



مرکز اطلاعات و مدارک علمی ایران

**فاز مطالعاتی و امکان‌سنجی شناسایی شماره
صفحات پایان‌نامه‌ها در بخش فهرست مطالب و
صفحات اصلی جهت زمینه‌سازی برقراری ارتباط
خودکار بین آنها**

مجری: اسماعیل فرامرزی

داور: دکتر احسان‌الله کبیر

ناظر: مهندس امیر پارس‌ی اصفهانی

دیماه ۱۳۸۲

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

فصل اول

طرح پڑوہش

پیدایش علوم و فنون جدید، جوامع بشری را با شکلهای مختلفی از اطلاعات روبرو نموده است. سطح توسعه یک جامعه را می توان با مقدار اطلاعات و دانش تولید شده در آن ارزیابی کرد. تولید فزاینده اطلاعات به شکلهای مختلف صورت می گیرد و با درجات متفاوتی از پیچیدگی همراه می باشد. در نتیجه نیاز به سیستمهای پردازش اطلاعات بصورت روزافزون افزایش می یابد. یکی از مسائل مهم در طراحی سیستمهای مدرن اطلاعاتی، بازشناسی خودکار الگوها است [۸۲].

۱-۱- شناسایی الگو^۱

شناسایی الگو شاخه ای از هوش مصنوعی^۲ است که با طبقه بندی (کلاسه بندی)^۳ و توصیف مشاهدات سروکار دارد. شناسایی الگو به ما کمک می کند تا داده ها (الگوها) را با تکیه بر دانش قبلی یا اطلاعات آماری استخراج شده از الگوها، طبقه بندی نماییم. الگوهایی که می بایست کلاسه بندی شوند، معمولاً گروهی از سنجشها یا مشاهدات هستند که مجموعه نقاطی را در یک فضای چند بعدی مناسب تعریف می نمایند.

یک سیستم شناسایی الگوی کامل متشکل است از یک حسگر^۴ که مشاهداتی را که می بایست توصیف یا کلاسه بندی گردند جمع آوری می نماید، یک مکانیزم استخراج ویژگیها^۵ که اطلاعات عددی یا نمادین (سمبولیک) را از مشاهدات محاسبه می کند و یک نظام کلاسه بندی یا توصیف که وظیفه اصلی طبقه بندی یا توصیف الگوها را با تکیه بر ویژگیهای استخراج شده عهده دار است.

شکل (۱-۱) بلوک دیاگرام یک سیستم شناسایی الگو را نشان می دهد [53]. همانطوریکه از پیکانه های برگشتی مشخص است، این بلوکها لزوماً مستقل نیستند و بسته به نتایج حاصله ممکن است لازم باشد که بلوهای اولیه مجدداً طراحی گردند تا راندمان کلی سیستم بهبود یابد.

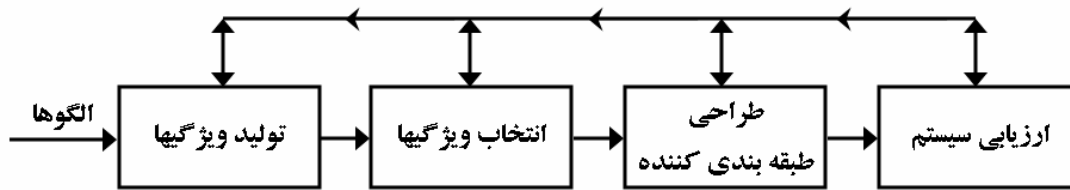
¹ Pattern Recognition

² Artificial intelligence

³ Classification

⁴ Sensor

⁵ Feature extraction



شکل ۱-۱- بلوک دیاگرام یک سیستم شناسایی الگو [53]

نظام کلاسه‌بندی یا توصیف معمولاً مبتنی بر وجود یک مجموعه از الگوهای است که قبلاً کلاسه‌بندی یا توصیف شده‌اند. این مجموعه الگوها را مجموعه آموزشی^۱ و قانون یادگیری منتج شده را قانون یادگیری باسرپرستی (با نظارت)^۲ می‌نامند. همچنین یادگیری می‌تواند بصورت بدون نظارت^۳ باشد و این در حالی است که الگوهای از قبل به سیستم داده نشده‌اند و در مقابل، سیستم خود بر اساس قواعد آماری الگوها، کلاسها را پایه‌گذاری می‌کند.

۲-۱- کاربردهای بازشناسی الگو

بازشناسی الگو در بسیاری از زمینه‌ها نقش کاربردی دارد [۸۲]. بازشناسی حروف، بازشناسی نویسنده، تصدیق امضاء، طبقه‌بندی اثر انگشت و بازشناسی گفتار نمونه‌هایی از این کاربردها هستند. شناسایی الگو برای تحلیل داده‌های پزشکی نیز بکار گرفته شده است. برای مثال تفسیر الکتروکاردیوگرام، تحلیل تصاویر اشعه X و طبقه‌بندی کروموزمها را می‌توان نام برد. نمونه‌های دیگری از این کاربردها شامل طبقه‌بندی مناطق زراعی، مطالعه آلودگی آبها، آشکار کردن منابع زیرزمینی و پیش‌بینی آب و هواست. در این نوع کاربردها، تصاویر ارسال شده از ماهواره و تصاویر هوایی به کمک روشهای بازشناسی الگو تفسیر می‌شوند. بازرسی تصویری و بازشناسی قطعات ماشینی، از کاربردهای صنعتی شناسایی الگو هستند. تحلیل بافت، آشکارسازی هدف در سیگنالهای برگشتی رادار یا سونار^۴، طبقه‌بندی امواج زلزله و تشخیص ذرات شیمیایی کاربردهای دیگری از بازشناسی الگو می‌باشند.

^۱ Training set

^۲ Supervised learning

^۳ Unsupervised

^۴ Sonar

۱-۳- بازشناسی نوری حروف (OCR)^۱

در چند دهه گذشته مسأله بازشناسی الگوهای نوشتاری شامل حروف، ارقام و سایر نمادهای متداول در اسناد مکتوب شده به زبانهای مختلف، توسط گروههای مختلفی از محققین مورد مطالعه و بررسی قرار گرفته است. نتیجه این تحقیقات منجر به پیدایش مجموعه‌ای از روشهای سریع و تا حد زیادی مطمئن موسوم به OCR یا «بازشناسی نوری حروف» بمنظور وارد نمودن اطلاعات موجود در اسناد، مدارک، کتابها و سایر مکتوبات چاپی یا تایپ شده و حتی دستنویس^۲ به داخل کامپیوتر شده است [۶۸].

هنگامی که یک سند متنی اسکن می‌شود، کامپیوتر این متن را بصورت یک تصویر گرافیکی تشخیص می‌دهد. در نتیجه کاربران قادر نخواهند بود که متن موجود در تصویر سند را ویرایش نمایند و یا آن را مورد جستجو قرار دهند. یک نرم‌افزار OCR این متن اسکن شده را خوانده، محتویات آنرا شناسایی نموده و بصورت یک فایل در کامپیوتر ذخیره می‌سازد. چنین قابلیتی امکان استفاده گسترده از کامپیوتر را در پردازش سریع حجم وسیعی از داده‌های مکتوب تولید شده توسط شرکتها و مؤسسات مختلف نظیر بانکها، شرکتهای بیمه، مؤسسات خدمات عمومی، اداره پست و سایر نهادهایی که سالیانه با میلیونها مورد پرداخت، دریافت و حسابرسی امور مشتریان خود مواجهند، فراهم می‌آورد.

بدنبال پیشرفتهای حاصله در زمینه ساخت پردازنده‌های سریع، اثر محدودکننده نحوه ورود داده‌ها به درون کامپیوتر بر سرعت انتقال اطلاعات، بیش از پیش نمود یافته است. با بکارگیری دستگاههای OCR علاوه بر دستیابی به سرعت بالاتر در مرحله داده‌خوانی، امکان استفاده از قابلیتهای پیش‌پردازشی آنها، ویرایشگری و تغییر فرمت داده‌ها نیز فراهم می‌آید.

۱-۳-۱- تاریخچه سیستمهای OCR

از جنبه تاریخی، سیستمهای OCR تا کنون سه مرحله تکاملی را پشت سر گذاشته‌اند [6]:

^۱ Optical Character Recognition

^۲ Handwritten

الف- مرحله تکوین (از ۱۹۰۰ تا ۱۹۸۰): رد پای اولیه اقدامات صورت گرفته در زمینه بازشناسی حروف را در سالهای اول دهه ۱۹۰۰ می توان یافت و آن زمانی است که Turing دانشمند روسی بر آن بود که به افراد مبتلا به نارساییهای بینایی کمک نماید. اولین اختراعاتی ثبت شده در این زمینه مربوط به سالهای ۱۹۲۹ و ۱۹۳۳ میلادی هستند [43]. این سیستمها حروف چاپی را با روش تطابق قالبی^۱ شناسایی می کردند. ماسکهای مکانیکی مختلفی از مقابل تصویر حرف عبور می کردند و نور از یک سو به آن تابانده شده و از سوی دیگر توسط یک آشکارساز نوری دریافت می شد. وقتی یک انطباق کامل صورت می گرفت، نور به آشکارساز نمی رسید و حرف ورودی بازشناسی می شد. این اختراع به دلیل تکنولوژی ایتومکانیکی مورد استفاده از آنها کاربردی نبود. تصور دسترسی به دستگاهی برای بازشناسی حروف تا دهه ۱۹۴۰ میلادی و ظهور کامپیوترهای دیجیتال بصورت یک رؤیا باقی ماند.

