح خاکستری را استفاده می کنند بلکه وابستگی پیکسل ها در یک همسایگی را نیز به کار می برند. برای مثال احتمالات متن، توابع همبستگی، احتمالات هم رخداد، مدل های وابستگی خطی محلی پیکسل ها، آنتروپی دو بعدی و . . . . یکی از اولین نویسندگانی که اطلاعات فضایی را کشف کرد، [۶۷] بود که میانگین محلی سطوح خاکستری برای آستانه گذاری را در نظر گرفت. دیگران هم این کار را با استفاده از relaxation و بهبود نگاشت دودویی، لاپلاسین تصاویر برای بهبود هیستوگرام ها، آستانه گذاری quadtree و محاسبات آماری درجه دوم دنبال کردند. احتمالات هم رخداد هم به عنوان نشان هایی از وابستگی فضایی استفاده شده اند.

جدول ۲-۵: الگوریتم‌های آستانه گذاری برای آستانه گذاری فضایی

|   |  |
|---|--|
| Spatial_Pal_1 and Spatial_Pal_2 <sup>94</sup> | $T_{opt} = \arg \max [H_{bb}(T) + H_{ff}(T)] \quad \text{or} \quad T_{opt} = \arg \max [H_{fb}(T) + H_{bf}(T)]$ <p>where <math>H_{fb}(T)</math>, <math>H_{bf}(T)</math>, <math>H_{ff}(T)</math>, <math>H_{bb}(T)</math> are the co-occurrence entropies</p> $(T_{opt}, \bar{T}_{opt}) = \arg \min \{ \log [P(T, \bar{T}) [1 - P(T, \bar{T})] + H_f / P(T, \bar{T}) + H_b / [1 - P(T, \bar{T})]] \}$ <p>where</p> $H_f = - \sum_{i=1}^T \sum_{j=1}^{\bar{T}} \frac{p(g, \bar{g})}{P(T, \bar{T})} \log \frac{p(g, \bar{g})}{P(T, \bar{T})}$ <p>and</p> |
| Spatial_Abutaleb <sup>97</sup>                | $H_b = - \sum_{i=T+1}^G \sum_{j=T+1}^G \frac{p(g, \bar{g})}{[1 - P(T, \bar{T})]} \log \frac{p(g, \bar{g})}{[1 - P(T, \bar{T})]}$ $T_{opt} = \arg \min \left[ - \sum_{k=0}^{s \times s} p_k^{block}(T) \cdot \log p_k^{block}(T) \right]$   |
| Spatial_Beghdadi <sup>104</sup>               | <p>where <math>p_k^{block}(T)</math> is the probability of <math>s \times s</math> size blocks containing <math>k</math> whites and <math>s^2 - k</math> blacks (<math>s = 2, 4, 8, 16</math>)</p> $T_{opt} = \max_{a,b,c} \{ H_{fuzzy}(foreground) + H_{fuzzy}(background) \}$  |
| Spatial_Cheng <sup>101</sup>                  | <p>where <math>H_{fuzzy}(A) = - \sum_{x,y} \mu_A(x,y) p(x,y) \log p(x,y)</math>, <math>\{a,b,c\}</math>, the S-function parameters; <math>\{A = foreground, background\}</math>; <math>\{x,y\} = \{ \text{pixel value, local average value within } 3 \times 3 \text{ region} \}</math>.</p>   |

## ۲-۱-۱-۵-۱ روش های آستانه گذاری هم رخداد

[۶۸] با درک این موضوع که دو تصویر با هیستوگرام های یکسان می توانند آنتروپی های مرتبه  $n$  متفاوت به خاطر ساختار فضایی شان داشته باشند، فرض کردند احتمال هم رخداد مقادیر خاکستری  $g_1$  و  $g_2$  روی همسایه های افقی و عمودی شان هستند. بنابراین پیکسل هایی که ابتدا با مقدار آستانه  $T$  دوسطحی شده اند به نواحی پیش نما و پس زمینه تقسیم می شوند. هم رخداد سطوح خاکستری  $k$  و  $m$  به شکل زیر قابل محاسبه است:

$$c_{k,m} = \sum_{\text{all pixel } c} \delta, \text{ where } \delta = 1 \text{ if}$$

$$\{[I(i, j) = k] \wedge [I(i, j + 1) = m] \vee [I(i, j) = k] \wedge [I(i + 1, j) = m]\} \quad (۲ - ۳۶)$$

Pal دو روش برای استفاده از احتمالات هم رخداد پیشنهاد داده است. در اولین عبارت تصویر دوسطحی شده باید تا حد ممکن گذرهای پیش نما به پس زمینه و بالعکس داشته باشد. در رویکرد دوم، برعکس آن درست است که در آن احتمال پیکسل های همسایگی که در همان کلاس باقی می ماند تقویت می شود.

[۶۹] آستانه را طوری پایه ریزی کرد که احتمالات هم رخداد تصویر اصلی و تصویر دودویی به طور کمینه واگرا هستند. به عنوان معیاری از تشابه، واگرایی (دیورژانس) مستقیم فاصله ی Kullback-Leibler استفاده می شود. همانطور که در شکل می بینید، ماتریس هم رخداد ۴ بخشی است. بخش اول گذرهای پس زمینه-پس زمینه است (bb)، بخش سوم گذرهای پیش نما به پیش نماست (ff)، به طور مشابه بخش های دوم و چهارم گذرهای پس زمینه به پیش نما و پیش نما به پس زمینه را نشان می دهند. با استفاده از احتمالات هم رخداد  $P_{ij}$  (که امتیاز گذرهای سطوح خاکستری  $i$  به  $j$  نرمالیزه شده به وسیله کل تعداد گذرهاست) احتمالات چهارتایی به شکل زیر به دست می آیند:

$$P_{bb}(T) = \sum_{i=0}^T \sum_{j=0}^T p_{ij}, \quad P_{bf}(T) = \sum_{i=0}^T \sum_{j=T+1}^G p_{ij}, \quad (۲ - ۳۷)$$

$$P_{fb}(T) = \sum_{i=T+1}^G \sum_{j=0}^T p_{ij}, \quad P_{ff}(T) = \sum_{i=T+1}^G \sum_{j=T+1}^G p_{ij}, \quad (۲ - ۳۸)$$

برای تصویر آستانه گذاری شده، کمیت ها به شکل زیر هستند:

$$Q_{bb}(T) = \sum_{i=0}^T \sum_{j=0}^T q_{ij}, \quad Q_{bf}(T) = \sum_{i=0}^T \sum_{j=T+1}^G q_{ij}, \quad (۲ - ۳۹)$$

$$Q_{fb}(T) = \sum_{i=T+1}^G \sum_{j=0}^T q_{ij}, \quad Q_{ff}(T) = \sum_{i=T+1}^G \sum_{j=T+1}^G q_{ij}, \quad (۲ - ۴۰)$$

$$T_{opt} = \operatorname{argmin}[P_{bb}(T)Q_{bb}(T) + P_{bf}(T)\log Q_{bf}(T) + P_{ff}(T)\log Q_{ff}(T) + P_{fb}(T)\log Q_{fb}(T)] \quad (۲ - ۴۱)$$

## ۲-۱-۱-۵-۲ آستانه گذاری آنتروپی درجه بالاتر

[۷۰] آنتروپی اتصال دو متغیر تصادفی مرتبط یعنی مقدار خاکستری تصویر  $g$  در یک پیکسل و میانگین مقدار خاکستری  $\bar{g}$  در یک همسایگی با مرکزیت آن پیکسل را در نظر گرفته است. با استفاده از هیستوگرام دو بعدی  $p(g, \bar{g})$  برای هر جفت آستانه  $(T, \bar{T})$  می توان توزیع تجمعی  $p(T, \bar{T})$  را محاسبه کرد و سپس آنتروپی پیش نما را به صورت زیر تعریف نمود:

$$H_f = - \sum_{i=1}^T \sum_{j=1}^{\bar{T}} \frac{p(g, \bar{g})}{p(T, \bar{T})} \log \frac{p(g, \bar{g})}{p(T, \bar{T})} \quad (۲ - ۴۲)$$

به طور مشابه می توان آنتروپی درجه دوم ناحیه ی پس زمینه را تعریف کرد. تحت این فرض که عبارات Off diagonal به نام های  $[(\cdot, T), (\bar{T}, G)]$  و  $[(T, G), (\cdot, \bar{T})]$  قابل چشم پوشی هستند و شامل عناصری هستند که تنها به دلیل لبه های تصویر و نویز وجود دارند، جفت بهینه ی  $(T, \bar{T})$  می تواند به عنوان مقدار مینیمم کننده ی تابع دو بعدی آنتروپی پیدا شود.

از سوی دیگر [۷۱] با استفاده از آنتروپی پیکربندی بلوک به عنوان یک منبع نمادین، همبستگی فضایی پیکسل ها را استخراج می کند. برای هر مقدار آستانه  $T$ ، تصویر یک  $N = 2^{S^T}$  مجموعه از بلوک های دودویی کنار هم با سایز  $S \times S$  پیکسل می شود.  $B_k$  زیر مجموعه ای از  $S \times S$  بلوک شامل  $k$  سفید و  $K-k$  سیاه را نشان می دهد. جمعیت نسبی آنها احتمالات منبع دودویی می شود که با  $p_k^{\text{block}}$  نشان می دهیم. در اینجا  $p_k^{\text{block}}$  احتمال بلوک شامل  $k$  سفید بی توجه به پیکربندی های دودویی پیکسل می باشد. یک آستانه ی سطح خاکستری بهینه، به وسیله ی ماکزیمم کردن تابع آنتروپی احتمالات بلوک به دست می آید. همان طور که سایز بلوک بزرگتر می شود، تعداد پیکر بندی ها به سرعت افزایش می یابد و در طرف دیگر، بلوک های کوچک ممکن است برای توصیف محتوای هندسی تصویر کافی نباشد. بهترین اندازه ی بلوک به وسیله ی جستجو حول بلوک های با سایز  $2 \times 2$ ،  $4 \times 4$ ،  $8 \times 8$  و  $16 \times 16$  به دست می آید.

## ۲-۱-۱-۶ آستانه گذاری افقی محلی

در این دسته از الگوریتم ها، یک آستانه در هر پیکسل محاسبه می شود که به بعضی از اطلاعات آماری مانند گستره، واریانس یا پارامترهای انطباق سطح همسایگی پیکسل بستگی دارد. در ادامه آستانه  $T(i,j)$  به عنوان تابعی از مختصات های  $(i,j)$  در هر پیکسل نشان داده شده است. یا اگر این ممکن نیست، تصمیمات شیء/پس زمینه بودن به وسیله ی متغیر منطقی  $B(i,j)$  نشان داده می شوند. روش های متعددی در این دسته از الگوریتم ها پیشنهاد شده است. جدول ۲-۶ خلاصه ای از این روش ها را نشان می دهد.

جدول ۲-۶: الگوریتم های آستانه گذاری برای آستانه گذاری وقتی محلی

|                               |  |
|-------------------------------|--|
| Local_Niblack <sup>110</sup>  | $T(i,j) = m(i,j) + k \cdot \sigma(i,j)$ where $k = -0.2$ and local window size is $b = 15$   |
| Local_Sauvola <sup>111</sup>  | $T(i,j) = m(i,j) + \left[ 1 + k \cdot \left[ \frac{\sigma(i,j)}{R} - 1 \right] \right]$ where $k = 0.5$ and $R = 128$  |
| Local_White <sup>112</sup>    | $B(i,j) = \begin{cases} 1 & \text{if } m_{w \times w}(i,j) < l(i,j) * \text{bias} \\ 0 & \text{otherwise} \end{cases}$ where $m_{w \times w}(i,j)$ is the local mean over a $w = 15$ -sized window and $\text{bias} = 2$ .   |
| Local_Bernsen <sup>113</sup>  | $T(i,j) = 0.5 \{ \max_w [l(i+m,j+n)] + \min_w [l(i+m,j+n)] \}$ where $w = 31$ , provided contrast $C(i,j) = l_{\text{high}}(i,j) - l_{\text{low}}(i,j) \geq 15$ .  |
| Local_Palumbo <sup>21</sup>   | $B(i,j) = 1 \text{ if } l(i,j) \leq T_1 \text{ or } m_{\text{neigh}} T_3 + T_5 > m_{\text{center}} T_4$ where $T_1 = 20$ , $T_2 = 20$ , $T_3 = 0.85$ , $T_4 = 1.0$ , $T_5 = 0$ , neighborhood size is $3 \times 3$ .   |
| Local_Yanowitz <sup>115</sup> | $\lim_{n \rightarrow \infty} T_n(i,j) = T_{n-1}(i,j) + R_n(i,j)/4$ where $R_n(i,j)$ is the thinned Laplacian of the image.   |
| Local_Kamel <sup>1</sup>      | $B(i,j) = 1 \text{ if } \{ [L(i+b,j) \wedge L(i-b,j)] \vee [L(i,j+b) \wedge L(i,j-b)] \} \{ [L(i+b,j+b) \wedge L(i-b,j-b)] \vee [L(i+b,j-b) \wedge L(i-b,j+b)] \}$ where $L(i,j) = \begin{cases} 1 & \text{if } [m_{w \times w}(i,j) - l(i,j)] \geq T_0 \\ 0 & \text{otherwise} \end{cases}, w = 17, T_0 = 40$ Define the optimal threshold value ( $T_{\text{opt}}$ ) by using a global thresholding method, such as the Kapur <sup>53</sup> method, then locally fine tune the pixels between $[T_0 - T_1]$ considering local covariance ( $T_0 < T_{\text{opt}} < T_1$ ). |
| Local_Oh <sup>13</sup>        | $B(i,j) = 1 \text{ if } m_{w \times w}(i,j) < T_3 \text{ or } \sigma_{w \times w}(i,j) > T_4$ where $w = 3$ , $T_1 = 50$ , $b = 16$ , $T_2 = 16$ , $T_3 = 128$ , $T_4 = 35$  |
| Local_Yasuda <sup>114</sup>   |  |

### ۲-۱-۱-۶ روش های واریانس محلی:

روش [۷۲] آستانه را با توجه به میانگین محلی  $m(i,j)$  و انحراف معیار  $B(i,j)$  تطبیق می دهد و یک پنجره با سایز  $6 \times 6$  را محاسبه می کند. در [۳] پنجره ای با سایز  $b = 15$  و تنظیمات با پالس  $K = 0.2$  قابل قبول و نتیجه رضایت بخش بود. در [۷۳] روشی ارائه شده که بهبود یافته روش Niblack می باشد مخصوصاً برای سندهای لکه دار و با روشنایی پایین، چون منطبق با مشارکت انحراف معیار است، آستانه کم تر می شود.

## ۲-۱-۱-۶-۲ روش های کنتراست محلی:

[۷۴] مقدار ارزش خاکستری پیکسل را با میانگین ارزش های خاکستری همسایگی حول آن پیکسل مقایسه می کند. (مثلاً یک پنجره  $10 \times 10$ ). اگر پیکسل به طور مشخصی از میانگین تیره تر است به عنوان نویسه و در غیر این صورت به عنوان پس زمینه نمایش داده می شود. در [۷۵] آستانه در مقدار وسط گستره تنظیم می شود که میانگین می نیمم  $I_{low}(i,j)$ ، و ماکزیمم  $I_{high}(i,j)$  ارزش های خاکستری در پنجره محلی پیشنهاد شده با سایز  $w=31$  است. با این وجود اگر کنتراست  $C(i,j)$  از یک آستانه مشخص کمتر باشد گفته می شود آن همسایگی تنها شامل یک کلاس است، به عبارتی پس زمینه یا پیش نما بودن به مقدار  $T(i,j)$  وابسته است.

در [۷۶] گستره دینامیکی تصویر، گسترده می شود سپس نرم کنندگی غیر خطی در آن اتفاق می افتد که در آن لبه های تیز حفظ می شود. نرم کنندگی شامل جایگزین کردن هر پیکسل با میانگین ۸ همسایه اش، گستره پیکسل محلی را ایجاد می کند که از یک آستانه  $T$  پایین تر است (به عنوان گستره ای میان ماکزیمم و مینیمم های محلی تعریف می شود). یک آستانه وقتی به کار گرفته می شود که در آن هر مقدار ارزش پیکسل به پس زمینه مرتبط است اگر رنج محلی از یک آستانه  $T_2$  پایین تر باشد یا مقدار پیکسل بالاتر از میانگین محلی باشد که هر دو روی پنجره های  $w \times w$  محاسبه می شوند. در غیر این صورت گستره دینامیکی گسترده می شود. نهایتاً تصویر با بیان این که یک پیکسل متعلق به یک شیء است اگر مینیمم آن روی پنجره  $3 \times 3$  کمتر از  $T_2$  است یا واریانس محلی آن بالای  $T_3$  است، دوسطحی می شود.

## ۲-۱-۱-۶-۳ طرح های مرکز - محیط

ایده ی منبع [۷۷] مقایسه ی میانگین مقادیر خاکستری در بلوک های متناسب با عرض شیء با محیط های اطرافشان است. اگر  $b$  عرض stroke تخمین زده شده باشد میانگین ها روی یک پنجره  $w \times w$  محاسبه می شوند که  $w=2b+1$

این رویکرد که استفاده از عملگر مقایسه کننده  $L(i,j)$  است شبیه مشتق های مستقیم نرم شده است. تنظیمات پارامتری که در ادامه می آیند مناسب به نظر می رسند.  $T=40$  و  $b=8$  (مقدار مقایسه است).

#### ۲-۱-۱-۶-۴ آستانه گذاری منطبق بر سطح

در [۷۸] لبه و اطلاعات سطح خاکستری با هم ترکیب شده اند تا یک سطح آستانه بسازند. ابتدا اندازه ی گرادیان تصویر نازک می شود تا ماکزیمم گرادیان محلی به دست آید. سطح آستانه با درون یابی با توابع پتانسیل سطح با استفاده از یک روش موفق overrelaxation ساخته می شود. آستانه مداوماً با استفاده از لاپلاسیان گسسته ی سطح به دست می آید.

[۷۹] از یک شبکه عصبی Hopfield برای یک نمونه سطح فعال استفاده کرده است.

#### ۲-۱-۱-۶-۵ روش کریگینگ

روش Oh یک الگوریتم دو بخشه است. در بخش اول با استفاده از یک روش غیر محلی مانند [۴۸] اکثریت جمعیت پیکسل ها به یکی از دو کلاس (پس زمینه با شیء) نسبت پیدا می کنند. با استفاده از واریانس تکنیک Kapur یک آستانه پایین تر  $T_0$  را در نظر می گیریم که مقادیر خاکستری زیر این عدد مطمئناً به کلاس ۱ مثلاً یک شیء، اختصاص دارند. یک آستانه ی بزرگتر  $T_1$  پیدا می شود، طوری که هر پیکسل با ارزش های خاکستری  $T_1 < 9$  به کلاس ۲، یعنی پس زمینه متعلق باشد. پیکسل های باقی مانده ی نامعین با مقادیر خاکستری  $T_1 < 9 < T_0$  به بخش دوم واگذار می شوند. در بخش دوم که مرحله ی نشانگر kriging نام دارد، این پیکسل ها با استفاده از کواریانس محلی نشانگرهای کلاس و روش رگرسیون خطی محدود شده در ناحیه ای از شعاع  $r=3$  (۱۸ پیکسل) به کلاس ۱ یا ۲ اختصاص می یابند.

## ۲-۲ شناسایی خطوط متن

### مقدمه

بخش بندی تصویر سند و تبدیل آن به خطوط متن مرحله‌ی بسیار مهمی در شناسایی سند توسط سیستم OCR است. زوایای چرخش مختلف در همه خطوط زمینه یا در طول یک خط، وجود هم پوشانی میان خطوط یا برخورد دو خط با هم، اندازه های متفاوت نویسه‌ها و یکسان نبودن قالب های نوشته از عواملی هستند که به عنوان چالش در زمینه استخراج خط زمینه به شمار می‌روند. استخراج خط زمینه به طور ایده آل باید روی تصاویر متنی صورت گیرد که بدون نویز و اعوجاج باشند. به همین منظور باید تا حد امکان با روش پیش پردازش، متن را برای ورود به مرحله‌ی استخراج خط آماده کرد. حذف انواع نویزها با روش های مختلف حذف نویز، حذف عناصر غیر متن تصویر و در آخر هم دوسطحی سازی تصویر به روش های عمومی یا محلی از جمله اقداماتی هستند که پیش از تشخیص خط متن باید صورت گیرد. [۸۰] روش‌های استخراج خط را به طور کل می‌توان به چند دسته تقسیم کرد که در ادامه با آنها آشنا می‌شوید.

### ۲-۲-۱ روش های مبتنی بر افکنش (projection based methods)

نمودار های افکنش<sup>۱</sup> عموماً برای جداسازی تصاویر چاپ شده متن به کار می‌روند. نمودار افکنش افقی به وسیله جمع کردن مقادیر پیکسل‌ها در طول محور افقی به ازای هر مقدار  $y$  به دست می‌آید. از روی نمودار افکنش افقی خلاهای میان خطوط متن در جهت افقی مشاهده می‌شوند. شکل ۲-۱ این موضوع را نشان می‌دهد.

$$\text{profile}(y) = \sum_{1 \leq x \leq M} f(x, y) \quad (۲ - ۴۳)$$

<sup>۱</sup> projection profiles

سایر روش‌های به دست آوردن منحنی افکنش شامل گذرهای سیاه به سفید یا تعداد مولفه‌های متصل به هم<sup>۱</sup>، و نه پیکسل‌ها می‌باشد. منحنی افکنش می‌تواند نرم شود مثلاً با یک فیلتر میان‌ه یا گوسی، تا ماکزیمم محلی حذف شود. سپس منحنی افکنش، آنالیز شده تا ماکزیمم و مینیمم آن پیدا شود. دو مانع برای این کار وجود دارد:

خطوط کوتاه پیک‌های پایین تولید می‌کنند و خطوط بسیار باریک و نیز خطوطی که اجزای هم پوشانی‌کننده در خود دارند، پیک‌های قابل توجهی تولید نمی‌کنند. در صورت وجود چرخش<sup>۲</sup> یا نوسانات ملایم<sup>۳</sup> روی خطوط، تصویر به نوارهای عمودی تقسیم می‌شود و افکنش‌ها در هر نوار بررسی می‌شوند. [۸۱] این افکنش‌های تکه‌ای، وسیله‌ای برای تعدیل نوسانات محلی در یک طرح عمومی تر هستند.

در منبع [۸۲] جهت عمومی صفحه‌ی دست نوشته، به وسیله‌ی به کار بردن تبدیل هاف<sup>۴</sup> روی کل تصویر به دست آمده است. هنگامی که این زاویه چرخش پیدا شد، افکنش‌ها حول این زاویه به دست می‌آیند و تعداد ماکزیمم-های نمودار، تعداد خطوط را می‌دهد. از ماکزیمم‌های کوچک در مقابل بزرگترین ماکزیمم صرف نظر می‌شود.

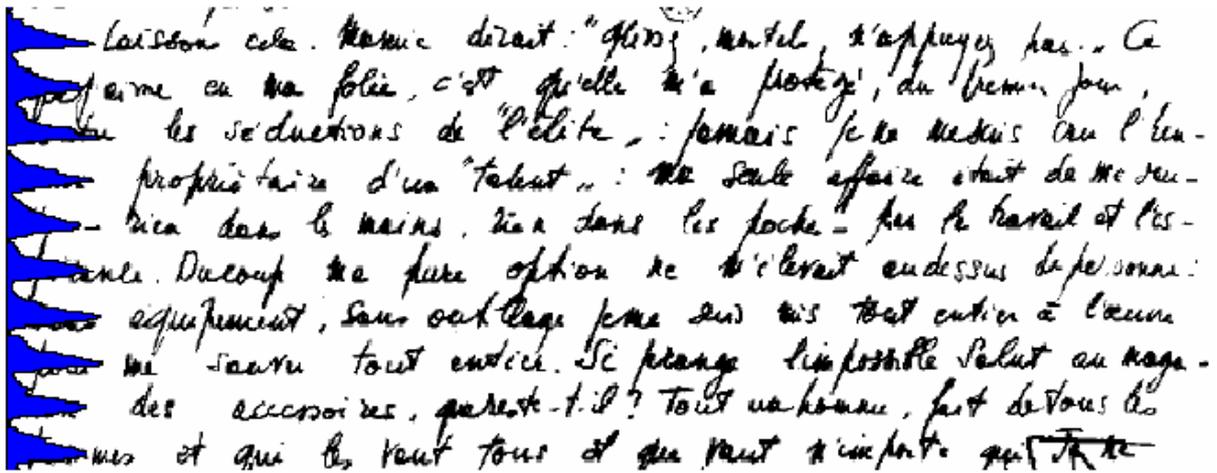
در منبع [۸۳] هر مینیمم در منحنی افکنش، یک نقطه‌ی پتانسیل جداسازی است. سپس نقاط پتانسیل، با توجه به فاصله‌شان به نقاط جداساز مجاور، امتیازبندی می‌شوند. فاصله‌ی مرجع، از هیستوگرام فاصله‌های میان نقاط پتانسیل جداسازی مجاور به دست می‌آید. نقطه‌ی پتانسیل جداسازی با بیشترین امتیاز به عنوان یک لنگر برای به دست آوردن سایر نقاط استفاده می‌شود.

روش قطع  $R_{xy}$  که در مقاله [۸۴] پیشنهاد شده، از افکنش‌های متناوب در طول محور  $X$  و  $Y$  استفاده می‌کند. این کار یک ساختار درختی سلسله‌مراتبی را نتیجه می‌دهد. قطع‌ها در فضاهای سفید پیدا می‌شوند. آستانه‌ها

---

<sup>۱</sup> connected components  
<sup>۲</sup> skew  
<sup>۳</sup> fluctuation  
<sup>۴</sup> hough transform  
<sup>۵</sup>  $R_{xy}$  cut

برای به دست آوردن فاصله های بین خطی یا بین بلوکی لازم هستند. این روش برای متون چاپی یا دست نوشته هایی با خطوط جدا شده از هم قابل استفاده است.



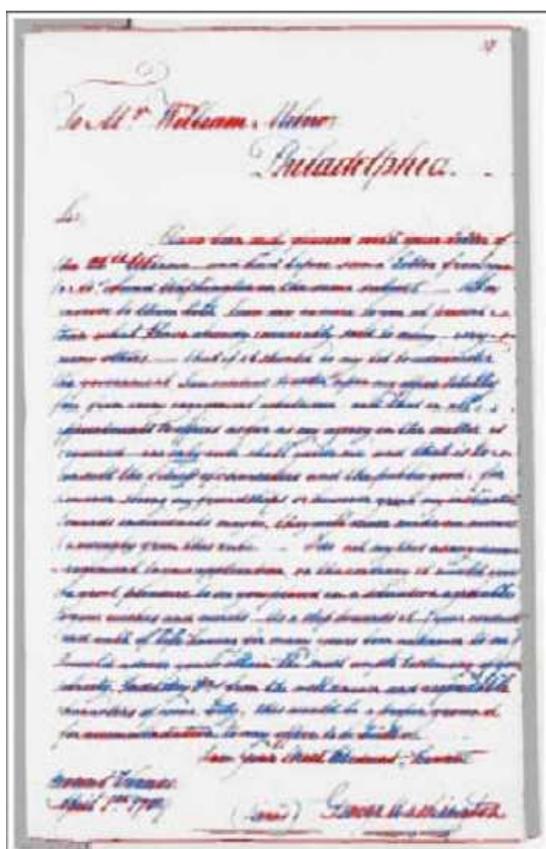
شکل ۱-۲: نمودار افکنش افقی روی یک دست خط [۸۰]

## ۲-۲-۲ روش های آغشته سازی

برای متون چاپ شده و دوسطحی شده، روش های آغشته سازی مانند run length smoothing [۸۵]، می تواند استفاده شود. پیکسل های سیاه متوالی در طول مسیر افقی آغشته می شوند. یعنی فضای سفید بین آن ها با پیکسل های سیاه پر می شود. اگر فاصله ی آن ها در یک محدوده ی از پیش تعیین شده باشد، جعبه های محدود کننده ی مولفه های متصل به هم در تصویر آغشته شده، خطوط متن را محصور می کنند.

الگوریتم دیگری برای این روش که با تصاویر سطح خاکستری سازگار است، برای کتاب های چاپ شده از قرن ۱۶ به کار برده شد، که جمع کردن گرادیان تصویر در طول مسیر افقی است. این روش با متون چاپی قدیمی سازگار است. الگو های خط متن در [۸۶] به وسیله ی ساخت یک ماتریس فازی run length پیدا می شوند. در هر پیکسل run length فازی، حد بیشینه ی پس زمینه، در طول مسیر افقی است. بعضی از پیکسل های پیش-نما ممکن است نادیده گرفته شوند و این در صورتی است که تعداد آنها از یک مقدار از پیش تعیین شده تجاوز

نکند. این ماتریس آستانه ای است برای ساختن تکه های خطوط زمينه ای که بدون قله ها و دره ها ظاهر می شوند. پارامترها باید به دقت و به طور پویا تنظیم شوند. شکل ۲-۲ نمونه ای از الگوهای خطوط متن در یک نامه را استخراج کرده است.



شکل ۲-۲: نمونه ای از الگوهای خطوط متن در یک نامه [۸۰]

### ۲-۲-۳ روش های گروهی کردن

این روش ها که به صورت پایین به بالا<sup>۱</sup> هستند بر مبنای ایجاد گروه ها از روی واحدهای کوچک مثل پیکسل ها هستند. این واحدها ممکن است پیکسل هایی از سطوح بالاتر باشند، مثل مولفه های متصل بهم بلوک ها یا دیگر ویژگی ها مانند نقاط برجسته. سپس واحدها به هم متصل می شوند تا صف بندی ها<sup>۲</sup> را تشکیل دهند. طرح اتصال

<sup>۱</sup> Bottom-up alignments

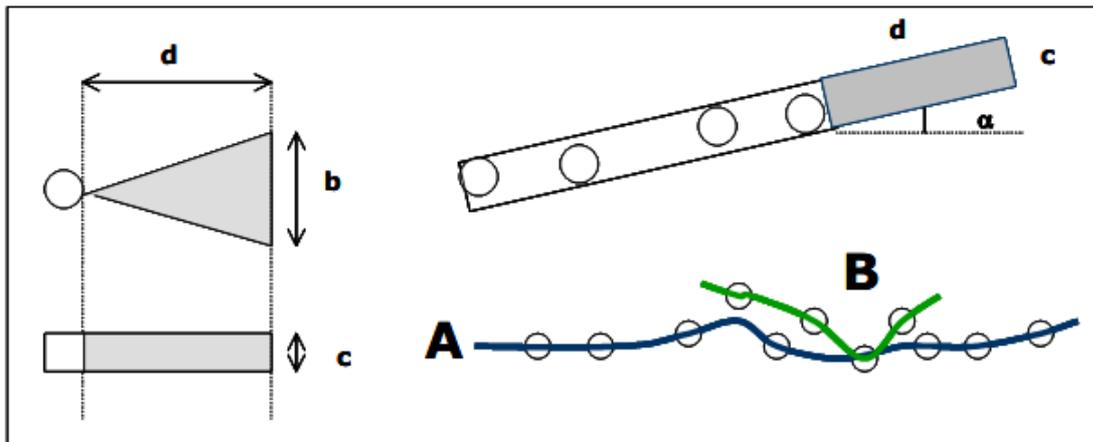
واحدها، هم وابسته به معیار محلی و هم معیار عمومی است، که برای چک کردن ثبات عمومی و محلی به کار می‌رود .

بر خلاف متون چاپ شده، یک طرح ساده اتصال نزدیک ترین همسایه، معمولاً در گروهی کردن واحدهای پیچیده‌ی متن دست نویس شکست می‌خورد. معیار اتصال استفاده شده در روش های زیر به نوع واحدها و مشخصه های متون تحت مطالعه سازگار است. اما هر روشی باید با موارد زیر روبرو شود:

-شروع کردن صف بندی ها: یک یا چند seed برای هر صف بندی

-تعریف همسایگی یک واحد برای به دست آوردن واحد بعدی که معمولاً یک ناحیه ی مستطیلی یا زاویه دار است. در شکل ۲-۳ این همسایه ها نشان داده شده اند.

-حل ناسازگاریها: به خاطر این که یک واحد ممکن است متعلق به چند صف بندی تحت ساخت باشد، باید یک انتخاب انجام شود. حذف یک صف بندی، نگه داشتن هر دو صف بندی، یا بریدن واحد به چند بخش. بنابراین این روش ها شامل یک یا چند معیار کیفیت هستند که تضمین می کنند خط متن، دقت بازسازی کیفیت خوبی دارد. هنگام مقایسه‌ی معیارهای کیفیت دو صف بندی متداخل، صف بندی با کیفیت پایین تر می تواند حذف شود. همچنین در حین پردازش گروهی کردن می توان بین واحدهای مختلفی که با یک همسایگی جمع می‌شوند، انتخاب کرد. این کار به وسیله‌ی ارزیابی کیفیت هر کدام از صف بندی‌های به اصطلاح شکل گرفته انجام می‌شود.



شکل ۲-۳: همسایه‌های زاویه دار و مستطیلی از نقطه و واحدهای مستطیلی. همسایه‌ها با خوشه‌ای از واحدها تعریف شده‌اند.

صف بندی‌های متضاد A, B. یک معیار کیفیت A را انتخاب و B را حذف می‌کند. [۸۰]

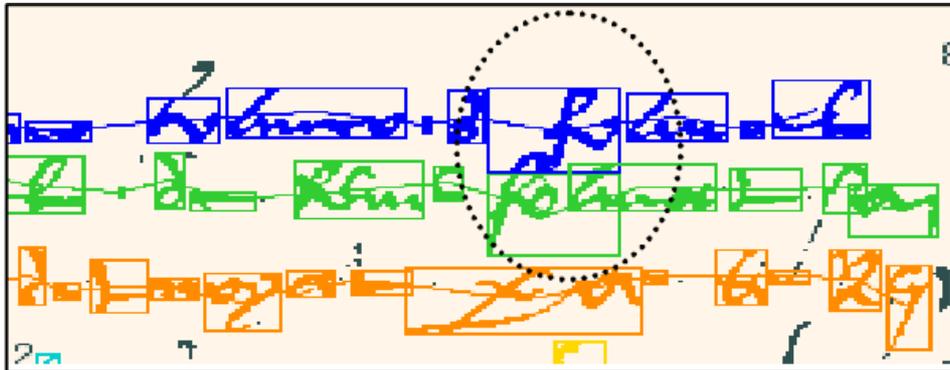
معیارهای کیفیت عموماً شامل استحکام صف بندی است، یعنی تعداد واحدهایی را که شامل می‌شود. دیگر عناصر کیفیت می‌تواند مربوط به اندازه‌ی مولفه و فاصله‌گذاری مولفه یا یک معیار استحکام صف بندی باشد.

منبع [۸۷]، یک روش مکرر<sup>۱</sup>، بر پایه‌ی گروهی کردن ادراکی برای شکل دادن صف بندی‌ها، پیشنهاد داده که بر روی دست نوشته‌ها، چک نویس‌های نویسنده و سندهای تاریخی اعمال شده است. لنگرها به وسیله‌ی انتخاب مولفه‌های متصل در طول جهت‌های ( $۰^{\circ}$  و  $۴۵^{\circ}$  و  $۹۰^{\circ}$  و  $۱۲۵^{\circ}$ ) شناسایی می‌شوند.

هر کدام از این لنگرها یک seed برای هر صف بندی می‌شوند، ابتدا هر لنگر و سپس هر صف بندی به سمت چپ و راست گسترده می‌شوند. این گسترش از سه معیار برای مولفه‌های گروهی‌شده استفاده می‌کند: مجاورت، شباهت، و پیوسته بودن جهت (مسیر). آستانه مکرراً افزایش می‌یابد تا مؤلفه‌ها را در یک همسایگی عریض تر گروه کند تا زمانی که تغییری رخ ندهد. در میان هر تکرار، کیفیت صف بندی به وسیله‌ی معیار کیفیت چک می‌شود که به صف بندی‌های طولانی امتیاز بالاتر و لنگرهایی در همان جهت می‌دهد. هنگامی که یک صف بندی، لنگرهایی صف بندی‌های دیگر را در بر گیرد، یک امتیاز منفی داده می‌شود. دو صف بندی

<sup>۱</sup> iterative

ممکن است همدیگر را قطع کنند یا با هم همپوشانی داشته باشند. مجموعه ای از این قوانین به کار گرفته می-شوند تا این کشمکش‌ها را حل کنند. شکل ۲-۴ این موضوع را نشان می دهد.



شکل ۲-۴: مولفه های برخورد کننده در یک متن [۸۰]

درمنبعی دیگر، خطوط زمینه‌ی بدنه در تصاویر، جستجو شدند. این سند ها شامل بسیاری نوسانات و خطوط هم‌پوشان هستند. واحدهای خطوط زمينه، نقاط مینیمم نوشته هستند، که در اینجا از اسکلت به دست آمده است. اولین بخش های خط زمينه (BLS) تشکیل می شوند، و هر نقطه مینیمم را به همسایگانش اتصال می دهد. این همسایگی به وسیله‌ی یک ناحیه زاویه دار ( $\pm 20^\circ$ ) برای اولین واحد گروهی شده تعریف می شود و این کار، سپس به وسیله‌ی یک ناحیه‌ی مستطیلی محصور کننده‌ی نقاطی که در آخر برای باقیمانده ها متصل می شوند، انجام می شود.

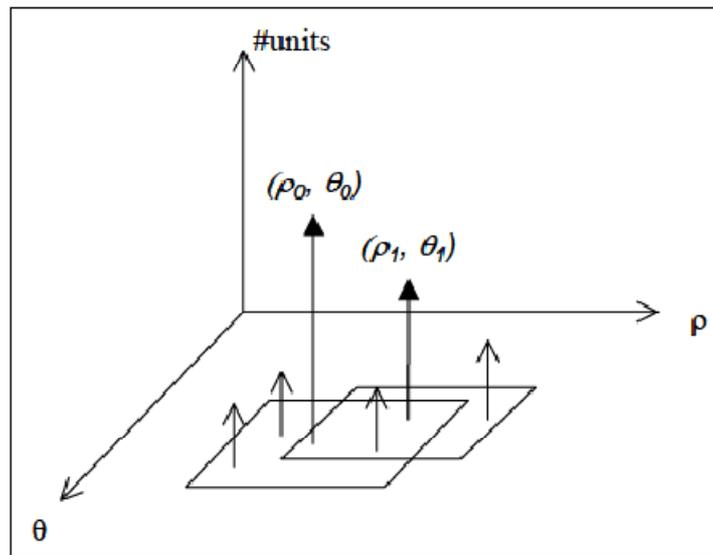
بخش های پایه‌ی ناخواسته از نقاط مینیمم که در ascender ها و descender ها شناسایی می شوند، پیدا می شوند. این بخش ها ممکن است جدا یا هم برخورد با دیگر بخش ها باشند. روش های مختلفی تعریف می شوند تا صف بندی ها را با توجه به اندازه شان یا فاصله بین خطی محلی و روی یک معیار کیفیت که موافق با صف بندی هایی است که واحدهای آن در مسیر یکسانی هستند تا واحدهای نزدیک تر که در مکانی بالاتر یا پایین تر قرار گرفته‌اند را حذف کند. صف بندی های متداخل با توجه به توپولوژی صف بندی های متداخل بازسازی می شوند. خط میانه از روی خط پایه و نقاط ماکزیمم پیدا می شود. پیکسل هایی که در یک خط زمينه و خط-

میان‌ه‌ی داده شده قرار گرفته اند، در حالتی که ascender ها و descender ها بخش بندی نشده باشند، در خط متن مرتبط خوشه بندی می‌شوند. نرخ بخش بندی‌های درست بین ۹۰٪ و ۹۹٪ با تنظیم پارامتر کافی گزارش شده است.

## ۲-۲-۴ روش های مبتنی بر تبدیل هاف

تبدیل هاف یک تکنیک بسیار محبوب برای پیدا کردن خطوط صاف در تصاویر است. در [۸۸]، روشی روی یک طرح hypothesis-validation ایجاد شده است. صف بندی های بالقوه، در دامنه‌ی هاف، فرض شده و در دامنه‌ی تصویر معتبر می‌شوند. بنابراین هیچ فرضی در مورد مسیرهای خط متن ساخته نمی‌شود. (ممکن است چند تا از آنها در یک صفحه وجود داشته باشند.) مراکز مؤلفه های متصل به هم، واحدهای تبدیل هاف هستند. مجموعه ای از واحدهای صف بندی شده در تصویر، در طول یک خط با پارامترهای  $(\rho, \theta)$  در سلول مرتبط  $(\rho, \theta)$  در دامنه‌ی هاف قرار دارند. صف بندی‌های شامل واحدهای زیاد، با سلول هایی با قله‌ی بالا در دامنه‌ی هاف مرتبط هستند. برای در نظر گرفتن نوسانات خطوط متن دست نویس، یعنی اینکه واحدهایی که در یک خط متن هستند، به طور دقیق صف بندی نشده اند، دو فرض برای هر صف بندی مد نظر قرار می‌گیرد و یک صف بندی از واحدهای ساختار سلولی یک سلول ابتدایی تشکیل شده است.

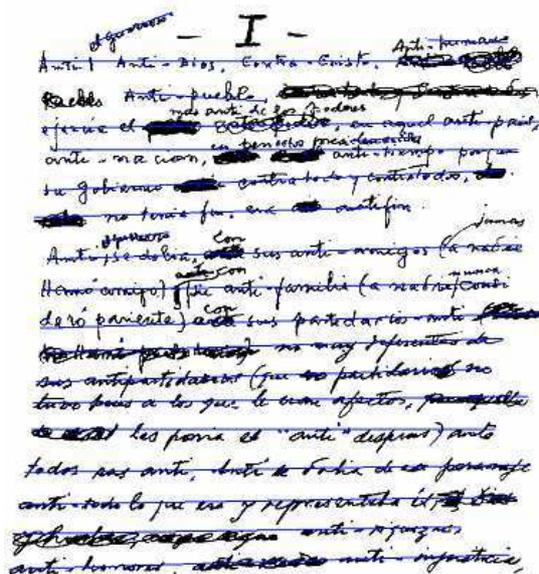
یک ساختار سلولی از یک سلول  $(\rho, \theta)$ ، شامل همه‌ی سلول‌های قرار گرفته در یک خوشه با مرکز  $(\rho, \theta)$  می‌باشد. سلول  $(\rho, \theta)$  شامل بیشترین تعداد واحدها را در نظر بگیرید. فرض دوم  $(\rho_1, \theta_1)$  در ساختار سلول  $(\rho, \theta)$  جستجو می‌شود. صف بندی انتخاب شده میان این دو فرضیه، قوی ترین آن‌هاست، یعنی آن صف بندی که شامل بیشترین تعداد واحد در ساختار سلولش است. و سلول مرتبط  $(\rho, \theta)$  یا  $(\rho_1, \theta_1)$  سلول ابتدایی است.



شکل ۲-۵: سلول‌های فرضی  $(\rho, \theta)$  و  $(\rho_1, \theta_1)$  در فضای هاف. هر قله به واحدهای به دقت صاف شده مرتبط است. [۸۰]

با این وجود، خطوط متن دقیق، به ندرت به صف‌بندی‌هایی با بیشترین تعداد واحد مربوط می‌شوند. مانند صف-بندی‌های متقاطع باید شامل واحدهای بیشتری از خطوط متن باشند. یک صف‌بندی بالقوه با استفاده از اطلاعات متنی معتبر می‌شود یعنی با در نظر گرفتن همسایه‌های داخلی و خارجی. یک همسایه درونی یک واحد  $z$  یک همسایه‌ی صف‌بندی درون هاف است. یک همسایه‌ی خارجی، یک همسایه‌ی خارج صف‌بندی هاف است که در دایره‌ای به شعاع  $\delta z$  از واحد  $z$  قرار می‌گیرد. فاصله‌ی  $\delta z$ ، فاصله‌ی میانگین فواصل همسایه‌ی داخلی از واحد  $z$  است. برای معتبر شدن، یک صف‌بندی بالقوه می‌تواند شامل واحدهای خارجی کمتری نسبت به واحدهای داخلی باشد. این، صفر رد کردن صف‌بندی‌هایی که ربط ادراکی ندارند را باعث می‌گردد.

این روش می‌تواند خطوط متن جهت‌دار شده و حاشیه‌نویسی‌های شیب‌دار را تحت این فرض که چنین خط‌هایی تقریباً صاف هستند را استخراج کند.



شکل ۲-۶: خطوط استخراج شده از یک دست نوشته، جهت‌های خطوط دنبال شده، به سلول‌های اولیه پیدا شده در فضای هاف

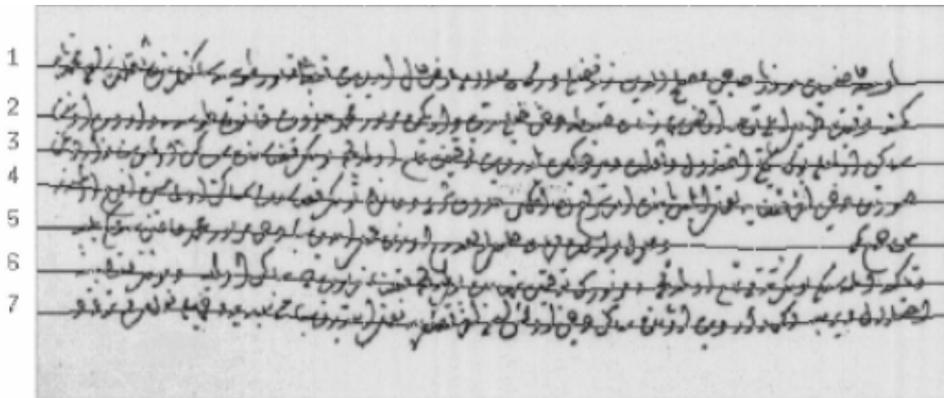
مرتبط هستند. [۸۰]

تبدیل هاف می‌تواند به خطوط نوسانی دست‌نویس‌ها هم اعمال شود. تبدیل هاف ابتدا به نقاط مینیمم (واحدها) در یک نوار عمودی روی سمت چپ تصویر اعمال می‌شود. صف‌بندی‌ها داخل دامنه هاف، با شروع از یک مسیر اصلی جستجو می‌شوند و بوسیله‌ی گروهی کردن سلول‌ها در یک جستجوی کامل در ۶ جهت این کار انجام می‌شود. پس یک پنجره‌ی متحرک که به طرح خوشه‌بندی در دامنه تصویر مرتبط است، واحدهای باقیمانده را به صف‌بندی‌ها اختصاص می‌دهد. طرح خوشه‌بندی، تولید خطوط جدید شروع شونده در وسط صفحه را ممکن می‌کند.

## ۲-۲-۵ روش شبکه‌ی دافعه-جاذبه:

یک رویکرد مبتنی بر نیروهای دافعه-جاذبه در مرجع [۸۹] ارائه شده است. این روش مستقیماً روی تصاویر سطح خاکستری کار می‌کند و شامل وفق دادن مکان  $y$  عدد از واحدهای پیش‌تعریف شده‌ی خط زمینه است. خطوط پایه یکی یکی از ابتدای تصویر تا انتهای تصویر ایجاد می‌گردند. پیکسل‌های تصویر به عنوان نیروهای

جاذب برای خطوط پایه عمل می‌کنند و خطوط پایه تازه استخراج شده به عنوان نیروهای دافعه ایفای نقش می‌کنند. خط پایه‌ای که قرار است استخراج شود تحت خط زمینه امتحان شده قبلی شروع می‌شود تا بوسیله آن دفع شود و بوسیله‌ی پیکسل‌های خط پایین جذب شود (اولین آنها در فضای خالی سر صفحه متن شروع می‌شود). خطوط باید طول‌های مشابهی داشته باشند. نتیجه، مجموعه‌ای از به اصطلاح خطوط زمینه‌هایی که هر کدام از میان بدنه کلمات عبور می‌کنند، می‌باشد. شکل ۲-۷ در این زمینه آمده است. روش پیشنهادی، بر روی سندهای باستانی و متون لاتین کاربردی است.