اقدامات اولیه در زمینه بازشناسی حروف، بر روی متون چاپی و یا مجموعه کوچکی از حروف و نمادهای دستنویس که براحتی قابل تشخیص بودند، متمرکز گردیده بود. سیستمهای بازشناسی حروف چاپی که در این مقطع زمانی معرفی شدند، عمدتاً از روش تطابق قالبی استفاده می نمودند که در آن، تصویر ورودی با کتابخانه ای از تصاویر حروف مورد مقایسه قرار می گرفت. در مورد متون دستنویس نیز الگوریتمهای پردازش تصویر که ویژگیهای سطح پایین^۲ را از تصاویر استخراج می کنند، به تصاویر دودویی اعمال می شد تا بردارهای ویژگی استخراج گردند. سپس این بردارهای ویژگی به طبقه کننده های آماری سپرده می شدند.

در این دوره، تحقیقات موفق اما مقیدی بیشتر بر روی حروف و اعداد لاتین انجام گرفت. با این وجود مطالعات چندی نیز در زمینه حروف ژاپنی، چینی، عبری، هندی، سیریلیکی، یونانی و عربی در هر دو زمینه حروف چاپی و دستنویس آغاز گردید. با ظهور صفحات رقومی کننده (دیجیتایزرها)^۳ در دهه ۱۹۵۰ که قادر به تشخیص مختصات حرکتی نوک یک قلم مخصوص بودند، سیستمهای OCR تجاری نیز امکان عرضه یافتند. این نوآوری سبب شد که محققان بتوانند در زمینه

^۱ Template matching

^۲ Low level features

^۳ Digitizers

بازشناسی برخط حروف دستنویس فعالیت خود را آغاز نمایند. مرجع [51] یک منبع مناسب در مورد اقدامات صورت گرفته بر روی بازشناسی برخط حروف تا سال ۱۹۸۰ می‌باشد.

ب- مرحله توسعه (از ۱۹۸۰ تا ۱۹۹۰): مطالعات صورت گرفته تا قبل از سال ۱۹۸۰ از فقدان سخت‌افزارهای کامپیوتری قدرتمند و دستگاههای اخذ داده‌ها رنج می‌بردند. در این دهه بواسطه رشد انفجارگونه تکنولوژی اطلاعات، وضعیت بسیار مناسبی برای زمینه‌های تحقیقاتی مختلف از جمله بازشناسی حروف فراهم گردید. روشهای ساختاری به همراه روشهای آماری در بسیاری از سیستمها ظاهر گردیدند. تحقیقات در زمینه OCR اساساً توجه خود را به روشهای بازشناسی اشکال بدون توجه به هرگونه اطلاعات معناشناختی معطوف نمود. این مسئله سبب گردید که نرخ بازشناسی نتواند از یک حد خاص فراتر رود که در بسیاری از کاربردهای OCR قابل قبول نبود. مروری بر تحقیقات و پیشرفتهای حاصل شده در مورد OCR در این برهه زمانی را می‌توان در مراجع [43] و [51] بترتیب برای بازشناسی برخط و برون‌خط جستجو نمود.

ج- مرحله بهبود (از ۱۹۹۰ به بعد): در این مقطع زمانی بود که با تکوین ابزارها و تکنیکهای پردازشی جدید، پیشرفت واقعی سیستمهای OCR محقق گردید. در اوایل دهه ۹۰، روشهای پردازش تصویر و بازشناسی الگو با تکنیکهای کارآمد هوش مصنوعی^۱ (AI) ادغام گشتند. محققان، الگوریتمهای بازشناسی حروف پیچیده‌ای را ابداع نمودند که قادر بودند داده‌های ورودی با تفکیک-پذیری^۲ بالا را دریافت کنند و در مرحله پیاده‌سازی، محاسبات بسیار زیادی را بر روی داده‌ها انجام دهند. امروزه علاوه بر وجود کامپیوترهای قدرتمندتر و تجهیزات الکترونیکی دقیقتر مانند اسکنرها، دوربین‌ها و صفحات رقومی‌کننده، استفاده از تکنیکهای پردازشی مدرن و توانمندی همچون شبکه-های عصبی^۳ (NNها)، مدل‌های مارکوف پنهان^۴ (HMMها)، منطقهای مجموعه فازی و مدل‌های پردازش زبانهای طبیعی^۵ (NLP) امکان‌پذیر گشته است. سیستمهای جدید OCR برون‌خط متون چاپی و برخط متون دستنویس با واژگان محدود و وابسته به نویسنده، برای کاربردهای محدود به

¹ Artificial Intelligence

² Resolution

³ Neural Networks

⁴ Hidden Markoff Model

⁵ Natural Language Processing

نحو کاملاً رضایت بخشی عمل می نمایند. با این وجود، بمنظور دستیابی به هدف نهایی در شبیه سازی ماشینین نگارش انسانی و متون چاپی، هنوز راه درازی در پیش است.

۱-۳-۲- بخشهای مختلف یک سیستم OCR کامل

یک سیستم OCR کامل معمولاً مرکب از ۵ قسمت اصلی زیر می باشد:

الف - پیش پردازش^۱: شامل کلیه اعمالی که روی سیگنال تصویری خام صورت می گیرند تا موجب تسهیل روند اجرای فازهای بعدی گردند؛ مانند باینری کردن تصویر، حذف نویز، هموارسازی، نازکسازی، تشخیص زبان و فونت کلمات و نظایر اینها.

ب - قطعه بندی^۲: عبارت است از روشهایی که بخشهای مختلفی همچون پاراگرافها، جملات یا کلمات، و حروف را از تصویر سند استخراج می نمایند.

ج - استخراج ویژگیها^۳: مجموعه کلیه محاسباتی است که روی الگوهای بدست آمده از مرحله پیش-پردازش انجام می شود تا بردار ویژگیهای متناظر با هر الگو تعیین گردد.

د - بازشناسی با یک یا چند طبقه بندی کننده^۴: شامل روشهایی برای متناظر ساختن هر یک از الگوهای بدست آمده از مرحله استخراج ویژگیها با یکی از کلاسهای فضای الگوهای مورد بحث است که از طریق کمینه ساختن فاصله بردار ویژگیهای هر الگو نسبت به یکی از بردارهای مرجع موجود در پایگاه داده های سیستم انجام می گیرد.

ه) بکارگیری اطلاعات جانبی (پس پردازش)^۵: مانند مجموعه لغات معتبر، اطلاعات آماری مربوط به رخداد حروف، اطلاعات دستوری و معنایی.

در فصل ۲ این قسمتها بطور جامع تشریح گردیده، روشهای مختلف توسعه داده شده برای آنها در مقالات، مورد بررسی قرار می گیرند. همچنین در فصل ۳ منحصراً به تعدادی از پایان نامه ها و مقالات مرتبط با حوزه سیستمهای OCR فارسی و عربی اشاره خواهد گردید. شکل (۱-۲) بلوک

^۱ Preprocessing

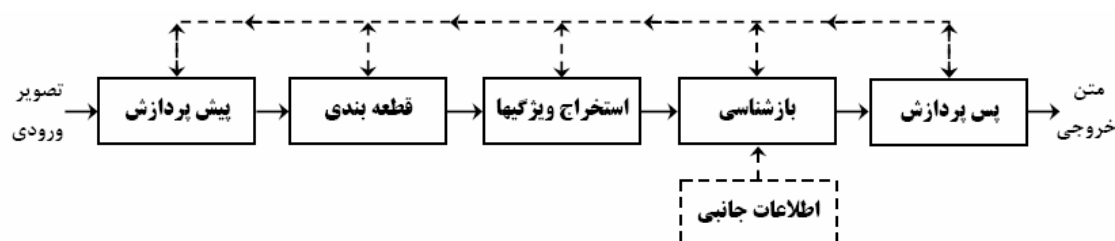
^۲ Segmentation

^۳ Feature Extraction

^۴ Classification and Recognition

^۵ Postprocessing

دیاگرام یک سیستم OCR را نمایش می‌دهد. البته توجه گردد که بسته به الگوریتم کلی بکار رفته و سطح انتظارات از عملکرد نرم‌افزار، ممکن است برخی سیستمها فاقد یک یا چند تا از مراحل فوق باشند.



شکل ۱-۲- بلوک دیاگرام یک سیستم OCR

۳-۳-۱- انواع سیستمهای OCR

در یک تقسیم‌بندی کلی می‌توان سیستمهای OCR را از لحاظ نوع الگوی ورودی به دو گروه اصلی تقسیم کرد:

الف- سیستمهای بازشناسی متون چاپی^۱.

ب- سیستمهای بازشناسی متون دستنویس^۲.

هر یک از این گروهها تکنیکهای خاص خود را دارند. همچنین از جنبه نحوه ورود اطلاعات، سیستمهای OCR به دو دسته زیر تقسیم‌بندی می‌شوند [۸۲]:

الف- سیستمهای برخط^۳.