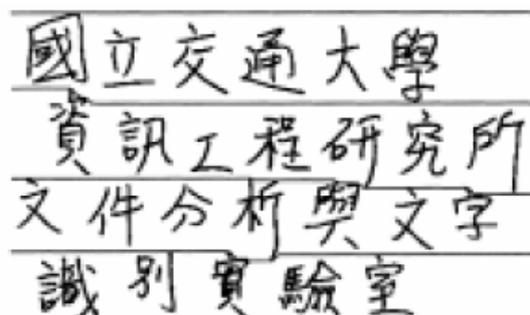


شکل ۲-۷: خطوط به اصطلاح زمینه که با کمک شبکه دافعه-جاذبه به دست آمده اند [۸۰]

## ۲-۲-۶ روش تصادفی

در اینجا روش بر مبنای الگوریتم احتمال viterbi ارائه شده است که مسیرهای غیر خطی میان خطوط متن همپوشان را استخراج می‌کند. با وجود اینکه این روش بر روی متون دست‌نوشته‌ی چینی بکار برده شده است، این اصل می‌تواند به متون تاریخی که عموماً خطوط همپوشان دارند تعمیم داده شود. خطوط بوسیله‌ی مدل hidden markov استخراج می‌شود. تصویر ابتدا به سلول‌های کوچکی تقسیم می‌شوند که هر کدام به حالتی از HMM مربوط می‌شود. بهترین مسیرهای جداسازی از چپ به راست جستجو می‌شوند. آنها به مسیرهایی که با نقاط سیاه زیادی تقاطع ندارند مرتبط می‌شوند و تا حد ممکن مستقیم هستند، با این وجود جانشین کردن گراف به چهار خانه‌هایی بالا یا پایین محدود می‌شود.

همه بهترین مسیرها که در هر موقعیت  $\gamma$  از تصویر خاتمه می‌یابند، ابتدا مورد ملاحظه قرار می‌گیرند. حذف بعضی از این مسیرها با استفاده از یک آستانه کیفیت  $T$  انجام می‌شود، مسیری که احتمال آن از  $T$  کمتر باشد حذف می‌گردد. مسیرهای تغییر محل یافته به آسانی حذف می‌شوند ( و مسیرهای بسته با توجه به معیار کیفیت حذف می‌گردند). این روش هنگامی موفق است که مسیر  $ground\ Truth$  میان خطوط متن به آرامی در طول مسیر  $\gamma$  تغییر می‌کند شکل ۸-۲ در این روش مسیرهای جداسازی را یافته است. . در صورت وجود مؤلفه- های متداخل، مسیر با بیشترین احتمال، مؤلفه‌ی همپوشان را در نقاطی با کمترین تعداد ممکن نقطه سیاه قطع می‌کنند. اما این روش ممکن است در نقاطی که تعداد زیادی پیکسل سیاه دارد شکست بخورد.



شکل ۸-۲: مسیرهای جداکننده که با الگوریتم تصادفی به دست آمده‌اند [۸۰]

فصل سوم

الگوریتم پیشنهادی

## مقدمه

همان گونه که در فصول قبل بیان گردید، اسناد فارسی که توسط دوربین، عکس برداری یا با اسکنر روبش شده اند، برای سیستم OCR فارسی، نقش ورودی را ایفا می کنند. تصاویر ورودی OCR غالباً باید دارای خطوط صاف و بدون اعوجاج و خالی از نویز باشند. از این رو، حذف اعوجاج غیرخطی موجود در خطوط سند فارسی از مراحل اصلی پیش پردازش در فرایند OCR است. یکی از این اعوجاجات که به شکل بالآمدگی در خطوط متن است و OCR را با مشکل مواجه می کند پیچش سه بعدی است. از بین بردن چرخش سه بعدی، باعث خوانایی بیشتر سند و افزایش دقت OCR در تشخیص صحیح متن می شود.

همان طور که در فصول قبل هم اشاره شد، کار بر روی حذف اعوجاج در اسناد لاتین بسیار بیشتر بوده است، اما رفع چرخش سه بعدی متن به عنوان امری ضروری در فرایند OCR کمتر مورد توجه قرار گرفته است. در این پایان نامه، یک روش جدید جهت رفع پیچش سه بعدی متن فارسی پیشنهاد شده است که در این فصل جزئیات این الگوریتم پیشنهادی مورد بررسی قرار می گیرد. نرم افزار استفاده شده جهت پیاده سازی این الگوریتم، MATLAB بوده است.

### ۱-۳ فلوجارت الگوریتم پیشنهادی

جهت بررسی الگوریتم پیشنهادی، روند کلی مراحل انجام آن، در قالب یک فلوجارت در شکل ۱-۳ نشان داده شده است.



شکل ۱-۳: فلوجارت الگوریتم پیشنهادی

در فلوجارت فوق یک بخش بندی کلی، از مراحل الگوریتم پیشنهادی، نشان داده شده است. در ادامه با توجه به ترتیب مراحل، به بررسی جزئیات مربوط به هر بخش فلوجارت می پردازیم.

۲-۳ دریافت داده‌ی اولیه

داده‌ی اولیه برای ورود به الگوریتم پیشنهادی، تصاویر سند فارسی هستند، که با دوربین دیجیتال عکس برداری شده یا توسط اسکنر روبش گردیده‌اند. این تصاویر شامل اعوجاج چرخش سه بعدی هستند و هدف رفع این اعوجاج است. در زمینه متون لاتین، مجموعه داده‌ی وجود دارد که شامل تصاویر دارای اعوجاج است. دو نمونه از تصاویر دارای اعوجاج در این مجموعه داده در شکل ۲-۳ آمده است.

CHAPTER 4. RELEVANCE FILTERING FOR WEAKLY LABELED VIDEO

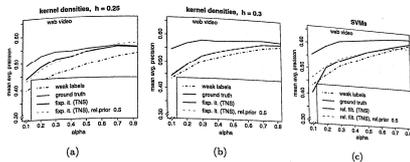


Figure 4.11: Comparing relevance filtering on raw web video training sets when using the correct relevance prior (dark blue) with using  $\alpha = 50\%$  (light blue). It can be seen that the latter choice leads to comparable results.

Finally, we address the question how to estimate the relevance prior  $\alpha$ . A simple solution is proposed, namely to set it to a "reasonable" value of 0.5 (which is typical for web-based training sets as shown in Section 4.3). Figure 4.11 compares relevance filtering (with temporal neighborhood suppression switched on) when using the true prior and when using  $\alpha = 0.5$ . It can be seen that setting  $\alpha = 0.5$  (i.e. by filtering half of all positive training samples) gives a performance comparable to the true value, at least for the range of  $\alpha = 0.2 - 0.5$  that is typical for web video. For SVMs and very noisy training sets, this choice even outperforms an aggressive filtering.

4.6 Discussion

In this chapter, *relevance filtering* has been presented, a novel approach for concept learning from user-tagged video. This method addresses the fact that web video tags are coarse, subjective, and context-dependent, such that significant amounts of training material do not show the target concept to be learned. This degrades the performance of standard concept detectors.

To address this problem, relevance filtering views web tags as weak indicators of true, latent concept labels and infers these during training. This approach can be used as a wrapper around generative base models (as demonstrated for kernel densities) as well as discriminative ones (as shown for SVMs).

PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dutra Lima, Departamento de Eletrônica e Sistemas - Universidade Federal de Pernambuco - Brazil

Abstract

This paper introduces PhotoDoc, a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an image plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.



Figure 02 - Document image acquired with the camera of HP (P4370) (1280x60 pixels) 162KB, without strobe flash

This new use of portable digital camera gave birth to a new research area [1] [2] that is evolving fast in many different directions.

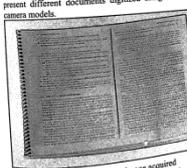


Figure 01 - Document image acquired with the camera of a LG cell phone LG-970 (1400x1200 pixels) 409KB, without strobe flash



Figure 03 - Document image acquired with camera Sony DSC-P40 (1.1 Mpixel)

شکل ۲-۳: نمونه‌هایی از یک مجموعه داده لاتین

برای زبان فارسی مجموعه داده‌ای در این زمینه موجود نیست و اسناد کمی در این باب جمع‌آوری شده‌اند. لذا بر آن شدیم تا مجموعه داده‌ای از تصاویر با اعوجاج سه بعدی تهیه کنیم. مجموعه داده تهیه شده، شامل ۱۲۰ عکس می‌باشد که با دوربین دیجیتال canon IXY 200F, 12.1 Mega pixels عکس برداری شده است و تعدادی دیگر نیز توسط اسکنر canon LiDE 110, اسکن شده‌اند. در تولید نمونه‌های ایجاد شده، از ۱۰ قلم فارسی معروف که توسط نرم افزار پرشیاانگار قابل شناسایی هستند، استفاده شده است. این قلم‌ها عبارتند از:

B-lotus, Bnazanin, mitra, Yagut, Zar, Traffic, Titr, Tahoma, Homa, Times New Romans

در تولید نمونه های مجموعه داده فارسی سعی شده است تا انواع اعوجاجات ممکن سه بعدی استفاده شود.

در شکل ۳-۳ نمونه هایی از این مجموعه داده نشان داده شده اند.

**نوار چسب ملکولی**

هنگامی که ما اجزای سازنده یک سیستم را در کنار هم جمع می کنیم و به آن ها اجازه برهمکنش با یکدیگر می دهیم، ساختارهای بسیار پیچیده ای شکل می گیرند. یک راه مناسب برای تعیین اینکه کدام اجزا به هم متصل شوند این است که در آن ها آرایشی از دی ان ای ها جاسازی کنیم که فقط به اجزایی اجازه اتصال می دهند که زنجیره دی ان ای آن ها مکمل یکدیگر باشد.

لانگ فنگ و همکارانش در دانشگاه نیویورک، مقاله ای را در فیزیکال ریویو لترز چاپ کرده اند که نشان می دهد روش دیگری نیز برای متصل کردن ذرات ریز به یکدیگر وجود دارد. روشی که آن ها معرفی کرده اند، از درهم تنیدگی حلقه های دی ان ای با ذرات مجاورشان استفاده می کند.

بعد از کودتای ۲ دسامبر ۱۸۵۱ به بروکسل گریخت و در تبعید در امدت خود، آثار بزرگی تدوین کرد. با سقوط ناپلئون سوم در سال ۱۸۷۰ به مین بازگشت. به مدت چندین سال او نماد مخالفت با پادشاهی و طرفدار جمهوری بود. در سال ۱۸۷۱ به مجلس ملی راه یافت ولی خیلی زود از نمایندگی مجلس کناره گرفت. در سال ۱۸۷۴ بی اعتنا نسبت به تقدیمات قاریخی طبیعت گویان، کتاب نود و سه را نوشت. در سن هفتاد و پنج سالگی کتاب دانشمین هنر پیدر بزرگ بودن را نوشت؛ اما همچنان به دنیای سیاست تمایل داشت و در سال ۱۸۷۶ به مجلس سنا راه یافت. در فوریه ۱۸۸۱ به سلامت ورود به هشتاد سالگی مراسم با شکوهی به افتخار وی بر پا گردید که کمتر کسی به زمال حیات خود، چنین افتخاری را کسب کرده است.

هنگامی که ما اجزای سازنده یک سیستم را در کنار هم جمع می کنیم و به آن ها اجازه برهمکنش با یکدیگر می دهیم، ساختارهای بسیار پیچیده ای شکل می گیرند. یک راه مناسب برای تعیین اینکه کدام اجزا به هم متصل شوند این است که در آن ها آرایشی از دی ان ای ها جاسازی کنیم که فقط به اجزایی اجازه اتصال می دهند که زنجیره دی ان ای آن ها مکمل یکدیگر باشد.

لانگ فنگ و همکارانش در دانشگاه نیویورک، مقاله ای را در فیزیکال ریویو لترز چاپ کرده اند که نشان می دهد روش دیگری نیز برای متصل کردن ذرات ریز به یکدیگر وجود دارد. روشی که آن ها معرفی کرده اند، از درهم تنیدگی حلقه های دی ان ای با ذرات مجاورشان استفاده می کند.

هنگامی که ما اجزای سازنده یک سیستم را در کنار هم جمع می کنیم و به آن ها اجازه برهمکنش با یکدیگر می دهیم، ساختارهای بسیار پیچیده ای شکل می گیرند. یک راه مناسب برای تعیین اینکه کدام اجزا به هم متصل شوند این است که در آن ها آرایشی از دی ان ای ها جاسازی کنیم که فقط به اجزایی اجازه اتصال می دهند که زنجیره دی ان ای آن ها مکمل یکدیگر باشد.

لانگ فنگ و همکارانش در دانشگاه نیویورک، مقاله ای را در فیزیکال ریویو لترز چاپ کرده اند که نشان می دهد روش دیگری نیز برای متصل کردن ذرات ریز به یکدیگر وجود دارد. روشی که آن ها معرفی کرده اند، از درهم تنیدگی حلقه های دی ان ای با ذرات مجاورشان استفاده می کند.

تیم فنگ کره هایی از جنس پلی استایرن با شعاع یک میکرون را که حاوی تعداد بسیار زیادی مولکول دی ان ای دارای زنجیره های خود مکمل بودند در یک آرایه قرار دادند. هر یک از این مولکول ها می تواند به مولکول دیگری که زنجیره اش به شکل تصویر آینه ای خود مولکول است متصل شود.

تیم فنگ از کره هایی استفاده نمودند که دارای هسته مغناطیسی بودند. این کره ها زمانی که در یک میدان مغناطیسی کوچک قرار می گیرند زنجیره هایی را تشکیل می دهند. این امر به معنای آن است که کره ها هنگام تشکیل حلقه ها (که با سرد کردن ترکیب اتفاق می افتد) به اندازه کافی به هم نزدیک هستند. آن ها دریافتند که حتی پس از حذف میدان مغناطیسی نیز کره ها نسبت به یکدیگر مقید باقی می مانند.

اما همان طور که برای حلقه های درهم تنیده نیز انتظار داریم، اضافه کردن یک آنزیم که به رشته دی ان ای ها اجازه می دهد بدون صدمه از میان یکدیگر رد شوند باعث از هم گسیختن زنجیره ها می شود. بر خلاف روش های قدیمی برای اتصال دی ان ای ها، این نوع اتصال چسب مانند به ترتیب مراحل فرآیند بستگی دارد و روش های نوینی را برای گردآوری ساختارهای پیچیده پیش روی ما قرار می دهد.

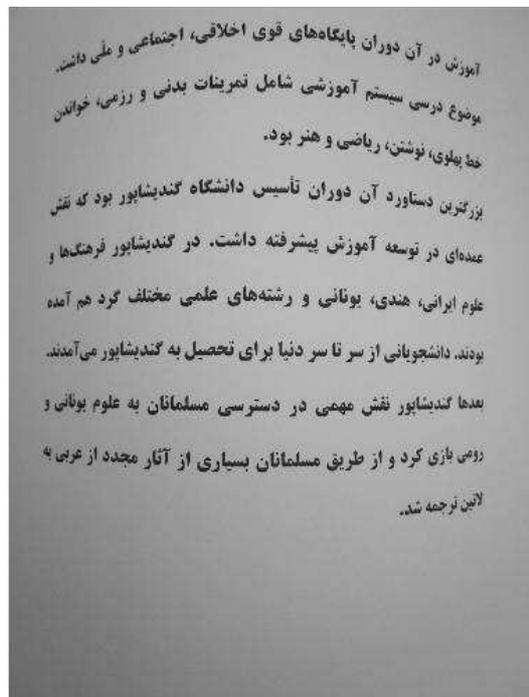
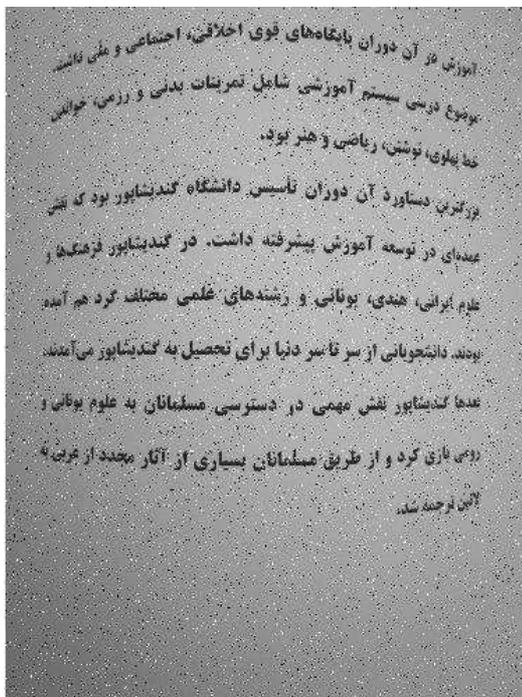
شکل ۳-۳: نمونه هایی از مجموعه داده تهیه شده

### ۳-۳ رفع نویز

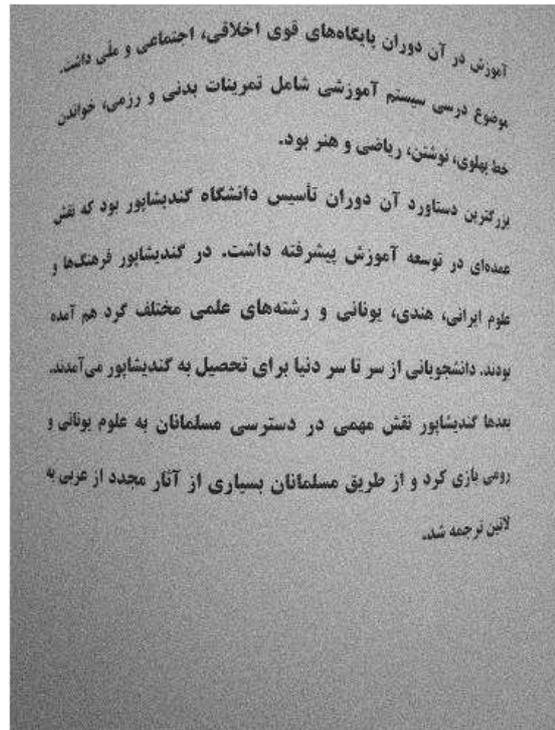
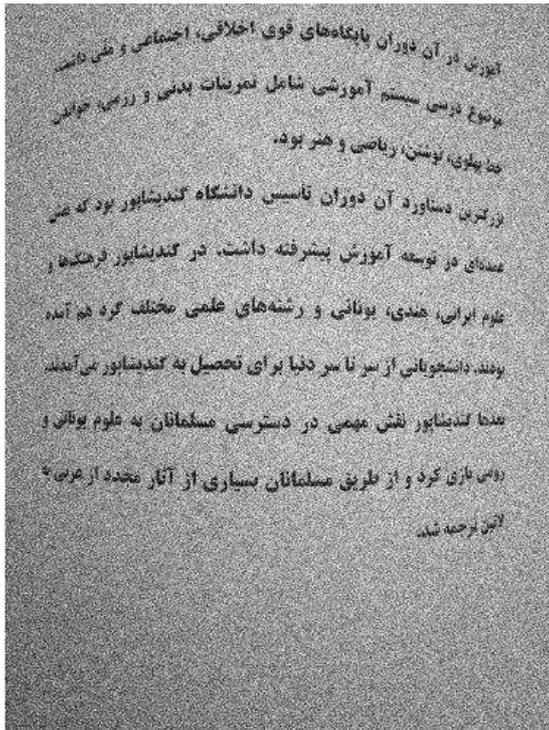
تصاویر دیجیتال در معرض انواع مختلف نویز هستند. در پروژه حاضر منبع نویز می تواند اسکنر یا دوربین باشد. انتقال الکترونیکی تصویر می تواند سبب ایجاد نویز در آن شود. بیشترین نویزهای دیده شده در تصاویر نویز فلغل نمکی و نویز گوسی هستند.

نویزهای فلغل نمکی به صورت نقطه های سیاه و سفید در تصویر دیده می شوند و همین تفاوت زیاد سطح روشنایی آن ها با نقاط مجاور در تصویر، معیار مناسبی برای شناسایی و حذف آن ها می باشد. در این بخش از فیلتر میانه استفاده شده است. این فیلتر یک عملگر غیر خطی برای کاهش این نوع نویز است. فیلتر میانه در حذف نویز فلغل نمکی از کانولوشن کارآمد تر است و لبه ها را حفظ می کند.

نویزهای گوسین موجود در تصاویر را هم می توان با نرم کردن تصویر توسط فیلتر میانگین تا حدودی برطرف کرد. در شکل های ۳-۴ و ۳-۵ دو تصویر که به آن ها نویز اضافه شده است و با فیلتر میانه و میانگین رفع نویز شده اند دیده می شود.



شکل ۳-۴: تصویر اصلی (سمت چپ)، و تصویر حاصل از فیلتر میانه (سمت راست)



شکل ۳-۵: تصویر اصلی (سمت چپ)، و تصویر حاصل از فیلتر میانگین (سمت راست)

البته لازم به ذکر است که در مجموعه داده‌ی ما نویز چندانی وجود ندارد و لذا این مرحله اعمال نمی‌شود.

### ۳-۴ دو سطحی سازی

بعد از برطرف کردن نویز تصویر، نوبت به دو سطحی سازی تصویر می‌رسد. در فصل قبل انواع روش‌های آستانه گذاری مطرح شد. در پروژه‌ی حاضر از الگوریتم Otsu برای دو سطحی سازی تصویر استفاده شده است. در روش Otsu به دنبال آستانه‌ای می‌گردیم که بتواند پیش‌نما را از پس زمینه جدا کند. این آستانه باید قادر باشد تا حد امکان گستره روشنایی هر کلاس را محدودتر کند و همپوشانی بین دو کلاس به کمترین مقدار ممکن برسد. در الگوریتم Otsu واریانس داخل کلاسی به صورت زیر تعریف می‌شود.

$$\sigma_{\text{within}}^2(T) = \omega_B(T)\sigma_B^2 + \omega_0(T)\sigma_0^2(T) \quad (1-4)$$

که در آن داریم: ( $T$  آستانه و  $L$  شدت روشنایی است)

$$\omega_B(T) = \sum_{i=0}^{T-1} p(i) \quad \omega_0(T) = \sum_{i=T}^{L-1} p(i) \quad (2-4)$$

و  $\sigma_B^2(T)$  واریانس پیکسل های پس زمینه (زیر آستانه) و  $\sigma_B^2(T)$  واریانس پیکسل های پیش نما (بالای آستانه) می باشد.

راه آسان تر برای واریانس بین کلاسی است یعنی:

$$\begin{aligned}\sigma_{\text{Between}}^2(T) &= \sigma^2 - \sigma_{\text{Within}}^2(T) \\ &= \omega_B(T)[\mu_B(T) - \mu]^2 + \omega_0(T)[\mu_0(T) - \mu]^2\end{aligned}\quad (3 - 4)$$

که با ساده کردن آن داریم:

$$\sigma_{\text{Between}}^2(T) = \omega_0(T)[\mu_B(T) - \mu_0(T)]^2 \quad (4 - 4)$$

که با ماکزیمم کردن رابطه، آستانه (T) بهینه به دست می آید.

این کار با استفاده از روابط بازگشتی ساده زیر نیز امکان پذیر است.

$$\omega_B(T + 1) = \omega_B(T) + p(T) \quad (5 - 4)$$

$$\omega_0(T + 1) = \omega_0(T) - p(T) \quad (6 - 4)$$

$$\mu_B(T + 1) = \frac{\mu_B(T)\omega_B(T) + p(T)\mu}{\omega_B(T+1)} \quad (7 - 4)$$

$$\mu_0(T + 1) = \frac{\mu_0(T)\omega_0(T) - p(T)\mu}{\omega_0(T+1)} \quad (8 - 4)$$

در شکل ۳-۶ و ۳-۷ تصویری از مجموعه داده به همراه تصویر دوسطحی شده آن و نیز هیستوگرام آن قبل و بعد از پیاده سازی الگوریتم دوسطحی سازی دیده می شود.

(ب)

حیات و استر، که طعم دهنده بسیاری از میوه‌ها است، ضروری است. ایزوتوپ  $^{14}\text{C}$  به طور متداول در سن یابی پرتوزایشی کاربرد دارد. کربن به دلایل زیادی قابل توجه است. اشکال مختلف آن شامل یکی از نرم‌ترین‌ها مثل گرافیت و یکی از سخت‌ترین‌ها الماس مواد شناخته شده توسط انسان می‌باشد.

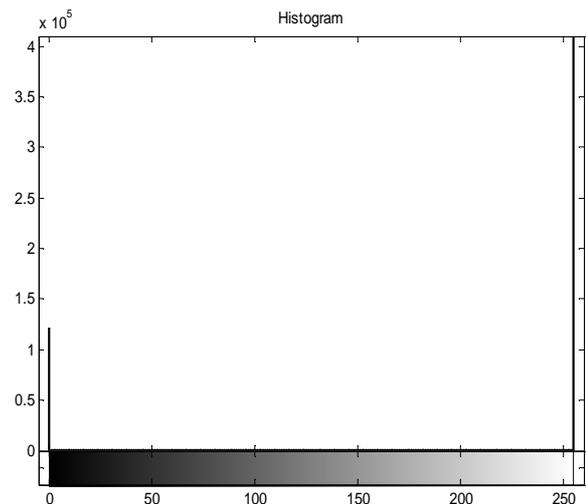
افزون بر این، کربن میل زیادی به پیوند با اتمهای کوچک دیگر در جمله اتمهای دیگر کربن، داشته و اندازه بسیار کوچک آن امکان پیوندهای متعدد را بوجود می‌آورد. این خصوصیات باعث شکل‌گیری ده میلیون ترکیبات کربنی شده است. ترکیبات کربن زیر بنای حیات در زمین می‌سازند و چرخه کربن - نیتروژن قسمتی از انرژی تولید شده توسط خورشید و ستارگان دیگر را تأمین می‌کند.

(الف)

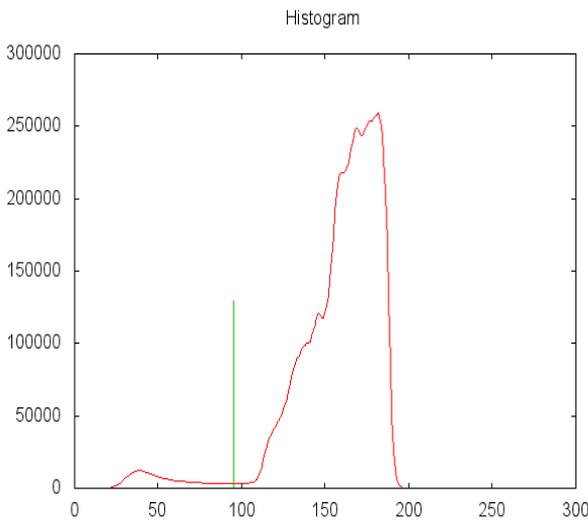
حیات و استر، که طعم دهنده بسیاری از میوه‌ها است، ضروری است. ایزوتوپ  $^{14}\text{C}$  به طور متداول در سن یابی پرتوزایشی کاربرد دارد. کربن به دلایل زیادی قابل توجه است. اشکال مختلف آن شامل یکی از نرم‌ترین‌ها مثل گرافیت و یکی از سخت‌ترین‌ها الماس مواد شناخته شده توسط انسان می‌باشد.

افزون بر این، کربن میل زیادی به پیوند با اتمهای کوچک دیگر از جمله اتمهای دیگر کربن، داشته و اندازه بسیار کوچک آن امکان پیوندهای متعدد را بوجود می‌آورد. این خصوصیات باعث شکل‌گیری ده میلیون ترکیبات کربنی شده است. ترکیبات کربن زیر بنای حیات در زمین می‌سازند و چرخه کربن - نیتروژن قسمتی از انرژی تولید شده توسط خورشید و ستارگان دیگر را تأمین می‌کند.

(د)



(ج)



شکل ۳-۶: (الف) تصویر اصلی، (ب) دوسطحی شده آن، (ج) هیستوگرام تصویر اصلی، (د) هیستوگرام تصویر دوسطحی

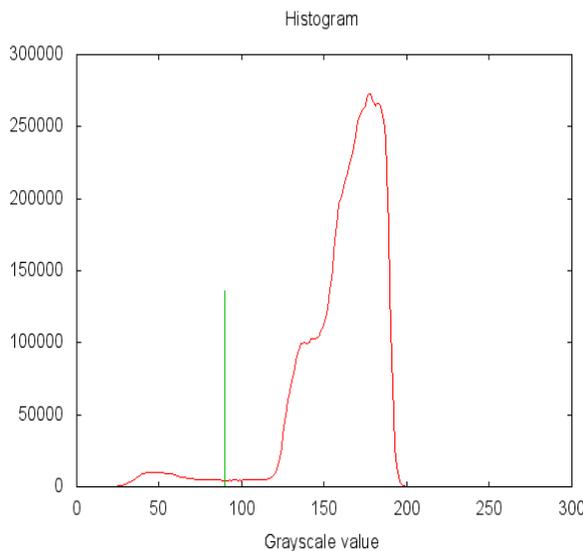
(الف)

انتقال مالکیت کالا و خدمات از یک شخص یا نهاد به دیگری به ازای دریافت چیزی از خریدار است. به طور کلی هرگونه عمل قابل سنجش و اندازه گیری مادی را که اشخاص در مقابل کالا یا خدمتی، کالا یا خدمتی را واگذار نموده و هر دو طرف به هنگام این عمل راضی باشند تجارت گفته می‌شود. تجارت به دو قسمت تجارت داخلی و تجارت خارجی تقسیم می‌شود. در عرف معمول به تبادل کالاها یا خدمات تجارت اطلاق می‌شود و برای بهتر معامله کردن در زمان لازم و محدود نیاز به اطلاعات و مدیریت کردن اطلاعات جهت رسیدن به هدف است. تجارت سازوکاری است که هسته سرمایه‌داری را تشکیل می‌دهد. مواد مورد استفاده برای ایجاد جواهرات در مصر از سال سه هزار قبل از میلاد معامله می‌شد. مسیرهای دور تجارت برای اولین بار در سه هزار سال قبل از میلاد ظاهر شد. زمانی که سومری‌ها در بین النهرین با تمدن هاراپان از دره سند معامله می‌کردند. فنیقی‌ها، معامله گران دریا اشاره شده‌اند. در

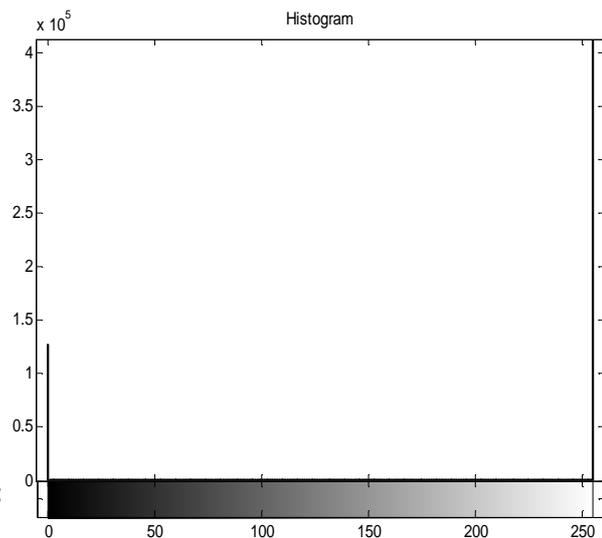
(ب)

انتقال مالکیت کالا و خدمات از یک شخص یا نهاد به دیگری به ازای دریافت چیزی از خریدار است. به طور کلی هرگونه عمل قابل سنجش و اندازه گیری مادی را که اشخاص در مقابل کالا یا خدمتی، کالا یا خدمتی را واگذار نموده و هر دو طرف به هنگام این عمل راضی باشند تجارت گفته می‌شود. تجارت به دو قسمت تجارت داخلی و تجارت خارجی تقسیم می‌شود. در عرف معمول به تبادل کالاها یا خدمات تجارت اطلاق می‌شود و برای بهتر معامله کردن در زمان لازم و محدود نیاز به اطلاعات و مدیریت کردن اطلاعات جهت رسیدن به هدف است. تجارت سازوکاری است که هسته سرمایه‌داری را تشکیل می‌دهد. مواد مورد استفاده برای ایجاد جواهرات در مصر از سال سه هزار قبل از میلاد معامله می‌شد. مسیرهای دور تجارت برای اولین بار در سه هزار سال قبل از میلاد ظاهر شد. زمانی که سومری‌ها در بین النهرین با تمدن هاراپان از دره سند معامله می‌کردند. فنیقی‌ها، معامله گران دریا اشاره شده‌اند. در

(ج)



(د)



شکل ۳-۷: (الف) تصویر اصلی، (ب) دوسطحی شده آن، (ج) هیستوگرام تصویر اصلی، (د) هیستوگرام تصویر دوسطحی

### ۳-۵ تشخیص خطوط متن

روش‌های تشخیص خطوط متن در مراجع گوناگون در فصل قبل بررسی شد. از میان همه‌ی این روش‌ها، در الگوریتم پیشنهادی از روش آغشته سازی استفاده شده است. در اینجا دو روش آغشته سازی<sup>۱</sup> گسترش و آغشته سازی با کمک الگوریتم DLCM<sup>۲</sup> مورد بحث قرار می‌گیرد.

#### ۳-۵-۱ آغشته سازی با استفاده از گسترش

جهت انجام آغشته سازی از المان ساختاری افقی استفاده شده است، زیرا خط متن چیدمانی افقی دارد. سپس با استفاده از عملگر مورفولوژی گسترش سعی شده هر خط متن آغشته شود، به این معنی که تمام فضاهای خالی بین حروف و کلمات پر شده و همه‌ی این اجزا به هم متصل شوند. این ایده در برخی موارد با مشکلاتی روبه روست. با توجه به این مشکلات مناسب‌تر دیده شد تا جهت آغشته سازی از الگوریتم دیگری استفاده شود. جزئیات مربوط به این الگوریتم در ادامه ارائه می‌شود و مواردی که منجر به انتخاب الگوریتمی جدید جهت آغشته سازی شد بیان می‌گردد.

#### ۳-۵-۲ آغشته سازی با استفاده از الگوریتم DLCM

جهت آغشته سازی با کمک الگوریتم DLCM، ابتدا باید ماتریس DLCM را تشکیل داد [90]. برای این کار لازم است صفحه‌ی سند دارای اعوجاج را، تحت مراحل قبلی گفته شده، رفع نویز و دوسطحی کرد. از آنجا که تصاویر سند دارای چرخش سه بعدی هستند، لازم است صفحه به چند ستون تقسیم شود تا تداخل بین خطوط دارای اعوجاج متن پیش نیاید. با آزمون و خطا مشخص شد که تعداد ستون بهینه برای این کار، سه ستون می‌باشد و تعداد ستون بیشتر تأثیری در نتیجه نداشته و تنها زمان اجرای برنامه را طولانی‌تر می‌کند.

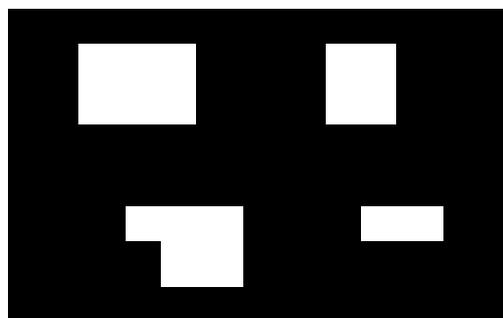
<sup>۱</sup> Dilation

<sup>۲</sup> Dynamic Local Connectivity Map

حال برای هر ستون، ماتریس DLCM تشکیل می‌شود. برای این کار ماتریسی هم سایز با ماتریس هر ستون از تصویر در نظر گرفته می‌شود، که درایه های اولیه ی آن همگی صفر هستند. ماتریس تصویر از سطر اول به صورت سطر به سطر روبش می‌گردد. فاصله ی هر درایه ی صفر تا اولین یک در هر سطر، تعیین شده و درایه ی مربوط در ماتریس DLCM برابر فاصله ی تعیین شده، قرار داده می‌شود. اگر درایه ای در ماتریس تصویر برابر یک باشد، در ماتریس DLCM نیز یک در نظر گرفته می‌شود.

به عنوان مثال یک تصویر نمونه و ماتریس آن در شکل ۳-۸ نشان داده‌اند.

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



شکل ۳-۸: تصویر نمونه (سمت راست)، ماتریس آن (سمت چپ)

ماتریس DLCM مربوط به تصویر فوق با توجه به توضیحات ارائه شده ایجاد می‌گردد. این ماتریس در شکل ۳-۹ نشان داده شده است.

|   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 1 | 1 | 1 | 3 | 2 | 1 | 1 | 1 | 0 | 0 | 0 |
| 2 | 1 | 1 | 1 | 1 | 3 | 2 | 1 | 1 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 2 | 1 | 1 | 1 | 1 | 3 | 2 | 1 | 1 | 1 | 0 | 0 |
| 4 | 3 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

شکل ۳-۹: ماتریس ساخته شده توسط الگوریتم DLCM بر روی تصویر نمونه شکل ۳-۸

حال، بر روی ماتریس DLCCM تشکیل شده برای هر ستون، یک آستانه اعمال می‌گردد، طوری که درایه‌های کوچکتر از این آستانه و البته غیر صفر ماتریس، همگی یک شوند. این به معنی پر شدن فضاهای خالی بین کلمات و پیوسته شدن همه‌ی اجزای هر خط در یک ستون یا همان آغشته سازی است. ماتریس حاصل از آستانه‌گذاری ماتریس شکل ۳-۹ با تعیین مقدار  $T=4$  به همراه تصویر این ماتریس در شکل ۳-۱۰ آمده است. همانگونه که در تصویر دیده می‌شود نقاط روی یک خط با انتخاب آستانه مناسب به هم متصل می‌گردند.

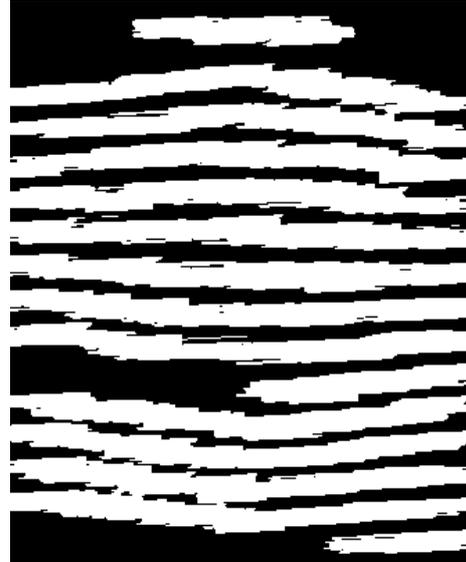
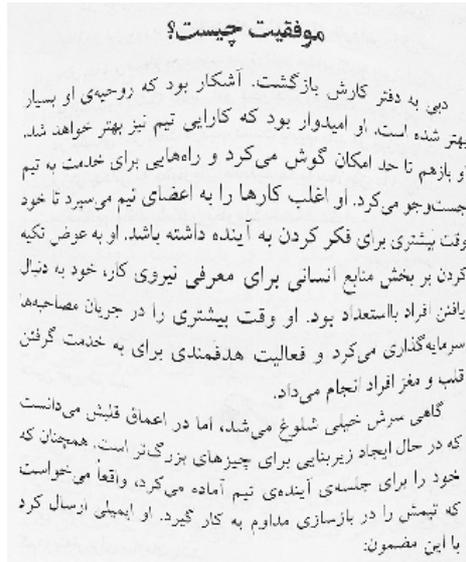
|   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0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



شکل ۳-۱۰: تصویر آغشته شده توسط الگوریتم DLCCM (سمت راست) و ماتریس آن (سمت چپ)

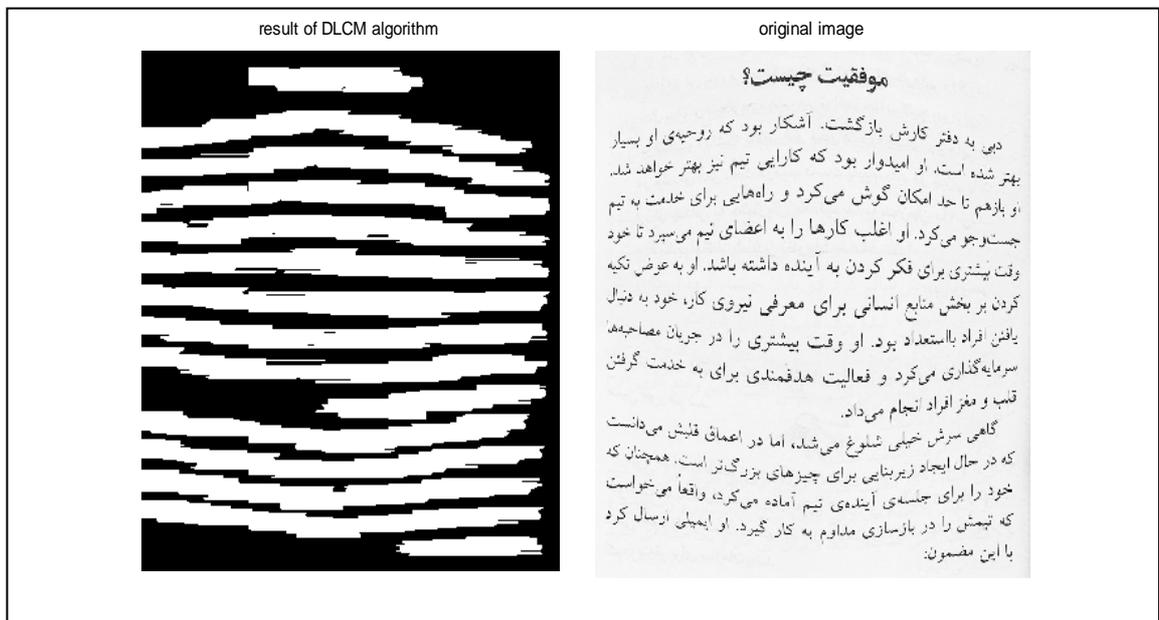
بنابراین جهت آغشته‌سازی توسط الگوریتم DLCCM همانگونه که بیان شد ابتدا تصویر به سه ستون تقسیم می‌گردد، اکنون برای تشکیل ماتریس آغشته شده‌ی نهایی، کافی است ماتریس‌های آغشته شده برای هر ستون به ترتیب در کنار هم در یک ماتریس واحد قرار گیرند.

علت انتخاب الگوریتم DLCCM، مشکلات الگوریتم گسترش بود. مثلاً پیدا کردن پارامتر بهینه برای المان ساختاری که بتواند هم همه‌ی خطوط را درست و مجزا تشخیص دهد و هم اجزای هر خط را به طور پیوسته به هم وصل کند، در بعضی تصاویر ناممکن است. به عنوان نمونه در شکل ۳-۱۱ تصویر اصلی انتخاب شده جهت آغشته‌سازی و نتیجه‌ی حاصل از کار با عملگر مورفولوژی dilation نشان داده شده است. همانگونه که در شکل دیده می‌شود امکان آغشته‌سازی تصویر به نحوی که تمامی خطوط مجزا گردند وجود ندارد.



شکل ۳-۱۱: تصویر اصلی (سمت چپ)، و تصویر آغشته شده با عملگر گسترش (سمت راست)

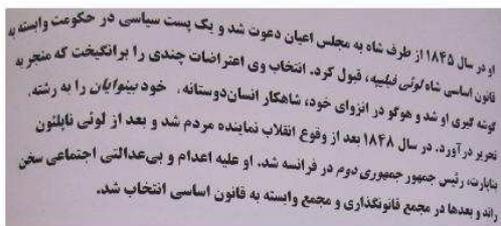
حال تصویر بالا را که توسط عملگر مورفولوژی dilation آغشته شده است توسط الگوریتم DLCCM نیز آغشته‌سازی می‌کنیم. نتیجه حاصل در شکل ۳-۱۲ نشان داده شده است، همانگونه که مشخص است در این روش آغشته‌سازی بدرستی انجام گرفته است.



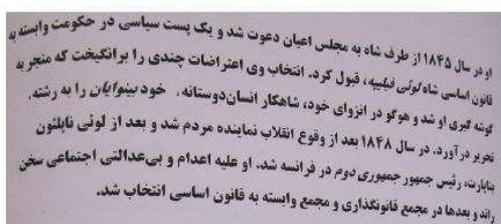
شکل ۳-۱۲: تصویر اصلی (سمت چپ)، و تصویر آغشته شده با DLCCM (سمت راست)

مشکل دوم در استفاده از گسترش، این است که مقدار بهینه‌ی انتخاب شده در یک تصویر به درستی خطوط را تشخیص داده، اما در تصویر دیگر همان مقادیر باعث تشخیص نادرست خط متن می‌شود. به عبارتی دیگر، در این روش با توجه به نوع تصویر (فاصله‌ی میان خطوط و کلمات)، مقدار بهینه‌ی پارامترهای المان ساختاری مدام در حال تغییر است..

در شکل ۳-۱۳ یک تصویر و آغشته شده آن که توسط الگوریتم گسترش آغشته شده است، دیده می‌شود. مقدار آستانه آغشته‌سازی در حدی انتخاب شده‌است که خطوط به درستی جداسازی شوند. در شکل ۳-۱۴ همان تصویر این بار توسط روش پیشنهادی و با کمک الگوریتم DLCCM آغشته‌سازی شده است. در اینجا نیز مقدار آستانه آغشته‌سازی در حدی انتخاب شده است که خطوط به درستی جداسازی شوند.



شکل ۳-۱۳: تصویر اصلی (سمت چپ)، و تصویر آغشته شده با گسترش (سمت راست)



شکل ۳-۱۴: تصویر اصلی (سمت چپ)، و تصویر آغشته شده با DLCCM (سمت راست)

حال با همان مقادیر آستانه انتخاب شده برای هر دو روش آغشته‌سازی، تصویر دیگری را آغشته‌سازی می‌کنیم، همانگونه که در شکل ۳-۱۵ دیده می‌شود، با آغشته‌سازی به روش گسترش تشخیص خطوط

بدرستی انجام نمی‌شود. اما در شکل ۳-۱۶ می‌توان دید که روش پیشنهادی با همان مقدار آستانه قبلی آغشته‌سازی را بدرستی انجام داده است.