ب- سیستمهای برون خط^۴.

در بازشناسی برخط، حروف در همان زمان نگارش توسط سیستم تشخیص داده می‌شوند. دستگاه ورودی این سیستمها یک قلم نوری است. در این روش علاوه بر اطلاعات مربوط به موقعیت قلم، اطلاعات زمانی مربوط به مسیر قلم نیز در اختیار است. این اطلاعات معمولاً توسط یک صفحه

^۱ Machine-Printed Character Recognition

^۲ Handwritten Character Recognition

^۳ Online

^۴ Offline

رقومی‌کننده^۱ اخذ می‌شوند. در این روش می‌توان از اطلاعات زمانی سرعت، شتاب، فشار و زمان برداشتن و گذاشتن قلم روی صفحه در بازشناسی استفاده کرد.

در بازشناسی برون‌خط، از تصویر دو بعدی متن ورودی استفاده می‌شود. در این روش به هیچ نوع وسیله نگارش خاصی نیاز نیست و تفسیر داده‌ها مستقل از فرآیند تولید آنها تنها براساس تصویر متن صورت می‌گیرد. این روش به نحوه بازشناسی توسط انسان شباهت بیشتری دارد.

۱-۳-۴- سیستمهای OCR فارسی

مسئله بازشناسی حروف الفبای فارسی سابقه‌ای نه چندان طولانی به همراه دارد. نخستین گزارشهای رسمی منتشر شده از تلاشهای انجام گرفته در این راه، مربوط به سالیان نخست دهه ۱۹۸۰ میلادی است [۶۸]. به رغم فراگیری نسبی کاربرد الفبای فارسی در میان ملل مختلف قاره آسیا، بررسی‌های انجام شده در خصوص یافتن روشهایی برای بازشناسی حروف این الفبا بسیار محدود بوده است. بواسطه وجود تفاوت‌های اساسی بین نحوه نگارش کلمات فارسی و کلمات لاتین نظیر چسبیده بودن حروف سازنده یک کلمه به یکدیگر و تغییر شکل حروف بر اساس موقعیت نسبی قرارگیری آن در یک کلمه فارسی، امکان اعمال مستقیم روشهای متداول در بازشناسی حروف انگلیسی بنظور شناسایی حروف تشکیل دهنده کلمات فارسی وجود ندارد.

جدول (۱-۱) جایگاه کنونی پیشرفتهای حاصل شده در زمینه سیستمهای OCR برای متون لاتین را به نمایش می‌گذارد [۶]؛ اما برای متون فارسی، نیاز به انجام تحقیقات بیشتر در همه قسمتها کاملاً احساس می‌گردد.

بنظر می‌رسد که در حال حاضر تنها نرم‌افزار تجاری کارآمد OCR که زبان فارسی را پشتیبانی می‌نماید، نرم‌افزار Automatic Reader محصول شرکت عربی Sakhr^۲ است که دارای دو نسخه Gold و Platinum می‌باشد. تفاوت بین این دو نسخه در این است که نسخه Platinum، SDK^۳ نرم‌افزار را نیز شامل می‌شود (SDK دسته‌ای از کدها و توابع کتابخانه‌ای است که توسط

^۱ Digitizer

^۲ <http://www.sakhr.com>

^۳ Software Developer's Kit (or Software Development Kit)

شرکت تولیدکننده نرم افزار ارائه می شود و امکان استفاده از قابلیت های آن نرم افزار را در سایر محصولات نرم افزاری فراهم می آورد؛ SDK نرم افزار OCR شرکت Sakhr در سه فرمت مختلف تولید شده است: OLE automation, COM object و DLL). در حال حاضر قیمت نسخه 7.0 آنها بترتیب در حدود ۱۴۰۰ و ۴۰۰۰ دلار است. این رقم در مقایسه با نرم افزار معروف OmniPage (نسخه 14 Pro) از شرکت ScanSoft^۱ که قیمت نسخه های معمولی و Office آن بترتیب در حدود ۱۵۰ و ۶۰۰ دلار است، واقعاً قابل ملاحظه می باشد. اما به هر حال بایستی توجه داشت که با وجود آنکه نرم افزار Automatic Reader زبان فارسی را پشتیبانی می کند، اما مجموعه لغات معتبر تعریف شده برای این نرم افزار که در بخش پس پردازش مورد استفاده قرار می گیرد (نرم افزار توسط این مجموعه لغات، خطاهای احتمالی را شناسایی و به کاربر اعلان می کند یا خود آنها را بصورت خودکار تصحیح می نماید)، تنها لغات عربی را شامل می گردد. همچنین شرکت "جیحون افزار" یک نرم افزار OCR فارسی بنام "شناسا" (ShenAsA) داشت که قیمت سال ۱۹۹۶ آن در حدود ۱۰۰۰ دلار بود^۲؛ اما ظاهراً از کارایی چندان مورد قبولی برخوردار نبوده است [۶۷].

جدول ۱-۱- جایگاه کنونی تحقیقات در زمینه سیستم های OCR لاتین

		متون چاپی			متون دستنویس		
		یک نوع فونت	چند نوع فونت	همه نوع فونت	گسسته	پیوسته	مخلوط
برخط	مقید						
	نامقید						
برون خط	بدون نویز						
	نویزی						

در حد مطلوب
 نیازمند بهبود
 نیازمند تحقیقات بیشتر

^۱ <http://www.scansoft.com>

^۲ <http://leb.net/archives/reader/1996/0138.html>

^۳ شرکت جیحون افزار نسخه بعدی برای این نرم افزار ارائه نداد.

۱-۳-۵- برخی ویژگیهای متون چاپی فارسی

نگارش فارسی ویژگیهای منحصر به فردی دارد که آن را کاملاً از نگارش لاتین متمایز می‌سازد [۸۲]. بمنظور فعالیت در زمینه OCR فارسی، آگاهی از قوانین نگارشی و نحوه چاپ حروف در این زبان امری ضروری است. در اینجا به ویژگیهای کلی نگارش فارسی اشاره می‌شود:

الف- متون فارسی برخلاف متون لاتین از چپ به راست نوشته می‌شود.

ب- در کلمات فارسی برخی از حروف از یک یا دو طرف به حروف مجاور خود اتصال داشته و برخی نیز بصورت مجزا نوشته می‌شوند. در نتیجه هر کلمه ممکن است شامل یک یا چند بخش متصل باشد که «زیرکلمه» نامیده می‌شوند (شکل ۱-۳-الف).

پ- حروف فارسی می‌توانند چهار موقعیت مجزا و در نتیجه چهار شکل متفاوت نگارش داشته باشند: حروف ابتدایی، میانی، انتهایی و مجزا (شکل ۱-۳-ب).

ت- حروف واقع در یک کلمه ممکن است همپوشانی داشته باشند، بدین معنی که نتوان با رسم خطوط عمودی، حروف را بطور کامل از یکدیگر مجزا نمود (شکل ۱-۳-ج).

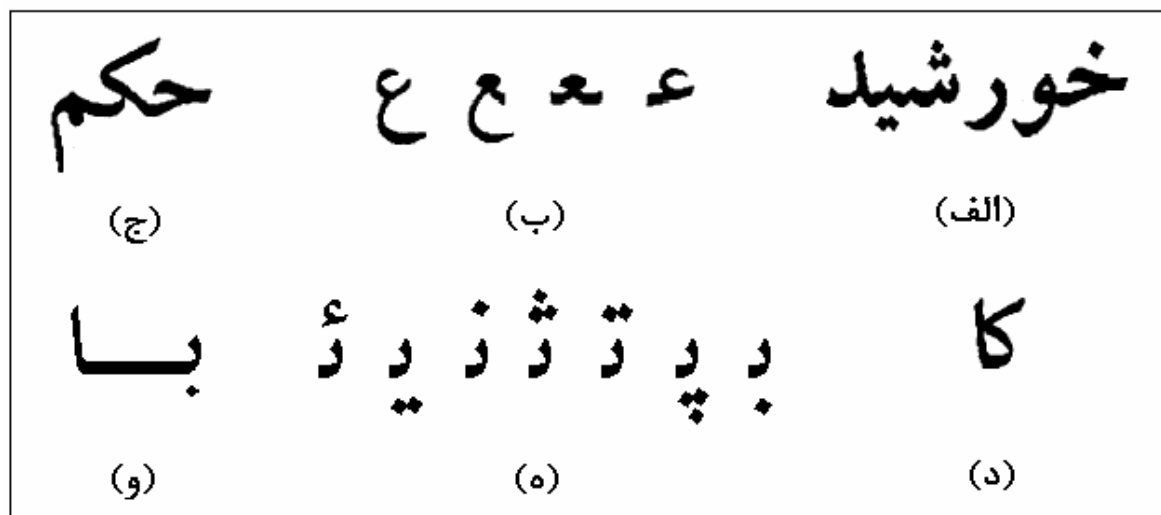
ث- در برخی از فونتها بعضی از حروف، در دو محل (از یک سمت) به یکدیگر اتصال دارند (شکل ۱-۳-د).

ج- برخی از حروف بین یک تا سه عدد نقطه دارند که ممکن است در بالا یا پایین بدنه حرف واقع باشند.