شکل ۳-۱۵: تصویر اصلی (سمت چپ)، و تصویر آغشته شده با گسترش (سمت راست)

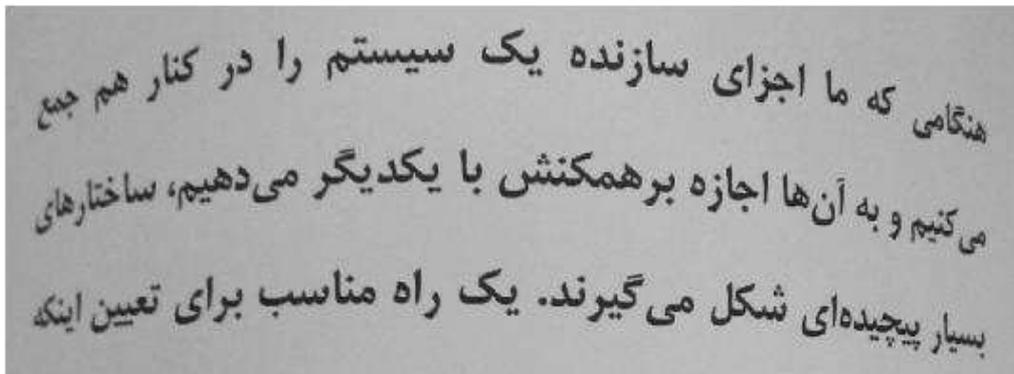


شکل ۳-۱۶: تصویر اصلی (سمت چپ)، و تصویر آغشته شده با DLCM (سمت راست)

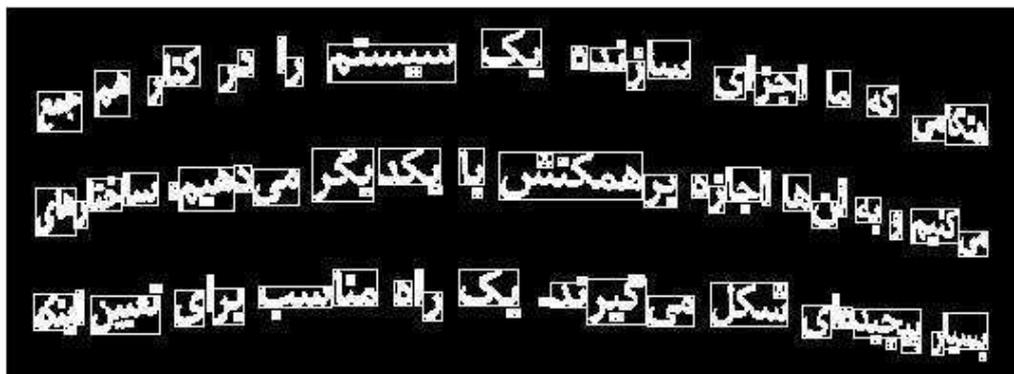
با توجه به نتایج دو روش آغشته‌سازی بیان شده، می‌توان گفت که هر چند در مورد برخی از تصاویر نتیجه حاصل از هر دو روش مشابه است اما در حالت کلی عملکرد روش DLCM مناسب‌تر می‌باشد، چراکه در برخلاف الگوریتم گسترش در مورد اکثر تصاویر آغشته‌سازی را به نحو مناسبی انجام می‌دهد و نسبت به الگوریتم گسترش نیاز کمتری به تغییر آستانه دارد. سرعت DLCM هم نسبت به الگوریتم گسترش بیشتر است. بنابراین در این پروژه جهت آغشته‌سازی از الگوریتم DLCM به نحوی که بیان گردید استفاده می‌شود.

### ۳-۶ تشخیص زیر کلمات و تعیین شیب آن‌ها

پس از مشخص شدن خطوط، باید به رفع اعوجاج سه بعدی در هر خط پرداخت. برای این کار ابتدا باید، تمامی زیر کلمات خط، مشخص شده و میزان چرخش در هر کلمه‌ی خط، تعیین گردد. بنابراین ابتدا کوچکترین مستطیل محاط بر هر جزء به هم پیوسته در خط را تشکیل داده می‌شود. در شکل ۳-۱۷ یک تصویر نمونه و در شکل ۳-۱۸ تصویر دوسطحی شده و مستطیل‌های دربرگیرنده‌ی زیر کلمات برای این تصویر نمونه، نشان داده شده اند.



شکل ۳-۱۷: تصویر نمونه

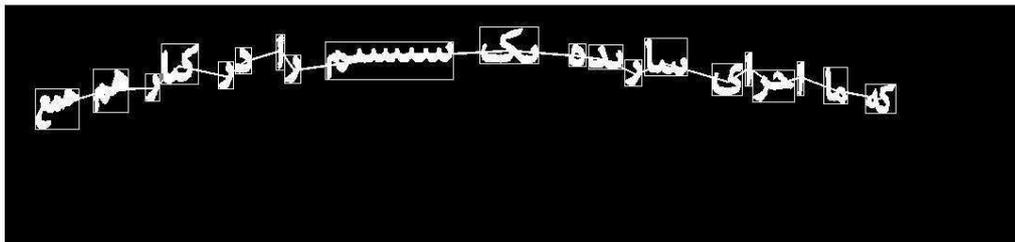


شکل ۳-۱۸: تصویر دوسطحی شده شکل ۳-۱۷ به همراه مستطیل‌های در بر گیرنده هر زیر کلمه

با توجه به این که هر زیر کلمه از خط دارای درجه‌ی خاصی از اعوجاج است، باید برای هر زیر کلمه میزان چرخش تعیین شود. برای این کار باید شیب (میزان چرخش) هر زیر کلمه را نسبت به خط افق محاسبه شود.

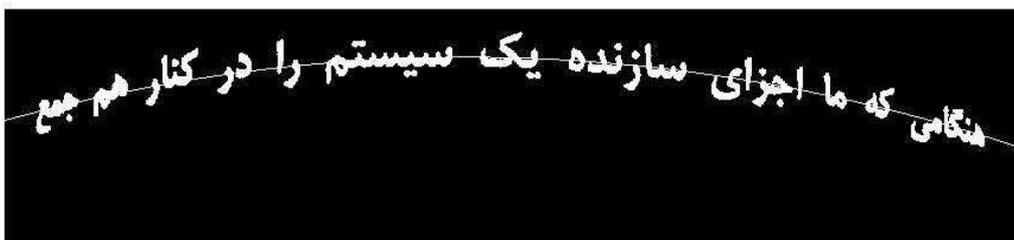
جهت تعیین شیب از دو روش استفاده شده است. در روش نخست، شیب موجود در هر زیر کلمه را از روی شیب خطوط متصل کننده‌ی مراکز جعبه‌ها تعیین می‌گردد. در این روش مناسب تر است که نقاط حذف شوند. زیرا باعث ایجاد خطای بسیار در تعیین شیب زیر کلمات می‌شوند. جهت حذف نقاط یک آستانه گذاری روی سطح جعبه‌ها انجام می‌شود و جعبه‌هایی را که سطح آن‌ها از آستانه‌ی خاصی کمتر بودند را حذف کرده و شیب برای سایر جعبه‌ها مشخص می‌شود. آستانه از طریق آزمون و خطا به دست می‌آید.

در شکل ۱۹-۳ برای خط اول تصویر نمونه شکل ۳-۱۷ نقاط خط با اعمال یک مقدار آستانه حذف شده و خطوط متصل کننده مرکز جعبه‌ها ترسیم شده‌اند. حال می‌توان میزان چرخش جعبه‌ها را با استفاده از شیب خطوط متصل کننده مراکز جعبه‌ها تعیین نمود.



شکل ۱۹-۳ : خطوط اتصال دهنده مرکز جعبه‌ها برای خط اول تصویر ۳-۱۷

در روش دوم، شیب از طریق برازش منحنی بر مرکز جعبه‌ها تعیین شده است. با توجه به اعوجاج‌های ممکن مناسب‌ترین منحنی جهت برازش، پس از سعی و خطا، چندجمله‌ای درجه‌ی ۲ می‌باشد. در شکل ۳-۲۰ منحنی برازش شده بر خط اول تصویر نمونه نشان شده است.



شکل ۳-۲۰ : منحنی برازش شده بر خط اول تصویر ۳-۱۷

در شکل قبل نقاط خط حذف نشده است. در شکل ۳-۲۱ منحنی برازش شده پس از حذف نقاط نشان داده شده است.



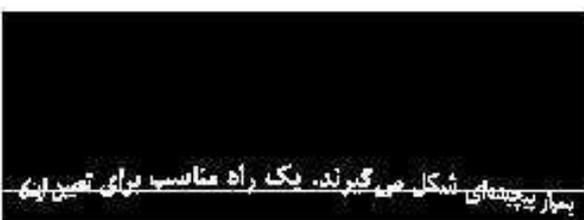
شکل ۳-۲۱: منحنی برازش شده بر خط اول تصویر ۳-۱۷ پس از حذف نقاط

همانگونه که در شکل‌های ۳-۲۰ و ۳-۲۱ دیده می‌شود، در این روش حذف نقاط یا عدم حذف آن‌ها تأثیر چندانی در منحنی حاصل ندارد. با مشخص شدن معادله‌ی منحنی و مشتق‌گیری از آن می‌توان شیب مربوط به هر زیر کلمه را تعیین کرد.

### ۳-۷ رفع چرخش کلمات و خطوط

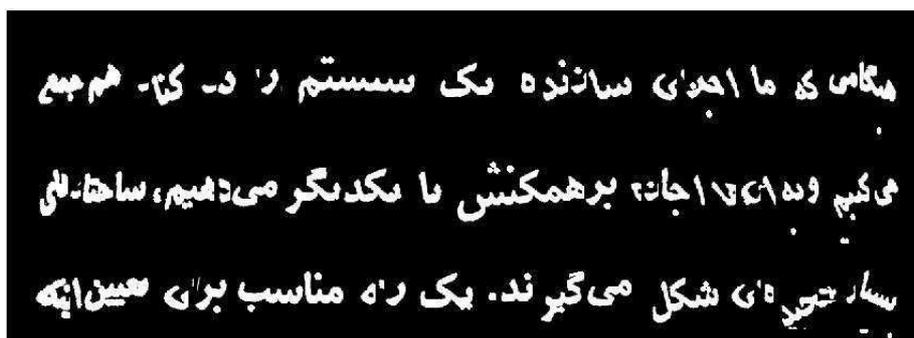
با مشخص شدن شیب زیر کلمات، باید خط زمینه هر خط مشخص شود و هر زیر کلمه ابتدا با توجه به شیب آن چرخانده شده و سپس در مکان مناسب بر روی خط زمینه قرار گیرد.

جهت تعیین خط زمینه‌ی هر خط، با توجه به شکل اعوجاج، مناسب‌تر دیده شد تا بالاترین نقطه و پایین‌ترین نقطه‌ی هر خط تعیین گردد و خط زمینه را در وسط پایین‌ترین و بالاترین نقطه در نظر گرفته شود. در شکل ۳-۲۲ خط زمینه برای تک تک خطوط نشان داده شده است.



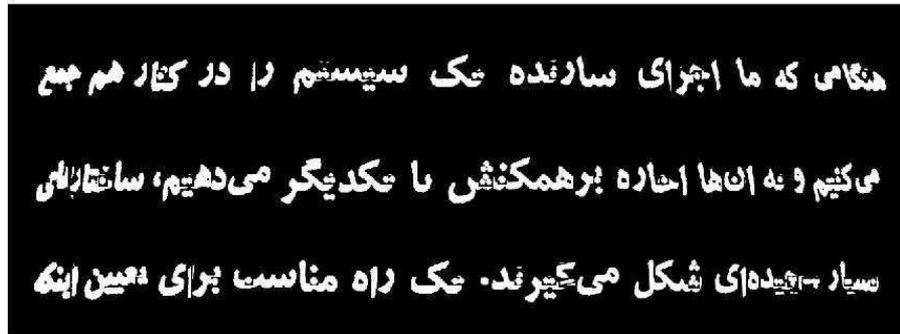
شکل ۳-۲۲: خط زمینه برای تک تک خطوط شکل ۳-۱۷

در شکل ۳-۲۳ شیب زیرکلمات، با توجه به روش یک تعیین و هر زیرکلمه ابتدا با توجه به شیب چرخانده شده و سپس بر روی خط زمینه قرار گرفته است.



شکل ۳-۲۳: تصویر چرخش یافته در حالتی که شیب از روش اول تعیین گردد

در شکل ۳-۲۴ نیز شیب زیر کلمات با توجه به روش دو تعیین و هر زیر کلمه ابتدا چرخانده شده و سپس بر روی خط زمینه قرار گرفته است.



شکل ۳-۲۴: تصویر چرخش یافته در حالتی که شیب از روش دوم تعیین گردد

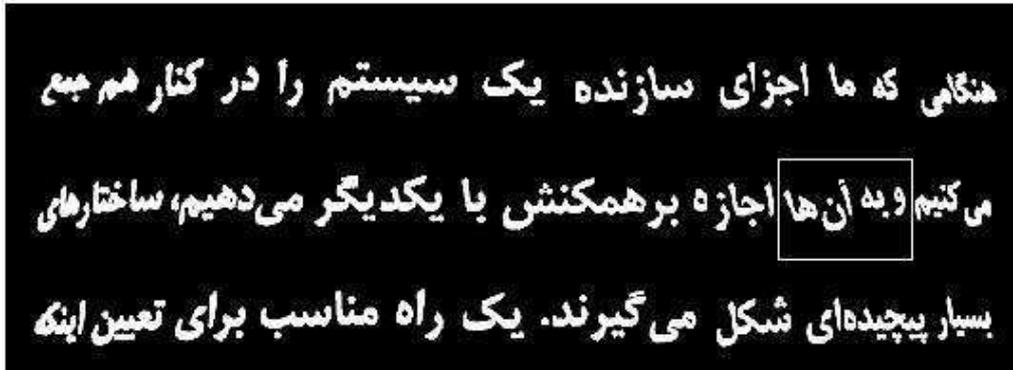
همانگونه که در شکل‌های ۳-۲۳ و ۳-۲۴ دیده می‌شود، تصویر خروجی خوانایی کافی را ندارد. نقاط در مکان مناسبی قرار ندارند و زیر کلمات مربوط به هر کلمه در جای درست خود نیستند. جهت رفع این مشکلات، ابتدا چند زیر کلمه‌ی متوالی با هم ترکیب شده و جعبه‌های کلمات شناسایی می‌گردند. برای این کار مستطیل محیطی در برگیرنده‌ی چند زیر کلمه را به مستطیل محیطی بزرگتر که در برگیرنده‌ی چند زیر کلمه است تبدیل می‌کنیم. در شکل ۳-۲۵ جعبه‌های زیر کلمات با هم ترکیب شده و برای هر سه خط تصویر نشان داده شده‌اند.



شکل ۳-۲۵: مستطیل‌های کلمات تشکیل شده از چند زیر کلمه



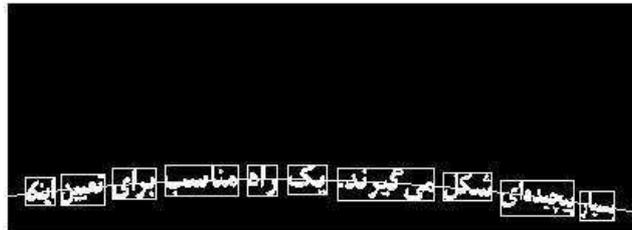
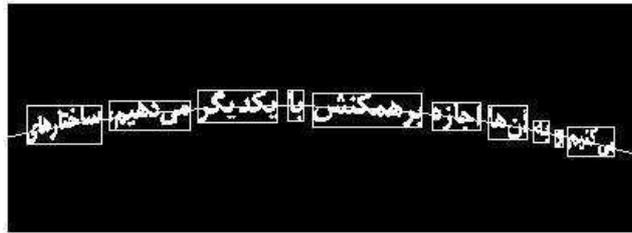
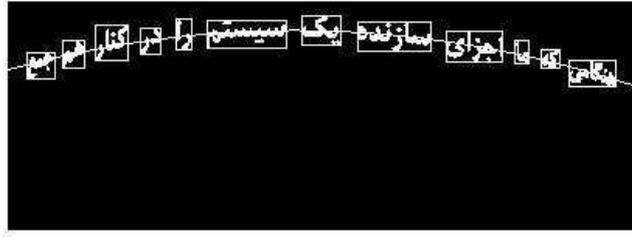
در شکل ۳-۲۷ ابتدا مستطیل‌های زیرکلمات با هم ترکیب شده و سپس میزان چرخش جعبه‌ها با استفاده از روش یک تعیین گردیده و هر جعبه ابتدا چرخانده شده و سپس در مکان مناسب بر روی خط زمینه قرار داده شده است.



شکل ۳-۲۷: تصویر چرخش یافته در حالتی که شیب از روش اول تعیین گردد

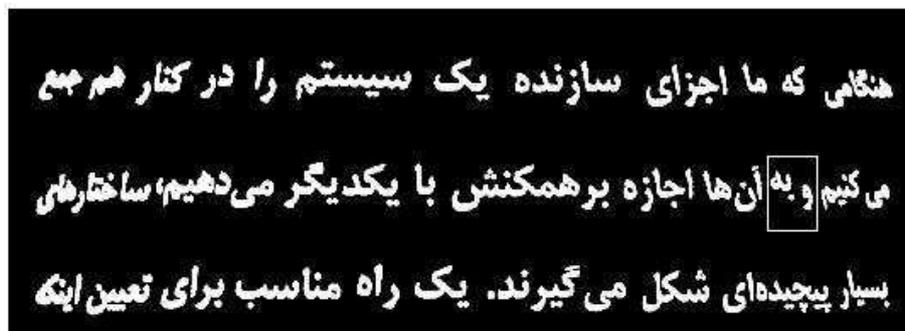
همانگونه که در شکل فوق دیده می‌شود با ترکیب جعبه‌ها تصویر خروجی بهبود چشمگیری پیدا می‌کند. و به جز بخش مشخص شده در شکل بقیه‌ی بخش‌ها در وضعیت مناسبی قرار گرفته‌اند و اعوجاج تصویر از بین رفته است.

در شکل فوق شیب کلمات، با توجه به روش یک تعیین گردید. جهت محاسبه شیب توسط روش دوم یعنی استفاده از برازش منحنی، یک چند جمله‌ای درجه‌ی دوم بر مراکز جعبه‌های ترکیب شده برازش می‌شود. منحنی‌های برازش شده برای تک‌تک خطوط تصویر نمونه در شکل ۳-۲۸ نشان داده شده است.



شکل ۲۸-۳: منحنی‌های برازش شده بر مراکز مستطیل‌های کلمات

همانگونه که در شکل‌های ۲۶-۳ و ۲۸-۳ دیده می‌شود نتایج حاصل از دو روش بیان شده جهت محاسبه میزان چرخش در حالتی که مستطیل‌های زیر کلمات ترکیب شده‌اند بسیار به یکدیگر شبیه است. در شکل ۲۹-۳ تصویر خروجی رفع اعوجاج شده برای حالتی که شیب کلمات با روش دو تعیین شده است، نشان داده شده است.



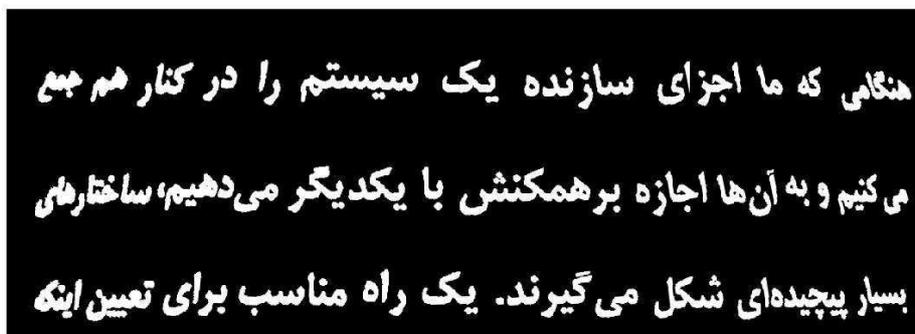
شکل ۲۹-۳: تصویر چرخش یافته در حالتی که شیب از روش برازش منحنی تعیین گردد

همان‌گونه که در شکل های ۳-۲۷ و ۳-۲۹ دیده می‌شود، روش دوم تعیین شیب مناسب تر بوده و تصویر خروجی مربوط به آن خوانایی بیشتری دارد. در شکل ۳-۲۹ اکثر بخش‌های متن به خوبی رفع اعوجاج گردیده‌اند. برای اینکه مشکل قسمت مشخص شده در شکل رفع گردد، باید مستطیل‌های زیر کلمات به نحوی با هم ترکیب شوند که علاوه بر اینکه تمامی زیرکلمات یک کلمه در یک مستطیل قرار می‌گیرند، کلمات کوچکی نظیر (و) و (به) نیز با هم در یک مستطیل قرار گیرند. برای این کار یک مقدار آستانه بر روی عرض مستطیل‌های اولیه اعمال می‌گردد تا دو کلمه کوچک متوالی نیز با هم در مستطیل قرار گیرند. در شکل مستطیل‌های ترکیب شده خط دوم متن پس از اعمال آستانه بر روی مستطیل‌ها نشان داده شده است. همانگونه که در شکل ۳-۳۰ دیده می‌شود با اعمال مقدار آستانه مناسب برای عرض مستطیل‌های متوالی و فاصله آنها دو کلمه (و) و (به) در یک مستطیل جای گرفته‌اند.



شکل ۳-۳۰: مستطیل‌های ترکیب شده بعد از آستانه‌گذاری بر روی عرض مستطیل‌ها

تصویر خروجی حاصل از آستانه‌گذاری بر روی عرض مستطیل‌ها در شکل ۳-۳۱ نشان داده شده است.



شکل ۳-۳۱: تصویر رفع اعوجاج شده پس از اعمال آستانه بر روی عرض مستطیل‌ها

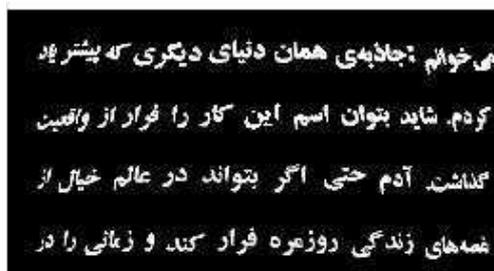
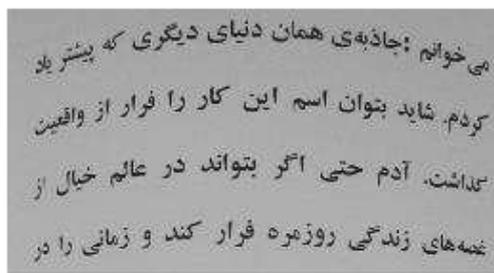
همانگونه که در شکل ۳-۳۱ دیده می‌شود با اعمال مقدار آستانه بر روی عرض مستطیل‌ها، تصویر خروجی بهبود یافته است و چرخش سه بُعدی حاصل برطرف شده است.

### ۳-۸ ارائه تصویر خروجی

در بخش‌های قبلی مراحل مختلف رفع اعوجاج تصویر متن بیان گردید. جهت از بین بردن چرخش سه بُعدی در تصاویر متون فارسی با اعوجاجات مختلف و داشتن تصاویر خروجی مناسب علاوه بر مواردی که در بخش‌های قبل بیان شد باید موارد زیر نیز در نظر گرفته شود.

### ۳-۸-۱ رفع اعوجاج تصویر اولیه

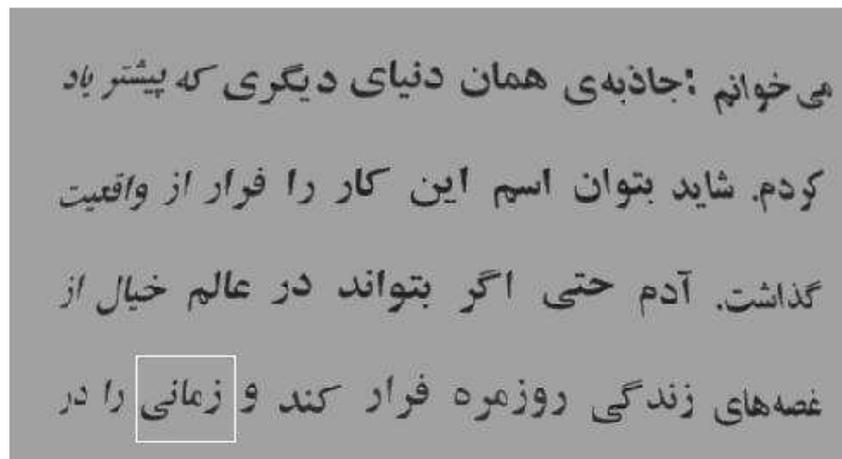
در بخش‌های قبلی ابتدا تصویر به یک تصویر دوسطحی تبدیل می‌گردید و سپس عملیات رفع چرخش سه بُعدی بر روی تصویر دوسطحی شده انجام می‌شد. در مورد تصاویری که وضوح تصویر کم است خروجی دودویی کیفیت خوبی ندارد. در شکل زیر یک تصویر با وضوح تصویر کم توسط الگوریتم ارائه شده رفع اعوجاج گردیده است. تصویر خروجی نیز در شکل ۳-۳۲ نشان داده شده است.



شکل ۳-۳۲: تصویر اولیه (بالایی) و تصویر رفع چرخش شده (پایینی)

جهت داشتن تصویر با کیفیت بهتر مناسب‌تر است که در مرحله جایگذاری بخش‌های رفع چرخش شده بر روی خط زمینه، مستطیل‌ها به جای تصویر دوسطحی شده از تصویر اولیه خاکستری انتخاب شوند. برای این کار باید زمینه تصویر خروجی نیز متناسب با تصویر خاکستری ورودی انتخاب گردد.