چ- بعضی از حروف بدنه مشابه دارند و تفاوت آنها تنها در تعداد و محل قرارگیری نقاط در آنهاست (شکل ۱-۳-ه).

ح- حروف فارسی می‌توانند در بالا یا پایین بدنه خود دارای اعراب باشند. سه اعراب - ، - ، و - در زبان فارسی، اعرابهای اصلی بوده و اعراب - در برخی کلمات عربی رایج در زبان فارسی دیده می‌شود (نظیر کلمات «عمداً» و «احتمالاً»). کلمات عربی دارای اعراب - و - در زبان فارسی عمومیت نیافته‌اند. هر چند کاربرد اعراب در زبان فارسی نسبت به زبان عربی بسیار محدودتر است، اما در

مواردیکه کلمه‌ای نامتداول باشد و یا بدلیل تشابه نگارشی آن با کلمه دیگر، تأکید بر تلفظ صحیح آن باشد، مورد استفاده قرار می‌گیرند.



شکل ۱-۳- برخی از ویژگیهای نگارش زبان فارسی (الف) کلمه خورشید از سه زیرکلمه تشکیل شده است (ب) چهار شکل مختلف حرف «ع» با توجه به موقعیت آن در کلمه (ج) همپوشانی دو حرف «ح» و «ک» در کلمه «حکم» (د) اتصال حروف «ک» و «ا» در دو محل (ه) تنوع حروف دارای بدنه مشابه (و) کشیدگی حرف «ب» در کلمه «با»

خ- در بالای بدنه یک حرف ممکن است علامت تشدید وجود داشته باشد.

د- برخی از حروف شامل همزه هستند ("ء"، "أ"، "ؤ"، "ة").

ذ- حروفی که از طرف چپ قابلیت اتصال به حرف مجاور خود را دارند، می‌توانند بصورت کشیده نوشته شوند (شکل ۱-۳-و).

۱-۴- آنالیز اسناد^۱

در بخشهای قبل، مبحث بازشناسی نوری حروف (OCR) مورد معرفی قرار گرفت. بازشناسی حروف خود زیرمجموعه‌ای از دسته بزرگی از تکنیکها بنام "آنالیز اسناد" (آنالیز مستندات) است که مبحث اصلی این طرح می‌باشد.

^۱ Document Analysis (DA) or Document Image Analysis (DIA)

به عنوان مثال فرض شود که از یک مجموعه مقالات یا پایان‌نامه‌ها، تنها نسخه چاپ شده یا کپی شده آنها موجود باشد و بخواهیم اطلاعات مختلف مورد نیاز نظیر عنوان، نام نویسندگان، نام مجله یا دانشگاه، سال انتشار، چکیده و سایر بخشهای آن را بصورت خودکار استخراج کرده و در یک پایگاه داده ذخیره نماییم، بگونه‌ای که امکان جستجو در هر یک از این فیلدها برای کاربر فراهم گردد. از طرف دیگر ممکن است هدف آن باشد که این مقالات یا مجلات را که به فرمتهای مختلفی تدوین شده‌اند، به یک قالب استاندارد تبدیل نموده و سپس در فرمت متداولی نظیر PDF بصورت تصویری (بدون نیاز به سیستم بازشناسی حروف) یا متنی ذخیره نماییم. برای این مقاصد لازم است سیستمی طراحی گردد که قادر به استخراج این اطلاعات از تصاویر اسکن شده اسناد باشد. مبحث آنالیز اسناد با مسائلی از این دست سر و کار دارد.

هدف از مبحث آنالیز اسناد (DIA یا DA)، شناسایی اجزاء متنی، گرافیکی و عکس در تصاویر اسناد و استخراج اطلاعات مورد نظر از آنها -مشابه با کاری که یک انسان انجام می‌دهد- می‌باشد [49]. آنالیز اسناد به عنوان یک زمینه تحقیقاتی، مشتمل بر کلیه مراحل پردازشی است که محتویات یک سند اسکن یا دورنگاری^۱ شده چند صفحه‌ای را به یک فرم الکترونیکی مناسب کد می‌کنند. این کد کردن می‌تواند چندین شکل داشته باشد: یک توصیف قابل ویرایش، یک نمایش فشرده که تصویر سند از آن قابل بازیابی باشد و یا یک توصیف معناشناختی سطح بالا که بمنظور پاسخگویی به پرس و جوها^۲ می‌تواند بکار رود [50]. در این راستا، سه دسته تکنیکهای پردازشی زیر مد نظر است (شکل ۱-۴):

الف- پردازش متن: این تکنیکها با اجزاء متنی تصویر سند سر و کار دارند. برخی از تکنیکهای این گروه عبارتند از: تعیین زاویه کجی سند، بازشناسی نوری حروف (OCR) و یافتن ستونها، پاراگرافها، خطوط متن و کلمات. این گروه از تکنیکها در بخش (۲-۵) مورد بررسی قرار می‌گیرند.

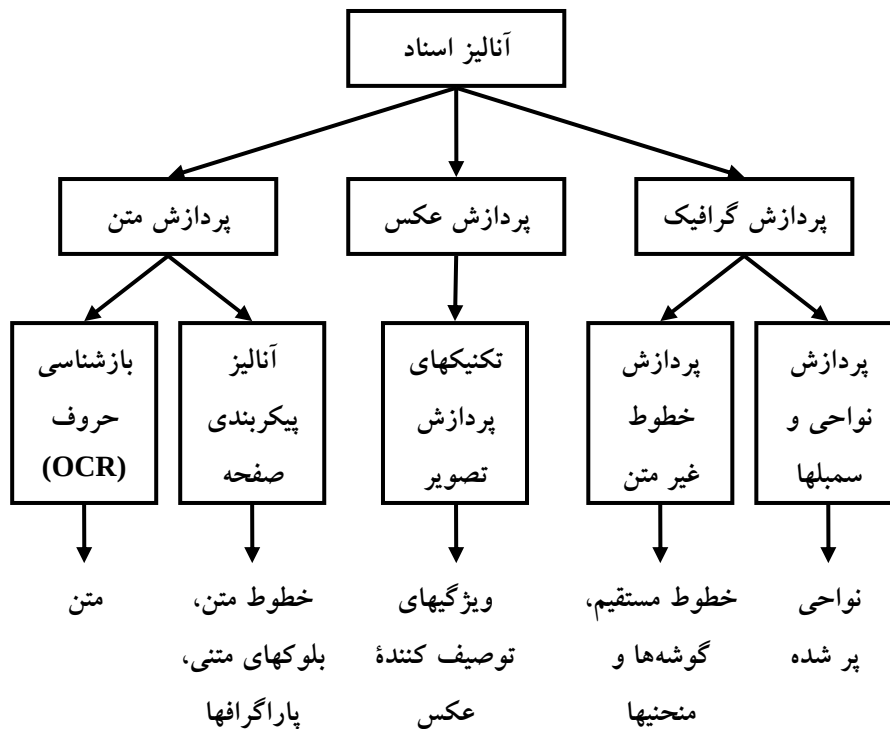
ب- پردازش گرافیک:^۳ تکنیکهای این دسته، با اجزاء خطی غیر متن و سمبلها که مواردی نظیر نمودارهای خطی، خطوط مستقیم مرزبندی بین قسمتهای متنی، آرم شرکتها و مواردی اینچنینی را

^۱ Facsimile

^۲ Queries

^۳ Graphical Processing or Graphics Processing

شامل می‌شوند، سر و کار دارند. از آنجائیکه بسیاری از اجزاء گرافیکی خطوط را شامل می‌شوند، این دسته از تکنیکهای پردازشی مشتمل بر نازک‌سازی خطوط، برازش خطی^۱ و آشکارسازی گوشه‌ها و منحنیها می‌باشد.



شکل ۱-۴- یک ساختار سلسله‌مراتبی از بخشهای مختلف مبحث آنالیز اسناد

ج- پردازش عکس^۲: عکسها سومین جزء اصلی اسناد هستند که البته بجز تشخیص موقعیت مکانی آنها در تصویر سند، آنالیزهای دیگر معمولاً به سایر تکنیکهای پردازش تصویر و بینایی ماشین مربوط می‌شود. مهمترین این پردازشها، فشردن سازی^۳ است.

مثالهای زیر علاوه بر مثالی که در ابتدای این فصل عنوان شد، ضرورت سرمایه‌گذاری بر روی مبحث آنالیز اسناد را بهتر نمایان می‌سازند [28]:

- با وجود آنکه امروزه بیشتر اسناد موجود در ادارات بوسیله کامپیوتر تولید می‌شوند، بواسطه بکارگیری نرم‌افزارهای مختلف سیستم‌عامل یا برنامه‌های کاربردی، فرمت نسخه الکترونیکی آنها در

^۱ Line Fitting

^۲ Picture Processing or Photographic Processing

^۳ Compression

همه موارد با یکدیگر سازگار نمی‌باشد. برخی از این اسناد شامل متن، جداول و حتی عبارات دستنویس نیز هستند. سیستمهای آنالیز اسناد بخشهای مختلف یک سند را تشخیص داده، بدین ترتیب امکان تبدیل فرمت سند به فرمت دلخواه (حتی مستقل از پلات فرم) فراهم می‌آید.