در شکل زیر خروجی تصویر قبلی بر روی یک زمینه خاکستری نشان داده شده است. همانگونه که در شکل ۳-۳۳ دیده می‌شود، تصویر خروجی در این حالت خوانایی و کیفیت بهتری نسبت به شکل ۳-۳۲ دارد.



شکل ۳-۳۳: تصویر خروجی شکل ۳-۳۲ با زمینه خاکستری

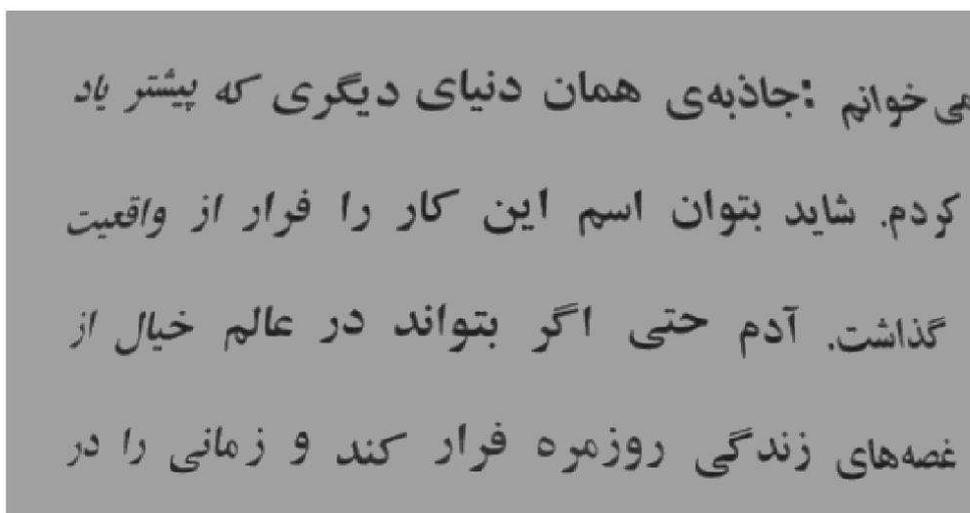
بنابراین می‌توان گفت جهت داشتن خروجی با زمینه خاکستری باید تمامی مراحل گفته شده در بخش‌های قبلی را به ترتیب انجام داده و پس از تعیین مختصات مستطیل‌های ترکیب شده، مستطیل‌ها را به جای تصویر دوسطحی شده از تصویر اولیه خاکستری انتخاب نمود. مستطیل‌ها پس از چرخش در یک زمینه خاکستری بر روی خطوط زمینه و در مکان مناسب قرار داده می‌شوند.

### ۳-۸-۲ رفع تداخل ایجاد شده در هنگام چرخش مستطیل‌ها

در تصاویری که میزان اعوجاج تصویر زیاد است، بعد از رفع چرخش و قرارگیری کلمات بر روی خط زمینه، به این دلیل که پس از چرخش اندازه مستطیل‌ها تغییر می‌کند، کلمات قرار گرفته بر روی خط زمینه با یکدیگر تداخل می‌یا بند. حتی در شکل ۳-۳۲ نیز که اعوجاج تصویر زیاد نیست در خط آخر، کلمه (زمانی)

دچار اختلال شده است. جهت رفع این مشکل عرض تصویر خروجی مقداری بیشتر از تصویر ورودی در نظر گرفته می‌شود تا بعد از قرارگیری کلمات چرخش یافته بر روی خط زمینه تداخلی ایجاد نگردد.

در شکل ۳-۳۳ با اینکه اعوجاج متن به خوبی رفع شده است، کلمه (زمانی) که در شکل مشخص گردیده به صورت ناقص بوده و دچار اختلال شده است. جهت رفع این تداخل، عرض تصویر خروجی مقداری بیشتر از تصویر ورودی در نظر گرفته می‌شود و با توجه به اندازه جدید یک مستطیل بعد از چرخش، مستطیل بعدی مقداری شیفت یابد تا تداخلی بین کلمات ایجاد نگردد. نتیجه حاصل از این کار بر روی تصویر نمونه قبلی در شکل ۳-۳۴ نشان داده شده است.



شکل ۳-۳۴: تصویر خروجی شکل ۳-۳۳ در حالتی که رفع تداخل گردیده است.

همانگونه که در شکل فوق دیده می‌شود با اعمال روش بیان شده در این بخش تداخل ایجاد شده رفع می‌گردد.

### ۳-۹ نتیجه‌گیری

در این فصل که می‌توان آن را فصل اصلی این پایان‌نامه نامید، الگوریتم پیشنهادی جهت رفع چرخش سه بُعدی تصاویر متون فارسی بیان شد. ابتدا فلوجارت کلی الگوریتم پیشنهادی نشان داده شد و سپس تک‌تک بخش‌های فلوجارت مورد بررسی قرار گرفت.

در بخش‌های مختلف این فصل چگونگی رفع اعوجاج بر روی تصاویر مختلف نشان داده شد. در فصل بعد نتایج تجربی حاصل از الگوریتم پیشنهادی بر روی تصاویر دیگر و با قلم‌های مختلف بررسی می‌شود.

فصل چہارم

نتایج تجربی

## مقدمه

در فصل قبل بخش‌های مختلف الگوریتم پیشنهادی به طور کامل مورد بررسی قرار گرفت و نمونه‌هایی از نتایج الگوریتم بر روی تصویرهای دارای اعوجاج نشان داده شد. در این فصل، عملکرد الگوریتم پیشنهادی بر روی نمونه‌های بیشتری از مجموعه داده تهیه شده و بر روی چند تصویر با قلم‌های مختلف بررسی می‌شود. همان‌گونه که در فصل قبل بیان شد، مجموعه داده ساخته شده شامل تصاویر اسناد با قلم‌هایی است که توسط نرم‌افزار نویسه خوان پرشیانگار قابل شناسایی هستند.

پرشیانگار، یک نرم‌افزار تجاری OCR فارسی است که قادر است تصاویر اسناد فارسی که شامل ۱۰ قلم متداول هستند را به متون قابل ویرایش تبدیل کند. این نرم‌افزار سال ۱۳۸۷ تولید شده است و ادعا شده که برای قلم‌های مذکور و درجه تفکیک ۳۰۰ نقطه بر اینچ، دقتی حدود ۹۷٪ دارد. قلم‌هایی که این نرم‌افزار پشتیبانی می‌کند عبارتند از نازنین، میترا، لوتوس، زر، ترافیک، یاقوت، هما، تیترا، تایمز و تاهما [۹۱]

در این فصل همچنین جهت سنجیدن صحت عملکرد الگوریتم پیشنهادی، تصاویر رفع اعوجاج شده را توسط نرم‌افزار پرشیانگار ارزیابی می‌کنیم تا مشخص گردد که الگوریتم پیشنهادی چه میزان می‌تواند در صحت عملکرد این نویسه خوان فارسی مؤثر واقع گردد.

## ۴-۱- نتایج تجربی الگوریتم پیشنهادی بر روی تصاویر نمونه

همان‌طور که گفته شد تصاویر نمونه از مجموعه داده تهیه شده انتخاب می‌گردند. شکل‌های ۴-۱ تا ۴-۵ هرکدام شامل ۴ تصویر است که در هر کدام، (الف): تصاویر اولیه دارای اعوجاج از مجموعه داده هستند. (ب): دو سطحی شده‌ی تصاویر ورودی می‌باشند. (ج) و (د) به ترتیب تصویر رفع اعوجاج شده دودویی و زمینه خاکستری هستند. تصاویر ورودی به ترتیب از قلم‌های traffic، nazanin، yagut، zar و Tahoma می‌باشند. تمامی مراحل الگوریتم پیشنهادی که در فصل قبل بیان شد روی این تصاویر ورودی پیاده سازی شده است.

(ب)

جلسات خیاطی را اداره می کرد، به برپایی کلاس هایی که روزهای یکشنبه تشکیل می شد، کمک زیادی می کرد و بهترین عضو جامعه ی کمک به کلیسا و دستیار هیئت مبلغان خارجی بود. اما با وجود همه ی آن کارها، خانم ریچل فرصت فراوانی داشت تا ساعت ها پشت پنجره ی آشپزخانه بنشیند و درحال بافتن لحاف های پنبه ای- طبق حرف هایی که خانم های خانه دار اوئی می زدند، او تا آن زمان کار بافتن شانزده عدد از آن ها را به اتمام رسانده بود. با چشمان تیزبینش، به دقت جاده اصلی اوئی را که از میان گودال می گذشت و شیب تپه ی قرمز را پشت سر می گذاشت، زیر نظر بگیرد. اوئی شامل شبه جزیره کوچک مثلثی شکلی بود که از میان خلیج سن لارنس سر بر آورده و دو طرفش را آب فرا گرفته بود. بنابراین هر کس می خواست وارد آن شود یا از آن بیرون برود، باید از جاده ی تپه ای عبور می کرد به این ترتیب

(الف)

جلسات خیاطی را اداره می کرد، به برپایی کلاس هایی که روزهای یکشنبه تشکیل می شد، کمک زیادی می کرد و بهترین عضو جامعه ی کمک به کلیسا و دستیار هیئت مبلغان خارجی بود. اما با وجود همه ی آن کارها، خانم ریچل فرصت فراوانی داشت تا ساعت ها پشت پنجره ی آشپزخانه بنشیند و درحال بافتن لحاف های پنبه ای- طبق حرف هایی که خانم های خانه دار اوئی می زدند، او تا آن زمان کار بافتن شانزده عدد از آن ها را به اتمام رسانده بود. با چشمان تیزبینش، به دقت جاده اصلی اوئی را که از میان گودال می گذشت و شیب تپه ی قرمز را پشت سر می گذاشت، زیر نظر بگیرد. اوئی شامل شبه جزیره کوچک مثلثی شکلی بود که از میان خلیج سن لارنس سر بر آورده و دو طرفش را آب فرا گرفته بود. بنابراین هر کس می خواست وارد آن شود یا از آن بیرون برود، باید از جاده ی تپه ای عبور می کرد به این ترتیب

(ج)

جلسات خیاطی را اداره می کرد، به برپایی کلاس هایی که روزهای یکشنبه تشکیل می شد، کمک زیادی می کرد و بهترین عضو جامعه ی کمک به کلیسا و دستیار هیئت مبلغان خارجی بود. اما با وجود همه ی آن کارها، خانم ریچل فرصت فراوانی داشت تا ساعت ها پشت پنجره ی آشپزخانه بنشیند و درحال بافتن لحاف های پنبه ای- طبق حرف هایی که خانم های خانه دار اوئی می زدند، او تا آن زمان کار بافتن شانزده عدد از آن ها را به اتمام رسانده بود. با چشمان تیزبینش، به دقت جاده اصلی اوئی را که از میان گودال می گذشت و شیب تپه ی قرمز را پشت سر می گذاشت، زیر نظر بگیرد. اوئی شامل شبه جزیره کوچک مثلثی شکلی بود که از میان خلیج سن لارنس سر بر آورده و دو طرفش را آب فرا گرفته بود. بنابراین هر کس می خواست وارد آن شود یا از آن بیرون برود، باید از جاده ی تپه ای عبور می کرد به این ترتیب

(د)

جلسات خیاطی را اداره می کرد، به برپایی کلاس هایی که روزهای یکشنبه تشکیل می شد، کمک زیادی می کرد و بهترین عضو جامعه ی کمک به کلیسا و دستیار هیئت مبلغان خارجی بود. اما با وجود همه ی آن کارها، خانم ریچل فرصت فراوانی داشت تا ساعت ها پشت پنجره ی آشپزخانه بنشیند و درحال بافتن لحاف های پنبه ای- طبق حرف هایی که خانم های خانه دار اوئی می زدند، او تا آن زمان کار بافتن شانزده عدد از آن ها را به اتمام رسانده بود. با چشمان تیزبینش، به دقت جاده اصلی اوئی را که از میان گودال می گذشت و شیب تپه ی قرمز را پشت سر می گذاشت، زیر نظر بگیرد. اوئی شامل شبه جزیره کوچک مثلثی شکلی بود که از میان خلیج سن لارنس سر بر آورده و دو طرفش را آب فرا گرفته بود. بنابراین هر کس می خواست وارد آن شود یا از آن بیرون برود، باید از جاده ی تپه ای عبور می کرد به این ترتیب

شکل ۴-۱: (الف) تصویر ورودی، (ب) دوسطحی شده تصویر ورودی (ج) تصویر خروجی دوسطحی (د) تصویر خروجی خاکستری

(الف)

هنگام عروسی محمدرضا شاه پهلوی و زن مصری اش فوزیه در سال ۱۳۱۷ خورشیدی چون مقرر بود، میهمانان مصری و همراهان عروس به وسیله راه آهن جنوب به تهران وارد شوند از طرف دربار و شهربانی دستور اکید صادر شده بود که دیوارهای تمام دهات طول راه و خانه های دهقانی مجاور خط آهن را سفید کنند.

در یکی از دهات چون گچ در دسترس نبود بخشدار دستور می دهد که با کشک و ماست که در آن ده فراوان بوده دیوارها را موقتاً سفید نمایند، و به این منظور با پولی که از کدخدای ده می گیرند و با خرید مقدار زیادی ماست کلیه دیوارها را ماست مالی کردند.

قدمت تاریخی این اصطلاح ماست مالی از شصت سال نمی گذرد، و اجرای این ماست مالی مدت ها موضوع اصلی شوخی های محافل و مجالس بود. حکیم ارد بزرگ می گوید: فرمانروایان تنها پاسخگوی زمان حال خویش نیستند آنها به گذشتگان و آیندگان نیز پاسخگو هستند.

(د)

هنگام عروسی محمدرضا شاه پهلوی و زن مصری اش فوزیه در سال ۱۳۱۷ خورشیدی چون مقرر بود، میهمانان مصری و همراهان عروس به وسیله راه آهن جنوب به تهران وارد شوند از طرف دربار و شهربانی دستور اکید صادر شده بود که دیوارهای تمام دهات طول راه و خانه های دهقانی مجاور خط آهن را سفید کنند.

در یکی از دهات چون گچ در دسترس نبود بخشدار دستور می دهد که با کشک و ماست که در آن ده فراوان بوده دیوارها را موقتاً سفید نمایند، و به این منظور با پولی که از کدخدای ده می گیرند و با خرید مقدار زیادی ماست کلیه دیوارها را ماست مالی کردند.

قدمت تاریخی این اصطلاح ماست مالی از شصت سال نمی گذرد، و اجرای این ماست مالی مدت ها موضوع اصلی شوخی های محافل و مجالس بود. حکیم ارد بزرگ می گوید: فرمانروایان تنها پاسخگوی زمان حال خویش نیستند آنها به گذشتگان و آیندگان نیز پاسخگو هستند.

(ب)

هنگام عروسی محمدرضا شاه پهلوی و زن مصری اش فوزیه در سال ۱۳۱۷ خورشیدی چون مقرر بود، میهمانان مصری و همراهان عروس به وسیله راه آهن جنوب به تهران وارد شوند از طرف دربار و شهربانی دستور اکید صادر شده بود که دیوارهای تمام دهات طول راه و خانه های دهقانی مجاور خط آهن را سفید کنند.

در یکی از دهات چون گچ در دسترس نبود بخشدار دستور می دهد که با کشک و ماست که در آن ده فراوان بوده دیوارها را موقتاً سفید نمایند، و به این منظور با پولی که از کدخدای ده می گیرند و با خرید مقدار زیادی ماست کلیه دیوارها را ماست مالی کردند.

قدمت تاریخی این اصطلاح ماست مالی از شصت سال نمی گذرد، و اجرای این ماست مالی مدت ها موضوع اصلی شوخی های محافل و مجالس بود. حکیم ارد بزرگ می گوید: فرمانروایان تنها پاسخگوی زمان حال خویش نیستند آنها به گذشتگان و آیندگان نیز پاسخگو هستند.

(ج)

هنگام عروسی محمدرضا شاه پهلوی و زن مصری اش فوزیه در سال ۱۳۱۷ خورشیدی چون مقرر بود، میهمانان مصری و همراهان عروس به وسیله راه آهن جنوب به تهران وارد شوند از طرف دربار و شهربانی دستور اکید صادر شده بود که دیوارهای تمام دهات طول راه و خانه های دهقانی مجاور خط آهن را سفید کنند.

در یکی از دهات چون گچ در دسترس نبود بخشدار دستور می دهد که با کشک و ماست که در آن ده فراوان بوده دیوارها را موقتاً سفید نمایند، و به این منظور با پولی که از کدخدای ده می گیرند و با خرید مقدار زیادی ماست کلیه دیوارها را ماست مالی کردند.

قدمت تاریخی این اصطلاح ماست مالی از شصت سال نمی گذرد، و اجرای این ماست مالی مدت ها موضوع اصلی شوخی های محافل و مجالس بود. حکیم ارد بزرگ می گوید: فرمانروایان تنها پاسخگوی زمان حال خویش نیستند آنها به گذشتگان و آیندگان نیز پاسخگو هستند.

شکل ۴-۲: (الف) تصویر ورودی، (ب) دوسطحی شده تصویر ورودی (ج) تصویر خروجی دوسطحی (د) تصویر خروجی خاکستری

(الف)

سراسر دریای مدیترانه، و تا بریتانیا در شمال با منابع قلع برای تولید برنز سفر می‌کردند. برای این منظور آنها مستعمرات یونانی تجارت به نام مرکز فروش را تاسیس کردند. از آغاز تمدن یونانی تا سقوط امپراطوری روم در قرن پنجم، تجارت ادویه ارزشمند به اروپا از شرق دور، تجارت پر سود، از جمله چین و هند است. امپراطوری روم به رشد و افزایش تجارت اجازه داد. امپراطوری روم، شبکه با ثبات و امن حمل و نقل را فعال کرد کالاهای تجاری بدون ترس از دزدی دریایی حمل و نقل می‌کردند. سقوط امپراطوری روم و قرون وسطی بی ثباتی برای اروپای غربی و فروپاشی شبکه‌های تجاری در جهان غرب را به ارمغان آورد. تجارت با این حال ادامه یافت: و در میان پادشاهی آفریقا، شرق میانه، هند، چین و جنوب شرق آسیا شکوفا شد. تجارت برخی در غرب رخ داد. به عنوان مثال، رادانیست‌های قرون وسطی و با گروهی به معنای دقیق کلمه به تاریخ از دست رفته از بازرگانان

(ب)

سراسر دریای مدیترانه، و تا بریتانیا در شمال با منابع قلع برای تولید برنز سفر می‌کردند. برای این منظور آنها مستعمرات یونانی تجارت به نام مرکز فروش را تاسیس کردند. از آغاز تمدن یونانی تا سقوط امپراطوری روم در قرن پنجم، تجارت ادویه ارزشمند به اروپا از شرق دور، تجارت پر سود، از جمله چین و هند است. امپراطوری روم به رشد و افزایش تجارت اجازه داد. امپراطوری روم، شبکه با ثبات و امن حمل و نقل را فعال کرد کالاهای تجاری بدون ترس از دزدی دریایی حمل و نقل می‌کردند. سقوط امپراطوری روم و قرون وسطی بی ثباتی برای اروپای غربی و فروپاشی شبکه‌های تجاری در جهان غرب را به ارمغان آورد. تجارت با این حال ادامه یافت: و در میان پادشاهی آفریقا، شرق میانه، هند، چین و جنوب شرق آسیا شکوفا شد. تجارت برخی در غرب رخ داد. به عنوان مثال، رادانیست‌های قرون وسطی و با گروهی به معنای دقیق کلمه به تاریخ از دست رفته از بازرگانان

(د)

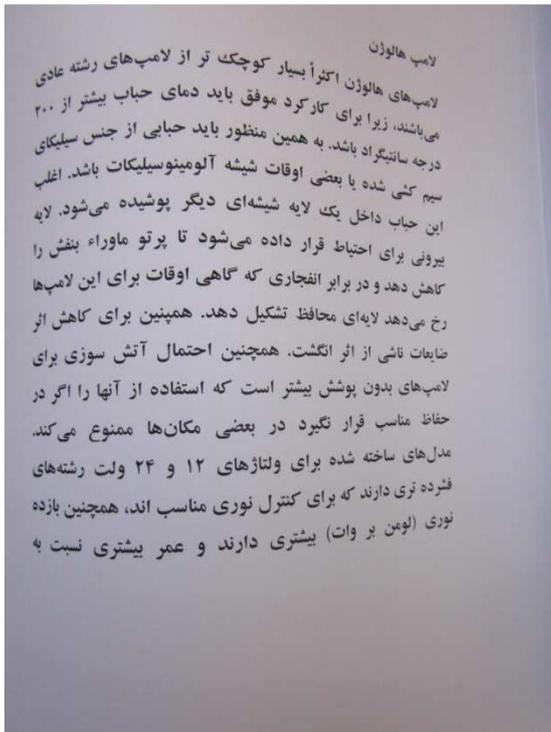
سراسر دریای مدیترانه، و تا بریتانیا در شمال با منابع قلع برای تولید برنز سفر می‌کردند. برای این منظور آنها مستعمرات یونانی تجارت به نام مرکز فروش را تاسیس کردند. از آغاز تمدن یونانی تا سقوط امپراطوری روم در قرن پنجم، تجارت ادویه ارزشمند به اروپا از شرق دور، تجارت پر سود، از جمله چین و هند است. امپراطوری روم به رشد و افزایش تجارت اجازه داد. امپراطوری روم، شبکه با ثبات و امن حمل و نقل را فعال کرد کالاهای تجاری بدون ترس از دزدی دریایی حمل و نقل می‌کردند. سقوط امپراطوری روم و قرون وسطی بی ثباتی برای اروپای غربی و فروپاشی شبکه‌های تجاری در جهان غرب را به ارمغان آورد. تجارت با این حال ادامه یافت: و در میان پادشاهی آفریقا، شرق میانه، هند، چین و جنوب شرق آسیا شکوفا شد. تجارت برخی در غرب رخ داد. به عنوان مثال، رادانیست‌های قرون وسطی و با گروهی به معنای دقیق کلمه به تاریخ از دست رفته از بازرگانان

(ج)

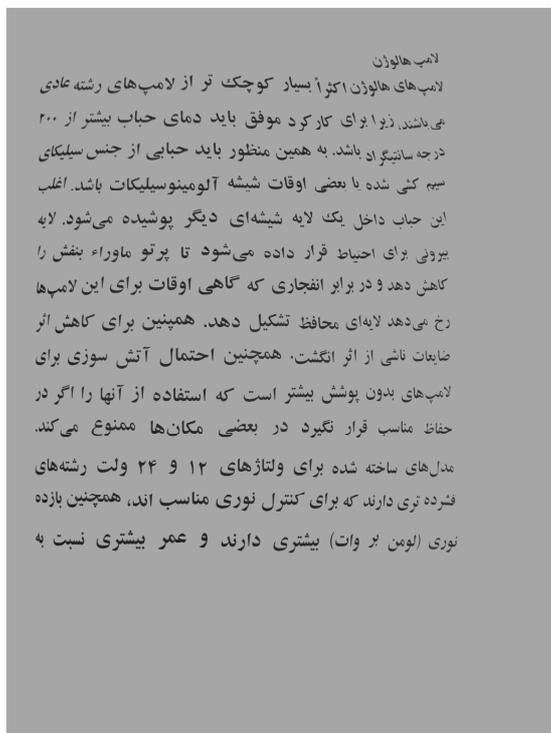
سراسر دریای مدیترانه، و تا بریتانیا در شمال با منابع قلع برای تولید برنز سفر می‌کردند. برای این منظور آنها مستعمرات یونانی تجارت به نام مرکز فروش را تاسیس کردند. از آغاز تمدن یونانی تا سقوط امپراطوری روم در قرن پنجم، تجارت ادویه ارزشمند به اروپا از شرق دور، تجارت پر سود، از جمله چین و هند است. امپراطوری روم به رشد و افزایش تجارت اجازه داد. امپراطوری روم، شبکه با ثبات و امن حمل و نقل را فعال کرد کالاهای تجاری بدون ترس از دزدی دریایی حمل و نقل می‌کردند. سقوط امپراطوری روم و قرون وسطی بی ثباتی برای اروپای غربی و فروپاشی شبکه‌های تجاری در جهان غرب را به ارمغان آورد. تجارت با این حال ادامه یافت: و در میان پادشاهی آفریقا، شرق میانه، هند، چین و جنوب شرق آسیا شکوفا شد. تجارت برخی در غرب رخ داد. به عنوان مثال، رادانیست‌های قرون وسطی و با گروهی به معنای دقیق کلمه به تاریخ از دست رفته از بازرگانان

شکل ۴-۳: (الف) تصویر ورودی، (ب) دوسطحی شده تصویر ورودی (ج) تصویر خروجی دوسطحی (د) تصویر خروجی خاکستری

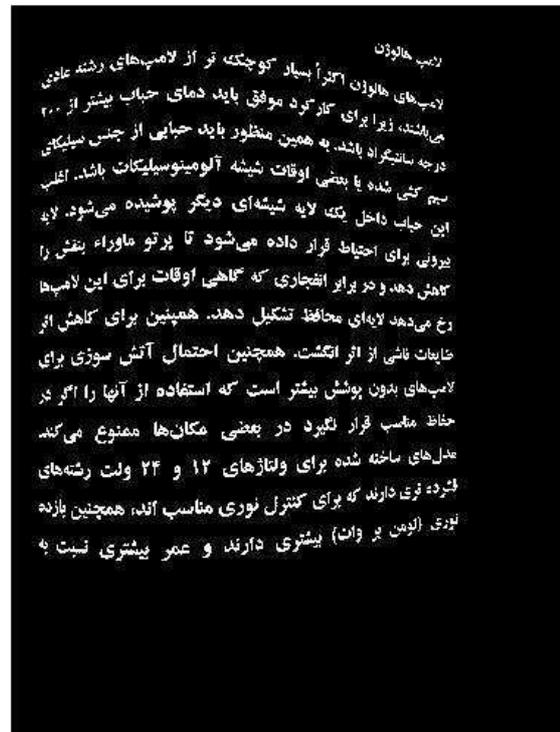
(الف)



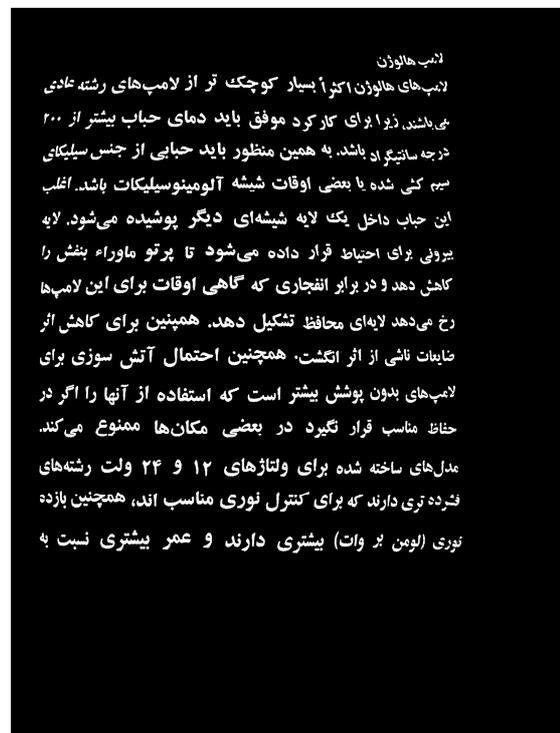
(د)



(ب)

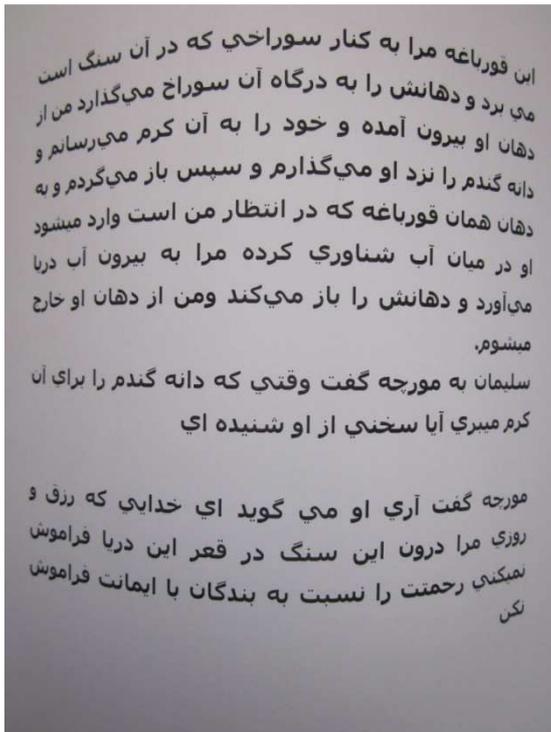


(ج)

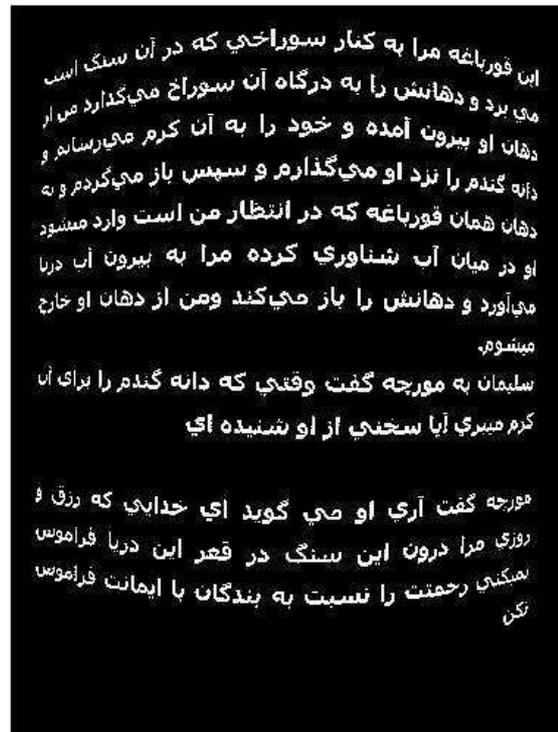


شکل ۴-۴: (الف) تصویر ورودی، (ب) دوسطحی شده تصویر ورودی (ج) تصویر خروجی دوسطحی (د) تصویر خروجی خاکستری

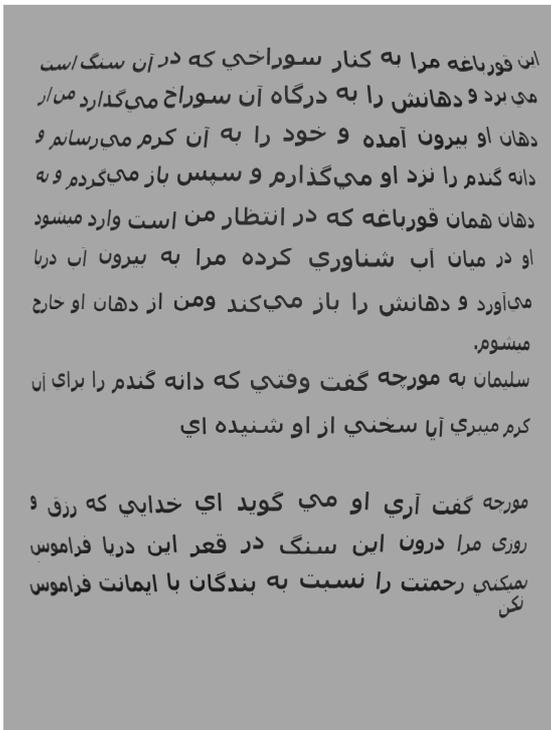
(الف)



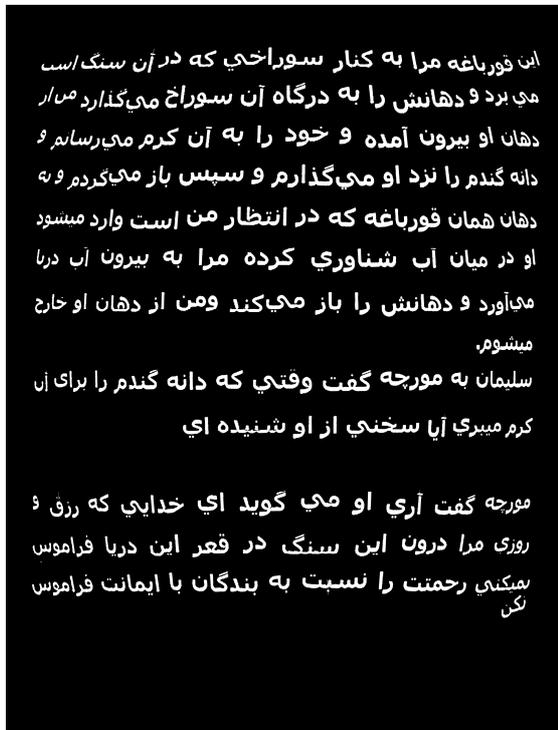
(ب)



(د)



(ج)

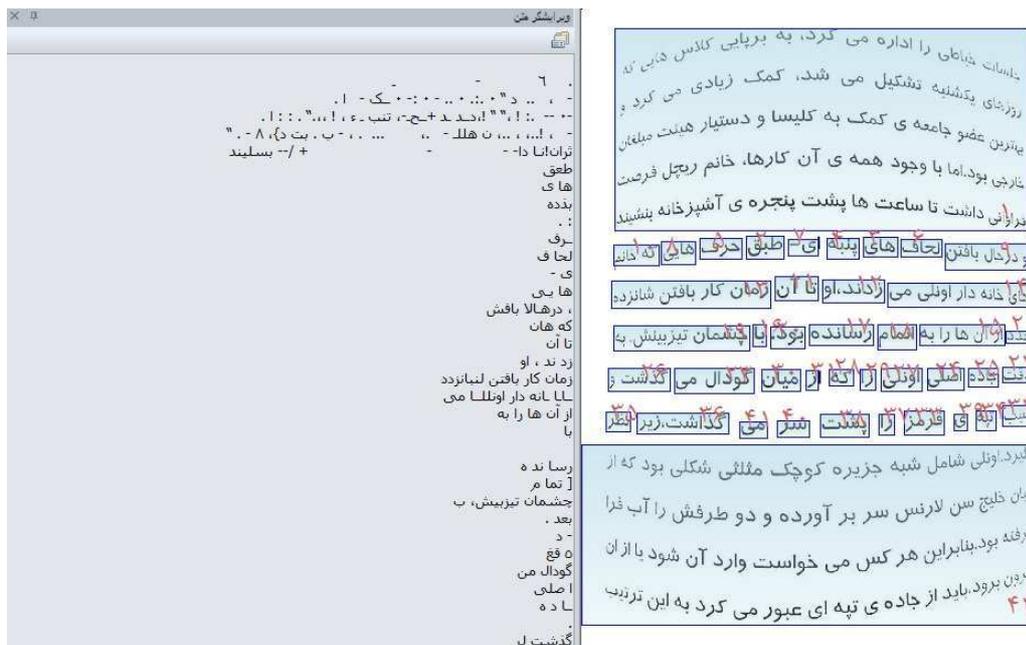


شکل ۴-۵: (الف) تصویر ورودی، (ب) دوسطحی شده تصویر ورودی (ج) تصویر خروجی دوسطحی (د) تصویر خروجی خاکستری

#### ۲-۴ بررسی عملکرد الگوریتم پیشنهادی توسط نرم افزار پرشیا نگار

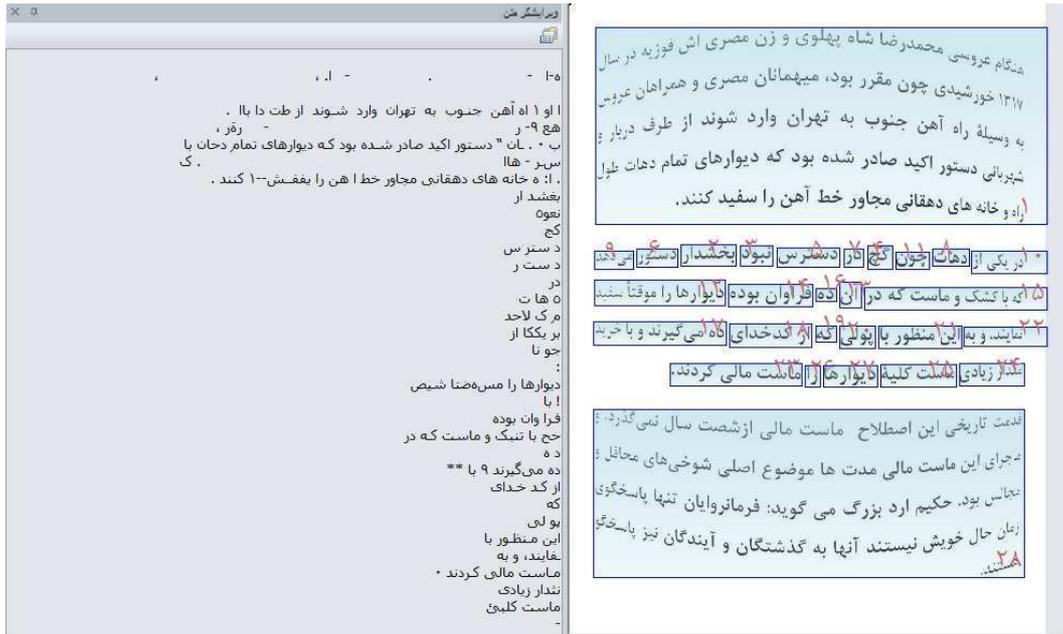
یکی از اهداف اصلی این پایان نامه رفع اعوجاجات غیرخطی اسناد فارسی است تا تصویر بعد از رفع اعوجاج با دقت خوبی توسط نرم افزارهای نویسه خوان نوری قابل خواندن باشد، از این رو جهت بررسی اینکه چقدر به هدف رسیده ایم، تصاویر خروجی الگوریتم توسط نرم افزار نویسه خوان فارسی پرشیا نگار مورد ارزیابی قرار گرفته است.

دو تصویر به عنوان ورودی به نرم افزار پرشیا نگار داده می شود. یکی تصویر اولیه که دارای چرخش سه بعدی است و تصویر دیگر تصویر نهایی رفع اعوجاج شده که نتیجه پیاده سازی الگوریتم پیشنهادی روی تصویر دارای اعوجاج است. شکل های ۴-۶ تا ۴-۱۰ هر کدام شامل ۲ تصویر است که در هر کدام، (الف): نتیجه خواندن تصاویر دارای اعوجاج در پرشیا نگار (ب): نتیجه خواندن تصاویر رفع اعوجاج شده توسط الگوریتم پیشنهادی در پرشیا نگار است.



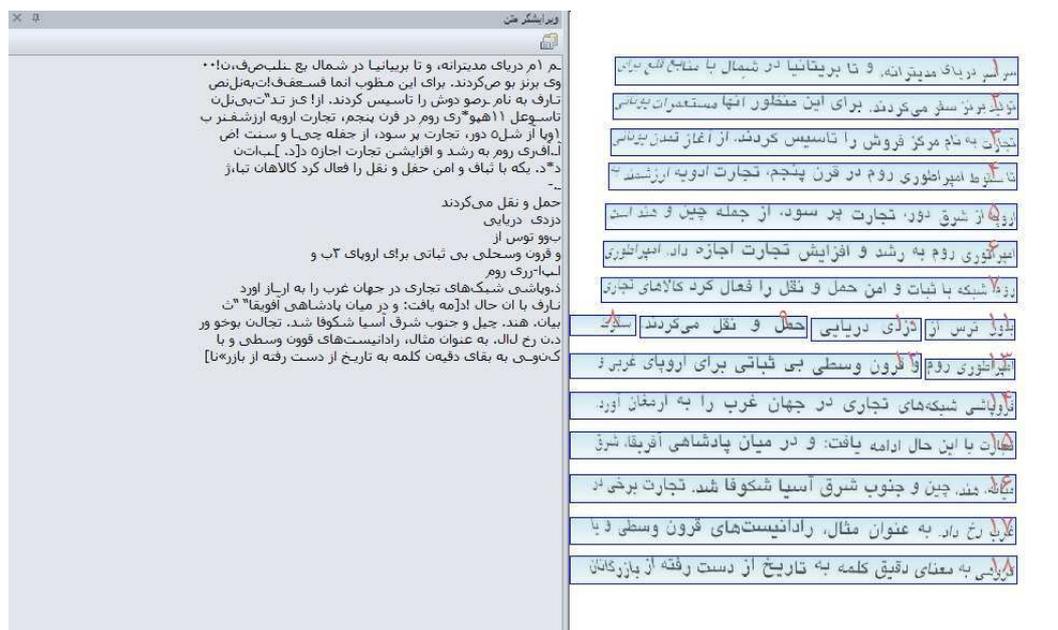
شکل ۴-۶: نتیجه پرشیا نگار برای تصویر دارای اعوجاج traffic (بالا) نتیجه پرشیا نگار برای تصویر بدون اعوجاج (پایین)

در شکل فوق نتیجه حاصل از OCR بر روی دو تصویر اعوجاج دار و بدون اعوجاج نشان داده شده است. همانگونه که در شکل دیده می شود بعد از رفع اعوجاج تشخیص خطوط متن توسط نرم افزار پرشیا نگار بسیار بهتر انجام می پذیرد و می توان گفت که برای قلم traffic، الگوریتم پیشنهادی عملکرد قابل قبولی داشته و تصویر رفع اعوجاج شده توسط الگوریتم پیشنهادی میزان دقت نویسه خوان را افزایش می دهد.



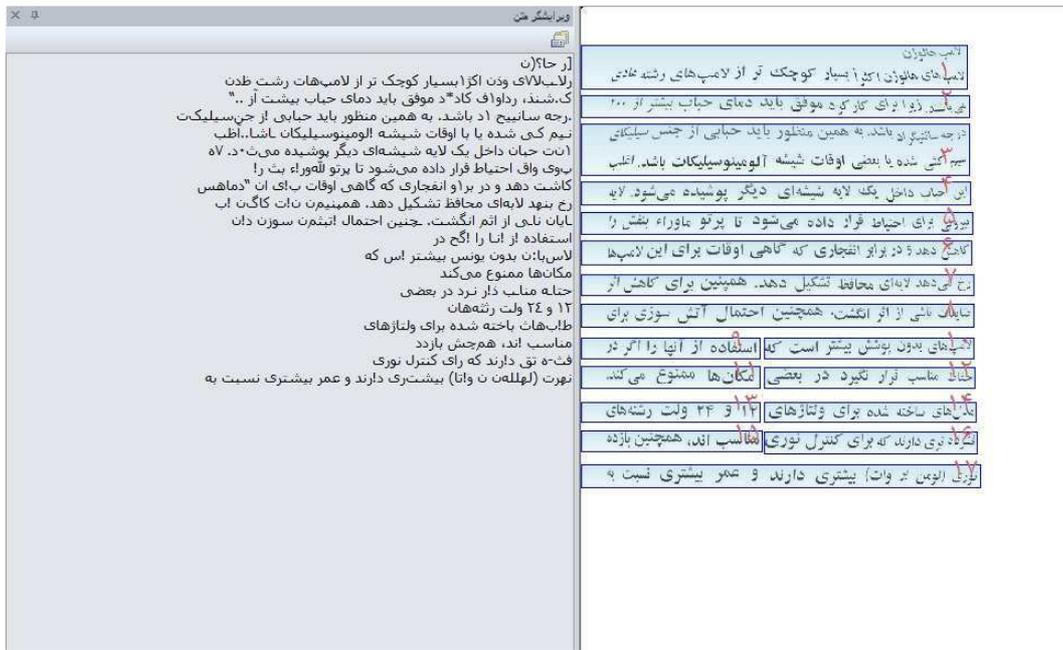
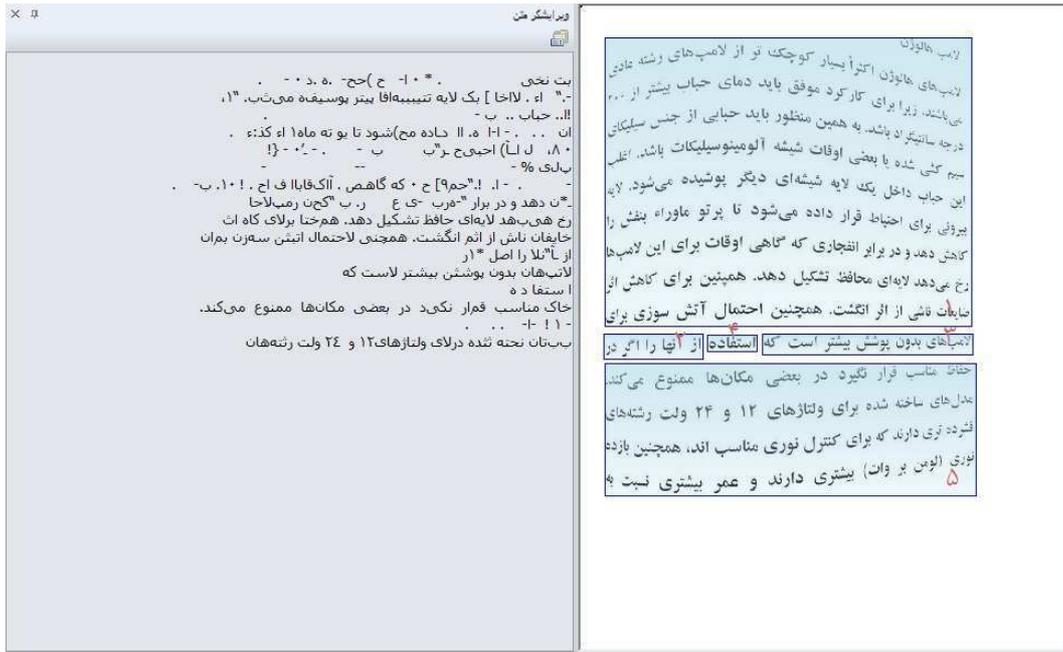
شکل ۴-۷: نتیجه پرشیا نگار برای تصویر دارای اعوجاج (بالا) نتیجه پرشیا نگار برای تصویر بدون اعوجاج (پایین)

در شکل فوق نتیجه حاصل از نرم افزار پرشین نگار بر روی دو تصویر دارای اعوجاج و بدون اعوجاج با قلم nazanin نشان داده شده است. همانگونه که در شکل دیده می شود نتیجه حاصل برای قلم nazanin نیز تاحد زیادی مشابه قلم traffic است و الگوریتم پیشنهادی در اینجا نیز عملکرد قابل قبولی داشته و تصویر رفع اعوجاج شده توسط الگوریتم پیشنهادی میزان دقت نویسه خوان را افزایش می دهد.



شکل ۴-۸: نتیجه پرسیا نگار برای تصویر دارای اعوجاج yagut (بالا) نتیجه پرسیا نگار برای تصویر رفع اعوجاج شده (پایین)

در شکل فوق نتیجه حاصل از OCR بر روی دو تصویر دارای اعوجاج و رفع اعوجاج شده با قلم yagut نشان داده شده است. برای قلم yagut نیز الگوریتم پیشنهادی خوانایی تصویر را برای نرم‌افزار پرسیانگار افزایش داده است، اما در جداسازی خطوط به خوبی دو قلم قبلی یعنی قلم‌های traffic و nazanin انجام نشده است.



شکل ۴-۹: نتیجه پرشیا نگار برای تصویر دارای اعوجاج zar (بالا) نتیجه پرشیا نگار برای تصویر رفع اعوجاج شده (پایین)

در شکل فوق نتیجه حاصل از نرم افزار پرشیا نگار بر روی دو تصویر دارای اعوجاج و رفع اعوجاج شده قلم zar نشان داده شده است. همانگونه که در شکل دیده می شود نتیجه حاصل برای قلم zar نیز تا حد زیادی مشابه قلم yagut است. در اینجا نیز الگوریتم پیشنهادی خوانایی تصویر را برای نرم افزار پرشیا نگار افزایش داده است، اما در جداسازی خطوط به خوبی قلم های traffic و nazanin انجام نشده است.