● دستگاههای نامه‌خوان اتوماتیک که کار مرتب‌سازی بسته‌ها و نامه‌های پستی را انجام می‌دهند، سالهاست که در پست‌خانه‌ها مورد استفاده قرار می‌گیرند؛ اما امروزه نیاز است که حجم بیشتری از نامه‌ها با دقت و سرعت بالاتر پردازش گردند.

● در یک کتابخانه سنتی، مسائلی چون مفقود شدن کتابها، بایگانی اشتباه، تعداد محدود هر یک از نسخه‌های موجود و حتی کاهش کیفیت کتابها در اثر گذشت زمان و استفاده مکرر، از جمله مشکلات متداول می‌باشد که می‌توان توسط تکنیکهای آنالیز مستندات راه حل مناسبی برای آنها یافت.

مبحث آنالیز اسناد در فصل ۵ مورد بررسی قرار می‌گیرد.

۱-۵- تعریف مسئله

نسخه الکترونیکی پایان‌نامه‌های مقاطع کارشناسی ارشد و دکترای انجام شده در دانشگاههای دولتی داخل کشور، بر روی وبسایت مرکز اطلاعات و مدارک علمی ایران (IRANDOC)^۱ موجود می‌باشند. فرایند تولید این اسناد بدین ترتیب است که ابتدا پایان‌نامه‌ها در چاپخانه مرکز برش داده شده و صفحات آن جدا می‌گردد. سپس این صفحات توسط اسکنرهای پرسرعت که بطور متوسط روزانه ۲۵ تا ۳۰ هزار صفحه را اسکن می‌نمایند، در حافظه کامپیوترهای متصل به اسکنرها در فرمت PDF^۲ بصورت تمام تصویر^۳ ذخیره شده (هر صفحه در یک فایل قرار می‌گیرد) و از آنجا به کامپیوتر سرور واقع در بخش خدمات ماشینی انتقال می‌یابند. پس از خاتمه مرحله اسکن، پایان‌نامه‌ها مجدداً صحافی می‌گردند. در بخش نمایه‌سازی، به هر پایان‌نامه یک شماره کاربرگ اختصاص داده می‌شود و اطلاعات آن شامل عنوان پایان‌نامه، نام نویسندگان، نام استاد راهنما، دانشگاه، سال انتشار و غیره به

^۱ <http://www.irandoc.ac.ir/>

^۲ Portable Document Format

^۳ Full image

همراه قسمت چکیده در درون یک فایل تایپ می‌گردد. پس از انتقال فایل به سرور، سرور چک می‌کند که برای یک پایان‌نامه آیا فایل PDFی با همان شماره کاربرگه وجود دارد یا خیر. اگر جواب مثبت باشد، فایل اطلاعات پایان‌نامه با فایل PDF (شامل متن کامل آن) به یکدیگر لینک شده و در پایگاه داده ذخیره می‌گردند. در غیر این صورت تنها فایل اطلاعات پایان‌نامه بر روی پایگاه داده قرار می‌گیرد تا در آینده فایل PDF پس از آماده شدن به آن ضمیمه گردد. ملاحظه می‌شود که در این سیستم تنها تصویر صفحات پایان‌نامه‌ها در دسترس کاربران قرار می‌گیرد و هیچ تلاشی برای شناسایی متن پایان‌نامه‌ها انجام نمی‌شود. بنابراین بجز اطلاعات معرفی و چکیده، هیچگونه اطلاعات متنی از پایان‌نامه‌ها در اختیار نمی‌باشد.

همانطور که گفته شد، هر یک از صفحات پایان‌نامه‌ها در یک فایل PDF تمام تصویر ذخیره شده‌اند. متأسفانه شماره اختصاص یافته به هر یک از این صفحات، با شماره واقعی صفحات پایان‌نامه‌ها مطابقت ندارد. دلیل این مسأله نیز آن است که در ابتدای هر پایان‌نامه صفحاتی نظیر عنوان، تقدیر و تشکر، تقدیم به ...، فهرست مطالب، فهرست جداول، فهرست نمودارها و غیره وجود دارند که یا شماره صفحه ندارند و یا شماره‌های آنها متفاوت از شماره صفحات اصلی می‌باشد. این مسأله سبب می‌شود که رجوع به یک صفحه خاص، با مقداری دشواری همراه باشد. چرا که به عنوان مثال کاربر شماره صفحه ۵۰ را وارد می‌کند، در حالی که شماره ۳۵ پایان‌نامه در معرض دید وی قرار می‌گیرد.

بخش فهرست مطالب (TOC)^۱ در هر سند (از جمله مجله، کتاب، پایان‌نامه) به نحو مناسبی ساختار منطقی آن را بیان می‌کند و توسط آن می‌توان مطالب مطروحه در سند و نحوه تنظیم آنها را با سرعت مشاهده نمود. بعلاوه کاربران با مشاهده این بخش براحتی می‌توانند مطالب مورد نظر خود را یافته و مستقیماً (بدون مرور صفحات قبل) به صفحه مورد نظر خود در سند پرش نمایند. برخی از پایگاه‌های داده دیجیتالی نیز از اطلاعات TOC بمنظور ساختارسنجی و شاخص‌گذاری^۲ مستندات بهره می‌جویند.

^۱ Table Of Contents (TOC)

^۲ Indexing

در نتیجه چنانچه این امکان فراهم آید که هر یک از عناوین موجود در بخش فهرست مطالب به عناوین مربوطه در سند لینک گردد، کاربر می‌تواند با کلیک کردن بر روی هر یک از عناوین مورد نظرش براحتی به آن صفحه پرش نماید. البته در برنامه Adobe Acrobat امکان لینک کردن یک بخش از سند به بخشی از آن سند یا سند دیگر بصورت دستی وجود دارد؛ اما انجام این کار برای حدود ۸۰ هزار (تا کنون) پایان‌نامه موجود بر روی وبسایت IRANDOC، بسیار وقتگیر و عملاً غیرممکن است.

یک روش عملی برای انجام این کار آن است که ابتدا توسط یک الگوریتم مناسب، بخش فهرست مطالب پایان‌نامه‌ها و مکان درج شماره صفحات در آنها را شناسایی کنیم و توسط یک الگوریتم بازشناسی حروف (OCR)، شماره صفحات را تشخیص دهیم. به همین طریق، شماره صفحات فصول اصلی پایان‌نامه‌ها را شناسایی می‌نماییم و سپس در مرحله بعد، بین شماره صفحات در بخش فهرست مطالب و صفحات اصلی لینک برقرار می‌کنیم. امکان‌سنجی این مسأله، هدف اصلی انجام این پروژه تحقیقاتی بوده است که در فصل ۶ به آن پرداخته می‌شود.

فصل دوم

تحلیل بخش‌های مختلف

سیستم‌های OCR

در این فصل سه بخش از پنج بخش اصلی یک سیستم OCR که در بخش (۱-۵) مورد اشاره قرار گرفت را تشریح نموده، به برخی از روشهای توسعه داده شده در مورد آنها برای متون لاتین و فارسی اشاره می‌نماییم. این سه بخش عبارتند از: مرحله پیش‌پردازش، مرحله قطعه‌بندی و مرحله استخراج ویژگیها یا بازنمایی. یک رویکرد تحلیلی از پایین به بالا^۱ آن است که روال بررسی را از سطح پیکسل شروع کرده و به یک متن معنی‌دار خاتمه دهیم.

۲-۱- پیش‌پردازش

این مرحله شامل کلیه پردازشهایی است که بر روی سیگنالهای تصویری خام انجام می‌شوند تا موجب تسهیل یا افزایش دقت روند اجرای فازهای بعدی گردند. از مجموعه این پردازشها، هدفهای زیر دنبال می‌شود [6]:

۱- کاهش نویز.

۲- نرمالیزه نمودن داده‌ها.

۳- فشرده‌سازی میزان اطلاعاتی که می‌بایست محفوظ بماند.

۲-۱-۱- کاهش نویز

نویز ایجاد شده بواسطه دستگاههای اسکنر نوری یا ابزارهای نگارشی منجر به ایجاد قطعه خطهای گسسته^۲، اتصال بین خطوط، فضاهای خالی در خطوط متن، پر شدن حفره‌های موجود در تصویر برخی حروف و غیره می‌گردد. همچنین اعوجاجهای مختلف شامل تغییرات محلی، منحنی شدن گوشه‌های حروف، تغییر شکل^۳ و یا خوردگی حروف^۴ را نیز بایستی مد نظر قرار داد. قبل از مرحله بازشناسی حروف لازم است که این نقایص برطرف شوند. تکنیکهای مختلف کاهش نویز را می‌توان به سه گروه اصلی دسته‌بندی نمود [6]:

^۱ Bottom-up approach

^۲ Disconnected line segments

^۳ Dilation

^۴ Erosion

الف - فیلتر کردن^۱

این روش به حذف نویز کمک می‌کند و ناصافیهای بدنه حروف را که معمولاً بوسیله سطوح نگارش ناهموار (در مورد متون دستنویس) و یا نرخ نمونه‌برداری ضعیف دستگاههای اخذ داده ایجاد می‌شوند، کاهش می‌دهد. فیلترهای حوزه مکانی یا فرکانسی متعددی را می‌توان برای این منظور طراحی کرد. ایده اصلی در این روش، کانولولو کردن^۲ یک ماسک از پیش تعریف شده با تصویر جهت تخصیص یک مقدار جدید به پیکسل برحسب تابعی از مقادیر پیکسلهای مجاور است. فیلترها را می‌توان برای مقاصد مختلفی چون هموارسازی^۳، شارپ کردن^۴، اعمال سطوح آستانه^۵، حذف پس-زمینه بافت‌گونه^۶ یا رنگی خفیف و تنظیم کنتراست (پادنمایی)^۷ طراحی نمود.