شکل ۴-۱۰: نتیجه پرشیا نگار برای تصویر دارای اعوجاج (tahoma) نتیجه پرشیا نگار برای تصویر رفع اعوجاج شده (پایین) در شکل فوق نتیجه حاصل از نرم افزار پرشیا نگار بر روی دو تصویر دارای اعوجاج وبدون اعوجاج با قلم tahoma نشان داده شده است. بدترین نتیجه در مورد پنج قلم مورد بررسی در مورد قلم tahoma ایجاد می گردد. در اینجا جداسازی خطوط تصویر به نحو مناسبی انجام نشده است. با این وجود در خروجی نرم افزار پرشیا نگار با اجرای الگوریتم پیشنهادی بر روی تصویر دارای اعوجاج بهبود نسبی ایجاد شده است.

## ۳-۴ نتیجه گیری

در این فصل نتایج تجربی حاصل از اجرای الگوریتم پیشنهادی بر روی تعدادی تصویر نمونه نشان داده شد و سپس توسط نویسه‌خوان پرشیانگار نتایج خروجی الگوریتم پیشنهادی مورد ارزیابی قرار گرفت. نتایج تجربی برای پنج تصویر نمونه با پنج قلم مختلف نشان داده شده و بررسی گردید.

از نتایج تجربی حاصل می‌توان نتیجه گرفت که الگوریتم پیشنهادی، برای تصاویر اسناد فارسی اعوجاج‌دار بدون توجه به نوع اعوجاج و قلم آن، دقت قابل قبولی داشته و چرخش سه بعدی متن را حذف می‌کند. اما میزان خوانایی تصویر رفع اعوجاج شده برای نرم‌افزار OCR برای تمامی قلم‌ها یکسان نیست.

با توجه به زمان صرف شده جهت آغشته‌سازی و سایر پردازش‌ها مدت زمان اجرای برنامه برای یک تصویر اسکن شده از یک کاغذ A4 با وضوح ۳۰۰ نقطه بر اینچ، حدود ۲۰ ثانیه است. بخش قابل توجهی از این زمان یعنی حدود ۵ ثانیه صرف آغشته‌سازی و تشخیص خطوط می‌گردد، البته این مساله برای تمامی روش‌های آغشته‌سازی صادق است.

در پایان می‌توان این‌طور نتیجه‌گیری کرد که الگوریتم پیشنهادی در رفع چرخش سه بعدی تصاویر اسناد فارسی دقت بالایی دارد. هر چند به دلیل عدم وجود یک مجموعه داده هماهنگ، نمی‌توان مقایسه دقیقی بین الگوریتم پیشنهادی و روش‌های قبلی انجام داد.

## فصل پنجم

### نتیجه گیری و پیشنهادات

## ۵-۱ نتیجه گیری

در این پایان نامه الگوریتم جدیدی جهت رفع چرخش سه بعدی در تصاویر اسناد فارسی پیشنهاد گردید. در فصل سوم مراحل مختلف الگوریتم پیشنهادی به طور جزء به جزء بررسی شد و نتایج حاصل از اعمال الگوریتم پیشنهادی بر روی تصاویر با متون و قلم‌های متفاوت ارائه گردید.

در فصل چهارم بعد از بررسی الگوریتم پیشنهادی بر روی تصاویر بزرگتر و متفاوت، یک ارزیابی توسط نرم‌افزار پرشیانگار انجام گردید و میزان بهبود ایجاد شده بر روی تصاویر پس از رفع اعوجاج توسط الگوریتم پیشنهادی نشان داده شد.

یافته‌های این پایان نامه را می‌توان به صورت زیر خلاصه کرد :

- ۱- جهت تشخیص خطوط متن از آغشته‌سازی به کمک الگوریتم DLCCM استفاده شده است که نسبت به روش‌های قبلی آغشته‌سازی دقت بالاتری داشته و جداسازی مناسب خطوط متن را در تصاویری که اعوجاج زیادی داشته و فاصله خطوط نزدیک به هم هستند، امکان پذیر می‌سازد.
- ۲- در تمامی کارهای قبلی انجام شده در این زمینه تصویر خروجی همواره به صورت دو سطحی بوده است. در این پایان نامه علاوه بر خروجی دو سطحی، یک خروجی با زمینه خاکستری نیز ارائه شده است. خروجی با زمینه خاکستری در مورد تصاویری که به دلیل وضوح پایین، تصویر دو سطحی مناسبی ندارند، خوانایی خوب و مناسبی دارد.
- ۳- از آنجا که مجموعه داده فارسی، در مورد تصاویر دارای چرخش سه بعدی وجود نداشت، یک مجموعه داده با بیش از یکصد تصویر و در ۱۰ قلم متفاوت ایجاد شد که می‌تواند در کارهای مشابه مورد استفاده قرار گیرد.
- ۴- بر خلاف کارهای قبلی، در این پایان نامه الگوریتم پیشنهادی بر روی تصاویر متنوع با اعوجاجات سه بعدی مختلف، درصد اعوجاجات مختلف و قلم‌های متفاوت مورد بررسی قرار گرفت.

۵- جهت بررسی میزان کارایی الگوریتم پیشنهادی، از نرم افزار توانمند نویسه خوان پرشیانگار استفاده شده است.

بطور خلاصه می توان این طور نتیجه گیری کرد که الگوریتم پیشنهادی برای انواع تصاویر اعوجاج دار با قلم های مختلف کارایی قابل قبولی دارد و تصویر رفع اعوجاج شده با الگوریتم پیشنهادی میزان دقت نرم افزار پرشیانگار را برای تصاویر اعوجاج دار افزایش می دهد.

## ۵-۲ پیشنهادات برای کارهای آینده

با توجه به الگوریتم پیشنهادی در این پایان نامه، موارد زیر برای ادامه کار پیشنهاد می گردد:

- ۱- علاوه بر مشکل اعوجاج غیر خطی که در این پایان نامه به آن پرداخته شد، تصاویر عکس برداری شده با دوربین و یا تصاویر اسکن شده مشکل دیگری نیز دارند. بدین صورت که روشنایی وسط تصویر سند با حاشیه های آن تفاوت زیادی دارد. این مساله باعث می شود که دوسطحی سازی تصویر به نحو مناسبی انجام نگیرد و در نتیجه رفع چرخش سه بعدی نیز با مشکل مواجه شود. از این رو رفع این مشکل برای کارهای آینده پیشنهاد می شود.
- ۲- در تصاویری که اعوجاج غیر خطی دارند، کلمات قرار گرفته در قسمت برآمدگی تصویر اندازه بزرگتری دارند. کوچک کردن این کلمات و همسان سازی اندازه آنها با سایر کلمات متن در هنگام رفع اعوجاج غیرخطی نیز برای کارهای آینده پیشنهاد می شود.
- ۳- با توجه به اینکه الگوریتم پیشنهادی به عنوان یک مرحله پیش پردازش برای فرآیند OCR است می توان گفت که زمان صرف شده جهت اجرای الگوریتم پیشنهادی نسبت به زمان صرف شده جهت خواندن تصویر توسط OCR زمان زیادی است، از این رو کاهش زمان اجرای الگوریتم به عنوان یکی از کارهای آینده پیشنهاد می شود.

# منابع

[۱] شورای پژوهشی OCR کارگروه خط و زبان فارسی، "پژوهشنامه نویسه خوان نوری (OCR) فارسی، ۱۳۸۶

- [2] H. Dehbovid, F.Razzazi, S.Alirezaii. " A novel method for de-warping in Persian document images captured by cameras".International Conference on Computer Information Systems and Industrial Management Applications(CISIM), Krackow, 2010.
- [3] M. Pilu. "Deskewing perspectively distorted documents: An approach based on perceptual organization". In HP Technical Reports, 2001.
- [4] Ulges, C. Lampert, and T. M. Breuel. "Document capture using stereo vision". In Proceedings of the ACM Symposium on Document Engineering, Milwaukee, Wisconsin, USA, 2004.
- [5] A. Yamashita, A. Kawarago, T. Kaneko and K.T.Miura. "Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system". In Proceedings of 17th International Conference on Pattern Recognition (ICPR) , Cambridge UK, 2004
- [6] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: a survey," Int.Journal on Document Analysis and Recognition, vol. 7, no. 2-3, pp. 84–104, 2005.
- [7] L. Zhang and C.L. Tan. "Warped image restoration with applications to digital libraries". In Proc. Eighth Int. Conf. on Document Analysis and Recognition, Washington, DC, USA,2005.
- [8] Filip Sadlo, Tim Weyrich, Ronald Peikert, Markus Gross, "A Practical Structured Light Acquisition system for Point-Based Geometry and Texture",In the Proceedings of the Eurographics Symposium on Point-Based Graphics ,2005.
- [9] Cao, H., Ding, X., and Liu, C., "A cylindrical surface model to rectify the bound document image", Int.'l Conf. Computer Vision, pp. 228-233,2003.
- [10] J. Liang, D. DeMenthon, and D. Doermann, "Geometric Rectification of Camera-captured Document Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. x, no. x, yy,2006
- [11] T.Wada, H. Ukida, and T. Matsuyama, "Shape from Shading with Interreflections under Proximal Light Source: 3D Shape Reconstruction of Unfolded Book Surface from a Scanner Image," Int'l Conf. Computer Vision (ICCV '95), pp. 66-71, 1995.
- [12] M.S. Brown and W.B. Seales, "Document Restoration using 3D Shape: a General Deskewing Algorithm for Arbitrarily Warped Documents", Int'l Conf. on Computer Vision, Volume 2, pp. 367-374, 2001
- [13] T. Wada, H. Ukida and T. Matsuyama. "Shape from Shading with Interreflections Under a Proximal Light Source: Distortion-Free Copying of an unfolded Book", Int'l Journal of Computer Vision, 24(2), pp. 125-135, 1997.
- [14] O. Lavaille, X. Molines, F. Angella, and P. Baylou, "Active contours network to straighten distorted text lines," in Proc. Of Int. Conf. on Image Processing, Thessaloniki, Greece, 2001, pp. 1074–1077.

- [15] D. C. Schneider, M. Block, and R. Rojas, "Robust document warping with interpolated vector fields," in Proc. 9th Int. Conf. Document Anal. Recognit., Curitiba, Brazil, 2007, pp. 113–117.
- [16] A. Masalovitch, L. Mestetskiy, "Usage of continuous skeletal image representation for document images dewarping". Proceedings of international conference PRIP (Pattern Recognition and Information Processing), 2007, pp. 279-284.
- [17] C. Wu and G. Agam, "Document image De-warping for Text/Graphics recognition," in Joint IAPR Int. Workshop Structural, Syntactic Statistical Pattern Recognit., Windsor, Canada, 2002, pp. 348–357.
- [18] Y. Weng and Q. Zhu. "Nonlinear Shape Restoration for Document Images", Computer Vision and Pattern Recognition Conference, pp. 568-573, 1996.
- [19] A. Ulges, C. H. Lampert, and T. M. Breuel, "Document Image Dewarping using Robust Estimation of Curled Text Lines" Proceedings of the eight international conference on document analysis and recognition. pp.1001-1005, USA. 2005
- [20] H. Cao, X. Ding, and C. Liu, "Rectifying the bound document image captured by the camera: A model based approach," in Proc. 7th Int. Conf. Document Anal. Recognition, 2003, pp. 71–75.
- [۲۱] هادی ده‌بوید، فرید رزازی، شاپور علی‌رضایی، "ارائه روشی نوین برای کاهش اعوجاج تصویربرداری در تصاویر متنی فارسی تصویربرداری شده توسط دوربین"، ششمین کنفرانس ماشین بینایی و پردازش تصاویر ایران، ۱۳۸۹
- [۲۲] مریم گل پرداز، حسین نظام آبادی پور، سعید صیدنژاد، "رفع اعوجاج هندسی در تصاویر متن فارسی گرفته شده با دوربین دیجیتال"، شانزدهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، تهران، ۱۳۸۹
- [23] M. Sezgin, B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", Journal of Electronic Imaging 13(1), 146–165, 2004.
- [24] S. U. Le, S. Y. Chung, and R. H. Park, "A comparative performance study of several global thresholding techniques for segmentation," Graph. Models Image Process. 52, 171–190, 1990.
- [25] J. S. Weszka and A. Rosenfeld, "Threshold evaluation techniques," IEEE Trans. Syst. Man Cybern. SMC-8, 627–629, 1978.
- [26] P. W. Palumbo, P. Swaminathan, and S. N. Srihari, "Document image binarization: Evaluation of algorithms," Proc. SPIE 697, 278–286, 1986.
- [27] O. D. Trier and A. K. Jain, "Goal-directed evaluation of binarization methods," IEEE Trans. Pattern Anal. Mach. Intell. PAMI-17, 1191–1201, 1995.
- [28] C. A. Glasbey, "An analysis of histogram-based thresholding algorithms," Graph. Models Image Process. 55, 532–537, 1993.
- [29] A. Rosenfeld and P. De la Torre, "Histogram concavity analysis as an aid in threshold selection," IEEE Trans. Syst. Man Cybern. SMC-13, 231–235, 1983.
- [30] J. Weszka and A. Rosenfeld, "Histogram modification for threshold selection," IEEE Trans. Syst. Man Cybern. SMC-9, 38–52, 1979.

- [31] R. J. Whatmough, "Automatic threshold selection from a histogram using the exponential hull," *Graph. Models Image Process.* 53, 592–600 .1991.
- [32] M. I. Sezan, "A peak detection algorithm and its application to histogram-based image data reduction," *Graph. Models Image Process.* 29, 47–59 .1985.
- [33] S. Boukharouba, J. M. Rebordao, and P. L. Wendel, "An amplitude segmentation method based on the distribution function of an image," *Graph. Models Image Process.* 29, 47–59 .1985.
- [34] D. M. Tsai, "A fast thresholding selection procedure for multimodal and unimodal histograms," *Pattern Recogn. Lett.* 16, 653–666,1995.
- [35] J. C. Olivo, "Automatic threshold selection using the wavelet transform," *Graph. Models Image Process.* 56, 205–218 ,1994.
- [36] N. Ramesh, J. H. Yoo, and I. K. Sethi, "Thresholding based on histogram approximation," *IEE Proc. Vision Image Signal Process.*142.5., 271–279 ,1995.
- [37] T. Kampke and R. Kober, "Nonparametric optimal binarization," *ICPR'98, Intl. Conf. Patt. Recog.*, pp. 27–29 ,1998.
- [38] J. Cai and Z. Q. Liu, "A new thresholding algorithm based on all-pole model," *ICPR'98, Intl. Conf. Patt. Recog.*, pp. 34–36 ,1998.
- [39] R. Guo and S. M. Pandit, "Automatic threshold selection based on histogram modes and a discriminant criterion," *Mach. Vision Appl.*10, 331–338 ,1998.
- [40] T. W. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Trans. Syst. Man Cybern. SMC-8*, 630–632,1978.
- [41] M. K. Yanni and E. Horne, "A new approach to dynamic thresholding," *EUSIPCO'94: 9th European Conf. Sig. Process.* 1,34–44,1994.
- [42] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Syst. Man Cybern. SMC-9*,62–66 ,1979.
- [43] D. E. Lloyd, "Automatic target classification using moment invariant of image shapes," *Technical Report, RAE IDN AW126, Farnborough,UK .Dec, 1985.*
- [44] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recogn.* 19,41–47 ,1986.
- [45] S. Cho, R. Haralick, and S. Yi, "Improvement of Kittler and Illingworths's minimum error thresholding," *Pattern Recogn.* 22, 609–617,1989.
- [46] N. Ramesh, J. H. Yoo, and I. K. Sethi, "Thresholding based on histogram approximation," *IEE Proc. Vision Image Signal Process.*142.5., 271–279 , 1995.
- [47] C. V. Jawahar, P. K. Biswas, and A. K. Ray, "Investigations on fuzzy thresholding based on fuzzy clustering," *Pattern Recogn.* 30.10.,1605–1613 , 1997.

- [48] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Graph. Models Image Process.* 29, 273–285 .1985.
- [49] P. Sahoo, C. Wilkins, and J. Yeager, "Threshold selection using Renyi's entropy," *Pattern Recogn.* 30,71–84 .1997.
- [50] T. Pun, "A new method for gray-level picture threshold using the entropy of the histogram," *Signal Process.* 2.3., 223–237 .1980.
- [51] C. H. Li and C. K. Lee, "Minimum cross-entropy thresholding," *Pattern Recogn.* 26, 617–625 .1993.
- [52] C. H. Li and P. K. S. Tam, "An iterative algorithm for minimum cross-entropy thresholding," *Pattern Recogn. Lett.* 19, 771–776.1998.
- [53] A. D. Brink and N. E. Pendock, "Minimum cross entropy threshold selection," *Pattern Recogn.* 29, 179–188 .1996
- [54] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inf. Theory* IT-26,26–37 .1980.
- [55] A. G. Shanbag, "Utilization of information measure as a means of image thresholding," *Comput. Vis. Graph. Image Process.* 56,414–419 .1994.
- [56] H. D. Cheng, Y. H. Chen, and Y. Sun, "A novel fuzzy entropy approach to image enhancement and thresholding," *Signal Process.* 75,277–301 .1999.
- [57] A. Kaufmann, *Introduction to the Theory of Fuzzy Sets: Fundamental Theoretical Elements*, Vol. 1, Academic Press, New York .1980.
- [58] W. H. Tsai, "Moment-preserving thresholding: A new approach," *Graph. Models Image Process.* 19, 377–393 .1985.
- [59] L. Hertz and R. W. Schafer, "Multilevel thresholding using edge matching," *Comput. Vis. Graph. Image Process.* 44, 279–295 .1988.
- [60] L. K. Huang and M. J. J. Wang, "Image thresholding by minimizing the measures of fuzziness," *Pattern Recogn.* 28, 41–51 .1995.
- [61] K. Ramar, S. Arunigam, S. N. Sivanandam, L. Ganesan, and D. Manimegalai, "Quantitative fuzzy measures for threshold selection," *Pattern Recogn. Lett.* 21, 1–7 .2000.
- [62] J. C. Russ, "Automatic discrimination of features in gray-scale images," *J. Microsc.* 148.3., 263–277 .1987.
- [63] C. K. Leung and F. K. Lam, "Maximum segmented image information thresholding," *Graph. Models Image Process.* 60, 57–76 .1998.
- [64] A. Rosenfeld, "The fuzzy geometry of image subsets," *Pattern Recogn. Lett.* 2, 311–317 .1984.

- [65] S. K. Pal and A. Rosenfeld, "Image enhancement and thresholding by optimization of fuzzy compactness," *Pattern Recogn. Lett.* 7,77–86.1988.
- [66] A. Kaufmann, *Introduction to the Theory of Fuzzy Sets: Fundamental Theoretical Elements*, Vol. 1, Academic Press, New York .1980.
- [67] R. L. Kirby and A. Rosenfeld, "A note on the use of .gray level, local average gray level. space as an aid in threshold selection," *IEEE Trans. Syst. Man Cybern. SMC-9*, 860–864 .1979.
- [68] N. R. Pal and S. K. Pal, "Entropic thresholding," *Signal Process.* 16,97–108 .1989.
- [69] C. Chang, K. Chen, J. Wang, and M. L. G. Althouse, "A relative entropy based approach in image thresholding," *Pattern Recogn.* 27,1275–1289 .1994.
- [70] A. S. Abutaleb, "Automatic thresholding of gray-level pictures using two-dimensional entropy," *Comput. Vis. Graph. Image Process.* 47,22–32 .1989.
- [71] A. Beghdadi, A. L. Negrata, and P. V. De Lesegno, "Entropic thresholding using a block source model," *Graph. Models Image Process.*57, 197–205 .1995.
- [72] W. Niblack, *An Introduction to Image Processing*, pp. 115–116, Prentice-Hall, Englewood Cliffs, NJ .1986.
- [73] 111. J. Sauvola and M. Pietaksinen, "Adaptive document image binarization," *Pattern Recogn.* 33, 225–236 .2000.
- [74] J. M. White and G. D. Rohrer, "Image thresholding for optical character recognition and other applications requiring character image extraction," *IBM J. Res. Dev.* 27.4.,400–411 .1983.
- [75] J. Bernsen, "Dynamic thresholding of gray level images," *ICPR'86:Proc. Intl. Conf. Patt. Recog.* , pp. 1251–1255 .1986.
- [76] Y. Yang and H. Yan, "An adaptive logical method for binarization of degraded document images," *Pattern Recogn.* 33, 787–807 .2000.
- [77] C. A. Murthy and S. K. Pal, "Fuzzy thresholding: A mathematical framework, bound functions and weighted moving average technique," *Pattern Recogn. Lett.* 11, 197–206 .1990.
- [78] S. D. Yanowitz and A. M. Bruckstein, "A new method for image segmentation," *Comput. Graph. Image Process.* 46, 82–95 .1989.
- [79] D. Shen and H. H. S. Ip, "A Hopfield neural network for adaptive image segmentation: An active surface paradigm," *Pattern Recogn.Lett.* 18,37–48 .1997.
- [80] L. Likforman-Sulem\*, A. Zahour\*\*, B. Taconet, "Text Line Segmentation of Historical Documents: a Survey", *International Journal on Document Analysis and Recognition*, Springer, 2006.
- [81] A. Zahour, , B. Taconet, , P. Mercy, S. Ramdane, , "Arabic hand-written text-line extraction", *Proceedings of the 6<sup>th</sup> ICDAR*, Seattle, pp. 281– 285, 2001

- [82] V. Shapiro, G. Gluhchev, V. Sgurev, "Handwritten document image segmentation and analysis", *Pattern recognition Letters*, 14:71-78, 1993
- [83] A. Antonacopoulos., D. Karatzas, "Document Image analysis for World War II personal records", *First Int. Workshop on Document Image Analysis for Libraries, DIAL'04, Palo Alto*, pp. 336-341, 2004
- [84] He J., Downton, A.C.. "User-Assisted Archive Document Image Analysis for Digital Library Construction", *Seventh International Conference on Document Analysis and Recognition, Edinburgh, 2003*
- [85] K. Wong, R. Casey, F. Wahl, "Document analysis systems", *IBM Journal of research and development*, Vol 26, no 6, 1982
- [86] Z. Shi, V. Govindaraju. "Line Separation for Complex Document Images Using Fuzzy Runlength" , *Proc. of the Int. Workshop on Document Image Analysis for Libraries, Palo, Alto, CA, January 23-24, 2004*
- [87] L. Likforman-Sulem, C. Faure, "Extracting lines on handwritten documents by perceptual grouping, in *Advances in Handwriting and drawing: a multidisciplinary approach*", Winter (Eds), pp. 21-38, Europia, Paris. 1994
- [88] L. Likforman-Sulem, A. Hanimyan, C. Faure, " A Hough Based Algorithm for Extracting Text Lines in Handwritten Document" , *Proc. of ICDAR'95, 774-777, 1995*
- [89] E. Oztop, A. Y. Mulayim, V. Atalay, F. Yarman-Vural, "Repulsive attractive network for baseline extraction on document images", *Signal Processing*, 75:1-10, 1999
- [90] Z. Shiand, V. Govindaraju, "Multi-scale Techniques for Document Page Segmentation". *Proceedings. Eighth International Conference on In Document Analysis and Recognition*, pp. 1020-1024 Vol. 2, 2005.
- [91] <http://p30codenevis.ir/Persian-Reader/>

## **Abstract**

Most of the works about removing document nonlinear distortion is in Latin document images. The purpose of removing nonlinear distortion is compensating page curl and flattening scanned or camera captured document, so that this documents are readable by OCR. In this thesis a new algorithm is proposed for removing 3D page curl in Persian document images. In the proposed algorithm noise reduction, binarization, smearing, baseline detection is done and then words of dewarped lines are located in the right place.

In the proposed algorithm, DLCM approach is used for line detection. The output of dewarped image is presented in both binary and grayscale layout. To evaluate the operation of the proposed algorithm ,Persia negar which is a Persian OCR software is used. Experimental results show good operation of the proposed algorithm.

Keywords: OCR, page curl, database, smearing, document image restoration.



**Shahrood University of Technology**

**Faculty of electrical engineering and robotics**

## **Removing nonlinear distortion in Persian document images**

**Samira Khosravirad**

Supervisor:

**Dr Hossein Khosravi**

**Date: Jan 2013**