ب - عملگرهای مورفولوژی^۸

ایده اصلی مستتر در این عملگرها، فیلتر کردن تصویر متن با بکارگیری عملگرهای منطقی بجای عمل کانولولوشن است. عملگرهای مورفولوژی متعددی را می‌توان بمنظور متصل ساختن زیرحروف شکسته شده، جدا کردن زیرحرف متصل، هموارسازی کانتورها، تقلیل نقطه‌های خطا، نازک‌سازی^۹ حروف و استخراج مرزها طراحی نمود. در نتیجه عملگرهای مورفولوژی می‌توانند نویز موجود بر روی تصاویر اسناد بواسطه کیفیت پایین کاغذ و جوهر و همچنین حرکت نامنظم دست نویسنده را با موفقیت حذف نمایند.

ج - مدلسازی نویز^{۱۰}

در صورتیکه مدلی برای نویز موجود وجود داشته باشد، می‌توان توسط پاره‌ای از تکنیکهای کالیبره کردن، نویز را حذف نمود. لیکن مدلسازی نویز در اکثر کاربردها امکان‌پذیر نیست. در مورد

1 Filtering

2 Convolute

3 Smoothing

4 Sharpening

5 Thresholding

6 Textured background

7 Contrast

8 Morphological operations

9 Thinning

10 Noise Modeling

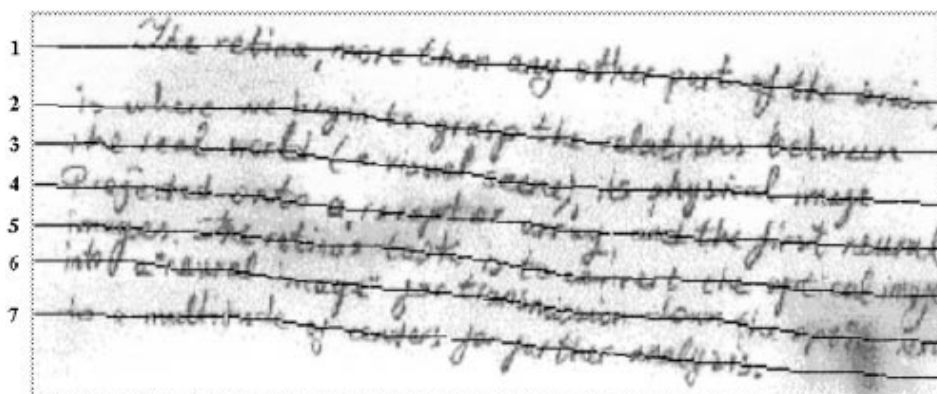
مدلسازی نویز ناشی از اختلالات اپتیکی نظیر ایجاد شدن لکه‌ها در تصویر اسکن شده، اریب شدن^۱ حروف و مات‌شدگی^۲ آنها تحقیقات بسیار کمی صورت گرفته است [6].

۲-۱-۲- نرمالیزه کردن^۳ داده‌ها

روشهای نرمالیزه کردن داده‌ها به حذف تغییرات نگارشی کمک نموده، داده‌های استاندارد شده‌ای را نتیجه می‌دهد. روشهای پایه نرمالیزه کردن عبارتند از:

الف- نرمالیزه کردن کجی^۴ متن و استخراج خطوط زمینه^۵

بدیل عدم دقت در مرحله اسکن و یا بی‌دقتی نویسنده در هنگام نگارش متن دستنویس، ممکن خطوط متن نسبت به تصویر اندکی انحراف یا چرخش داشته باشند (شکل ۱-۲). این مسئله می‌تواند کارایی الگوریتمهای بکار رفته در طبقات بعدی سیستم OCR را تأثیر قرار دهد؛ چرا که یکی از مفروضات بیشتر روشهای قطعه‌بندی، عدم کج بودن تصویر متن ورودی است و در نتیجه لازم است که این نقیصه آشکار و تصحیح گردد. آشکارسازی خط زمینه در بسیاری از تکنیکهای قطعه‌بندی و بازشناسی متون فارسی، عربی و لاتین نقش اساسی دارد. علاوه بر این، برخی از کاراکترها مانند «9» و «g» را می‌توان بواسطه موقعیت نسبی‌شان نسبت به خط زمینه آشکار ساخت.



شکل ۱-۲- استخراج خطوط زمینه در یک متن دستنویس دارای کجی

¹ Slant

² Blur

³ Normalization

⁴ Skew

⁵ Baselines

الگوریتمهای توسعه داده شده برای آشکارسازی کجی صفحه، تقریباً به همان تعداد الگوریتمهای باینری کردن تصویر (در ادامه بحث می‌شود) می‌باشند [46]. کلیه این روشها بر روی صفحات متنی با ترازبندی^۱ یکنواخت، دقیق عمل می‌کنند. الگوریتمی کاراتر است که بواسطه حضور مواردی در صفحه نظیر گرافیک، پاراگرافهای دارای کجی متفاوت، اعوجاجهای منحنی-خطی^۲ ظاهر شونده در کتابهای فوتوکپی شده، نواحی وسیع پیکسلهای سیاه نزدیک حاشیه صفحه و خطوط متنی مختصر و کوتاه، دقت آن کمتر دستخوش تغییر قرار بگیرد.

روشهای بکار رفته جهت تصحیح کجی خطوط زمینه در متون لاتین را می‌توان به چهار گروه اصلی دسته‌بندی کرد که عبارتند از [6 و ۷۹]:

۱- بکارگیری هیستوگرام^۳ (پروفایل تصویرنمایی^۴) تصویر.

۲- استفاده از روش خوشه‌بندی نزدیکترین همسایه‌ها^۵.

۳- روش همبستگی متقابل^۶ بین حروف.

۴- تبدیل هاف^۷.

در [۷۹] این چهار روش فوق مورد بررسی و مقایسه قرار گرفته‌اند که از میان آنها دو روش مبتنی بر همبستگی متقابل بین حروف و مبتنی بر تبدیل هاف بمنظور تصحیح کجی متون فارسی انتخاب و پیاده‌سازی شده‌اند. در [۷۸] دو روش برای تشخیص و تصحیح کجی اسناد فارسی و لاتین معرفی گردیده است. ایده روش اول بر این اصل شکل گرفته است که در یک متن فارسی یا عربی تعداد زیادی حرف «الف» مجزا و در یک متن انگلیسی به میزان کافی حروف «i» و «l» وجود دارد که راستای این حروف قائم بر راستای متن می‌باشد. لذا با شناسایی این حروف می‌توان راستای صحیح متن را تشخیص داد. در روش دوم، راستایی که در آن راستا مستطیل محیطی یک جزء متصل از

^۱ Alignment

^۲ Curvilinear Distortion

^۳ Histogram

^۴ Projection profile

^۵ Nearest neighbors clustering

^۶ Cross correlation

^۷ Hough transform

حروف درون یک خط از متن دارای کمترین مساحت باشند به عنوان جهت کجی در نظر گرفته می-شود.

اغلب پس از آشکارسازی کجی، تصویر صفحه به جهت اصلی چرخانده می شود تا عملیات تحلیل فرمت بندی متن و OCR بتواند با سهولت و دقت بیشتری انجام پذیرد. نمونه برداری مجدد^۱ مورد نیاز برای این منظور که بایستی بر روی صفحات باینری شده اعمال شود، ممکن است پترن کاراکترها را تغییر دهد. در این حالت بجای چرخاندن تصویر می توان الگوریتمهای پردازشی را به نحوی اصلاح نمود که اثر چرخش در آنها لحاظ گردد [26]. همچنین می توان تصویر سند را قبل از باینری کردن چرخش داد یا اینکه مقدار چرخش را از روی انتقالهای کوچک و بدون اعوجاج کل بلوکهای متنی تقریب زد [46].

ب- نرمالیزه کردن اریب شدگی^۲

در متون چاپی فارسی و لاتین، کاراکترهای دارای فرمت ایتالیک از راستای عمود انحراف دارند. همچنین در متون دستنویس برخی از نویسنده ها حروف را بصورت زاویه دار می نویسند. این پدیده تحت عنوان «اریب شدگی» شناخته می شود و می تواند دقت برخی از الگوریتمهای قطعه بندی یا بازشناسی را تحت تأثیر قرار دهد و لذا در این سیستمها لازم است که در مرحله پیش پردازش، میزان اریب بودن کاراکترها شناسایی و تصحیح گردد.

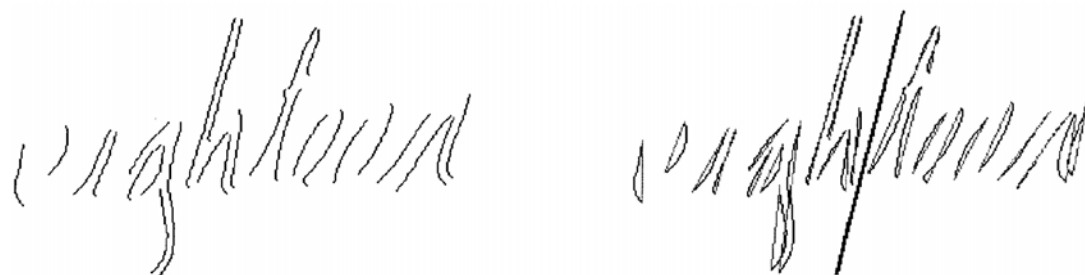
اریب شدگی بصورت زاویه شیب بین طویلترین زیرحرف^۳ در یک کلمه و جهت عمودی تعریف می شود. نرمالیزه کردن اریب، بنظر نرمالیزه نمودن کلیه کاراکترها به یک فرم استاندارد بکار می رود. معمولترین روش در تخمین میزان اریب شدگی، محاسبه زاویه متوسط اجزاء نزدیک به خط عمود است (شکل ۲-۲). در [35]، استخراج خطوط عمودی از کاراکترها بوسیله دنبال کردن مؤلفه های کد زنجیره ای توسط یک جفت فیلتر یک بعدی انجام می پذیرد. مختصات شروع و پایان هر خط، زاویه اریب را بدست می دهد. مرجع [22] از روشی استفاده می کند که در آن پروفایلهای تصویرنمایی برای تعدادی از زوایا که از جهت عمود انحراف دارند محاسبه می شوند. زاویه متناظر با

¹ Resampling

² Slant Normalization

³ Stroke

تصویرنمایی که از بزرگترین مشتق مثبت برخوردار است، بمنظور تشخیص زاویه اریب غالب بکار می‌رود. در [11] آشکارسازی میزان اریب‌شدگی، از طریق تقسیم تصویر به پنجره‌های افقی و عمودی انجام می‌شود. در این تحقیق، بر اساس مرکز جرم نیمه بالایی و پایینی هر پنجره، اریب‌شدگی تخمین زده می‌شود. بالاخره در [30] گونه‌ای از تبدیل هاف با اسکن کردن چپ به راست تصویر و محاسبه تصویرسازی در ۲۱ زاویه اریب مختلف بکار گرفته شده است. سه تصویرنمایی بالایی مربوط به هر زاویه اریب با یکدیگر جمع می‌شوند و زاویه متناظر با بیشترین مقدار این حاصل جمع، زاویه اریب قلمداد می‌گردد. از سوی دیگر در برخی از مطالعات انجام شده، سیستمهای بازشناسی زاویه اریب را تصحیح نمی‌کنند و در مرحله آموزش این نقیصه جبران می‌گردد.



شکل ۲-۲- تخمین زاویه اریب با ملاحظه الف- اجزاء نزدیک به خط عمودی ب- زاویه

شیب متوسط [6]

ج- نرمالیزه کردن (تغییر مقیاس دادن)^۱ اندازه

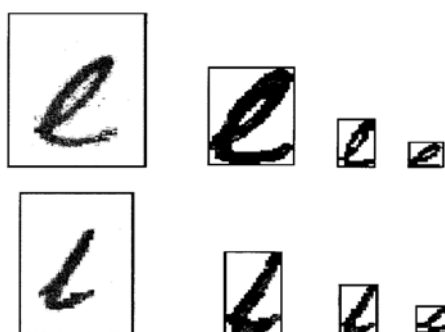
در سیستمهای OCR اغلب تصاویر کلمات خیلی کوچک یا خیلی بزرگ، به یک اندازه استاندارد نرمالیزه می‌شوند. این عمل معمولاً با نمونه‌برداری مجدد^۲ تصویر انجام می‌گیرد. روشهایی نظیر Bilinear یا Bicubic بر روی تصاویر سطح خاکستری بنحو مناسبی عمل می‌کنند؛ اما عملیات نمونه‌برداری موجود در آنها، کاراکترهای دوسطحی را دچار اعوجاج می‌سازند. یک راه‌کار استاندارد، اعمال یک روال دو مرحله‌ای است که در آن ابتدا کانتور هموار اصلی مربوط به کاراکتر نمونه‌برداری

^۱ Scaling

^۲ Resampling

شده، توسط یک فیلتر کانولوشن وزن دار^۱ و کوانتیزاسیون دوسطحی دامنه^۲ تقریب زده می شود و پس از آن عمل نمونه برداری مجدد انجام می گیرد [46].

روشهای بازشناسی حروف ممکن است نرمالیزه کردن اندازه را در هر دو جهت افقی و عمودی انجام دهند. در [61] هر کاراکتر به تعدادی ناحیه تقسیم می شود و هر یک از این نواحی، بصورت جداگانه تغییر مقیاس داده می شوند. همچنین می توان نرمالیزه کردن اندازه را به عنوان قسمتی از مرحله آموزش انجام داد و پارامترهای اندازه را برای هر یک از داده های آموزشی منفرد، جداگانه تخمین زد [4]. در شکل (۲-۳) دو کاراکتر نمونه بصورت تدریجی به اندازه بهینه کوچک می شوند که بدین ترتیب نرخ بازشناسی در داده های آموزشی حداکثر می گردد. از طرف دیگر در روشهای بازشناسی کلمات بصورت واحد، طول کلمات می تواند کمک شایانی به بازشناسی صحیح نماید و مطلوب است که با حفظ طول کلمات، اختلاف بین کلاسی حداکثر گردد. در نتیجه در اینجا تنها ارتفاع عمودی کلمات نرمالیزه می شود و یا اینکه نرمالیزاسیون اندازه افقی بر مبنای ضریب مقیاس محاسبه شده برای نرمالیزاسیون اندازه عمودی انجام می گیرد [30].



شکل ۲-۳- نرمالیزه کردن حروف e و l

د- هموارسازی کانتور^۳

در متون دستنویس، بواسطه لرزش یا حرکات ناخواسته دست نویسنده هنگام نگارش، ممکن است که کانتور حروف شکل ناصاف پیدا کند. همچنین این مسئله می تواند در سیستمهای بازشناسی متون چاپی و دستنویس بدلیل تغییر مقیاس حروف و یا وجود نویز در مرحله اسکن تصاویر نیز ظاهر

^۱ Weighted Convolution Filter

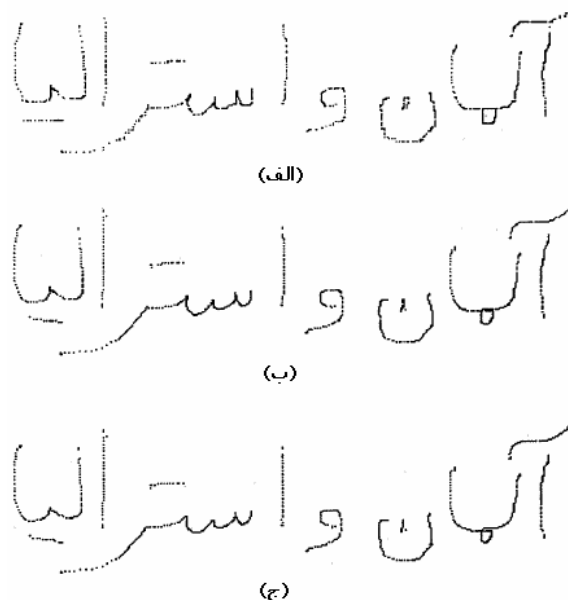
^۲ Bilevel Amplitude Quantization

^۳ Contour Smoothing

گردد. روشهای هموارسازی کانتور بمنظور جبران این نقیصه مورد استفاده قرار می گیرند. بطور کلی هموارسازی کانتور تعداد نقاط نمونه مورد نیاز برای بازنمایی کاراکتر را کاهش می دهد و در نتیجه کارایی مراحل پردازشی باقیمانده را بهبود می بخشد. در [۸۷] مقدار هر پیکسل از تصویر متن با مقدار میانگین وزنی پیکسلهای همسایه آن جایگزین می شود که با دو بار تکرار این عمل، تصویر هموارتری از متن دستنویس بدست می آید و در نتیجه اثر لرزش دست نویسنده کاهش پیدا می کند (شکل ۲-۴).

۳-۱-۲- فشرده سازی

این مسئله پذیرفته شده است که تکنیکهای کلاسیک فشرده سازی تصاویر که تصویر را از حوزه مکانی به حوزه های دیگر منتقل می کنند، برای بازشناسی حروف مناسب نمی باشند. در بازشناسی حروف، عمل فشرده سازی نیازمند آن دسته از تکنیکهای حوزه مکانی است که اطلاعات شکلی را حفظ می نمایند. دو تکنیک متعارف فشرده سازی، یکی تکنیک اعمال سطح آستانه^۱ (بمنظور باینری کردن تصاویر سطح خاکستری متون) و دیگری نازک سازی^۲ می باشند [6].



شکل ۲-۴- اعمال یک الگوریتم هموارسازی بر متن دستنویس الف- تصویر اصلی ب-

یک مرحله هموارسازی ج- دو مرحله هموارسازی [۸۷]

^۱ Thresholding

^۲ Thinning

الف - باینری (دوسطحی) کردن تصویر متن

بمنظور کاهش حجم ذخیره‌سازی مورد نیاز و افزایش سرعت پردازش، اغلب مطلوب است که با انتخاب یک سطح آستانه، تصاویر سطح خاکستری یا رنگی را به تصاویر باینری تبدیل نمود. دو روش اعمال سطح آستانه عبارتند از: سراسری^۱ و محلی^۲. در روش اعمال سطح آستانه سراسری، یک مقدار سطح آستانه برای کل تصویر سند انتخاب می‌شود. این مقدار اغلب بر مبنای تخمینی از سطح پس‌زمینه که از هیستوگرام سطح روشنایی تصویر محاسبه می‌گردد، سنجیده می‌شود. روش اعمال سطح آستانه محلی (تطبیقی) بر اساس اطلاعات نواحی محلی، از مقادیر متفاوتی برای هر پیکسل استفاده می‌کند. در [55] مقایسه‌ای بین تکنیکهای معمول اعمال سطح آستانه بصورت سراسری و محلی با استفاده از یک معیار ارزیابی^۳ انجام پذیرفته است. این معیار ارزیابی عبارت است از مقایسه دقت بازشناسی صحیح یک سیستم بازشناسی حروف با اعمال تکنیکهای مختلف سطح آستانه‌گذاری. در این بررسی نشان داده شده است که روش تطبیقی محلی Niblack [47] بهترین نتیجه را حاصل می‌نماید. اضافه بر این، بررسی اخیر در [60] یک روش تطبیقی منطقی را معرفی می‌کند که از طریق تحلیل ویژگیهای خوشه‌بندی و اتصالات کاراکترها در تصاویر مخدوش شده عمل می‌نماید.

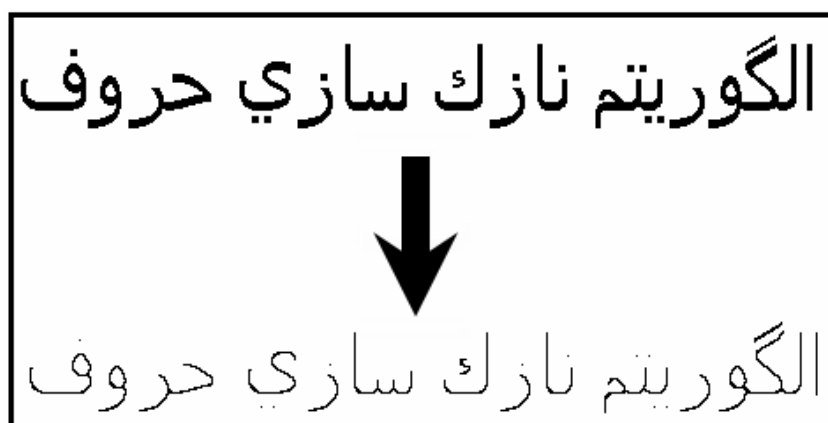
در سیستمهای OCR روال استخراج ویژگیها غالباً بر روی تصاویر باینری انجام می‌گیرد. لیکن باینری کردن یک تصویر سطح خاکستری ممکن اطلاعات توپولوژیکی مهمی را از کاراکتر ساقط نماید. بمنظور اجتناب از این مسئله، در بعضی از تحقیقات سعی گردیده که ویژگیها مستقیماً از تصاویر سطح خاکستری حروف استخراج گردند [56].

ب - نازک‌سازی^۴

این عمل درحالیکه کاهش قابل ملاحظه‌ای در حجم داده‌ها ایجاد می‌کند، اطلاعات شکلی کاراکتر را نیز استخراج می‌نماید. دو روش پایه برای نازک‌سازی عبارتند از: «نازک‌سازی از طریق پیکسل»^۵ و «نازک‌سازی غیر از طریق پیکسل». نازک‌سازی از طریق پیکسل بصورت محلی و تکراری

^۱ Global^۲ Local^۳ Evaluation criterion^۴ Thinning^۵ Pixel wise thinning

تصویر را مورد پردازش قرار می‌دهد تا وقتی که از تصویر کاراکتر تنها اسکلت آن به عرض یک پیکسل باقی بماند. این روش نسبت به نویز بسیار حساس بوده، ممکن است تصویر کاراکتر را مخدوش سازد. از سوی دیگر، روشهای نازک‌سازی غیر از طریق پیکسل، طی فرایند نازک‌سازی مقداری از اطلاعات سراسری درباره کاراکتر را مورد استفاده قرار می‌دهند. این روشها یک خط مرکزی یا میانه بخصوص از تصویر پترن را بدون آزمایش همه پیکسلها تولید می‌نمایند. بررسی الگوریتمهای نازک‌سازی از طریق پیکسل و غیر از طریق پیکسل در مرجع [32] انجام گرفته است. عملیات تکرار در نازک‌سازی می‌تواند بصورت الگوریتمهای ترتیبی یا موازی انجام پذیرد. الگوریتمهای ترتیبی نقاط کانتور را بوسیله اسکن در راستاهای افقی یا با دنبال کردن کانتور مورد پردازش قرار می‌دهند. الگوریتمهای موازی نسبت به الگوریتمهای ترتیبی برتری دارند؛ زیرا با استفاده از یک مجموعه شرایط یکسان برای عمل حذف، کل پیکسلها را بصورت همزمان آزمایش می‌کنند و می‌توان آنها با سخت‌افزارهای موازی پیاده‌سازی کرد. برآوردی از الگوریتمهای نازک‌سازی موازی برای بازشناسی حروف را می‌توان در مرجع [33] یافت. شکل (۲-۵) یک متن به همراه تصویر نازک شده آن را نشان می‌دهد.



شکل ۲-۵- اعمال عملیات نازک‌سازی بر روی یک تصویر متنی نمونه

تکنیکهای پیش‌پردازشی علاوه بر بازشناسی حروف، در بسیاری از زمینه‌های پردازش تصویر مورد بررسی و پژوهش قرار گرفته‌اند. توجه شود که تکنیکهای فوق سبب تغییر داده‌ها می‌شوند و در نتیجه امکان دارد که اعوجاجهای دور از انتظاری را متوجه تصویر سند سازند. بنابراین ممکن است

این تکنیکها سبب از دست رفتن برخی اطلاعات مهم گردند. در نتیجه در بکارگیری آنها می‌بایست دقت کافی مبذول داشت [6].

۲-۱-۴- بازشناسی خط^۱، زبان و فونت

بازشناسی خط، تعداد کلاسهای مختلف سمبل که بایستی مورد ملاحظه قرار گیرند را کاهش می‌دهد. شناسایی زبان متن، بمنظور بکارگیری مدل‌های متنی^۲ خاص ضرورت دارد. طبقه‌بندی فونتها، تعدد شکل‌های مختلف حروف در هر کلاس که می‌بایست در فرایند بازشناسی لحاظ گردد را کاهش می‌دهد و سبب می‌شود که امر شناسایی، تنها به یک کلاس فونت محدود گردد. بازشناسی خط و زبان و فونت در کاربردهایی مانند نمایه‌سازی^۳ و دستکاری اسناد نیز مطلوب می‌باشد. در [46] چهار مقاله مرتبط با این مبحث مورد اشاره قرار گرفته‌اند.

۲-۲- قطعه‌بندی^۴

مرحله پیش‌پردازش یک تصویر اصلاح شده از سند را نتیجه می‌دهد بگونه‌ای که مقدار کافی از اطلاعات شکلی، فشرده‌سازی بالا و نویز نویز پایین، از تصویر نرمالیزه شده سند قابل حصول است. قطعه‌بندی (شکل ۲-۶) یک مرحله بسیار با اهمیت در بازشناسی حروف مخصوصاً حروف فارسی و عربی که بصورت پیوسته نوشته می‌شوند می‌باشد؛ چرا که نتیجه بدست آمده از جداسازی کلمات، خطوط یا کاراکترها مستقیماً بر روی نرخ بازشناسی سیستم تأثیر می‌گذارد.

قطعه‌بندی غلط کاراکترها، عامل بسیاری از خطاهای OCR است (مانند: $rn \rightarrow m$ یا $m \rightarrow rn$). میزان دقت یک الگوریتم قطعه‌بندی به سبک نگارش حروف، کیفیت دستگاه پرینت (کاراکترهای ایتالیک لکه شده، برای قطعه‌بندی مشکل می‌باشند) و نیز نسبت اندازه فونت به رزولوشن دستگاه اسکنر (تابع گسترش نقاط^۵ و نرخ نمونه‌برداری مکانی) بستگی دارد [46].

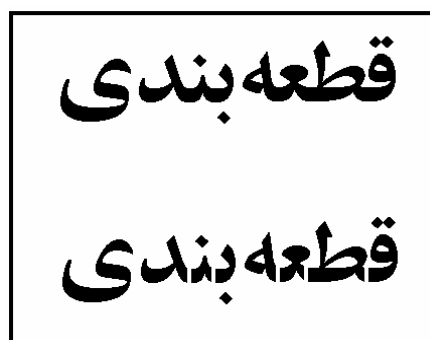
¹ Script

² Context models

³ Indexing

⁴ Segmentation

⁵ Point-spread Function



شکل ۲-۶- قطعه‌بندی یک کلمه به حروف

دو نوع قطعه‌بندی وجود دارد [6]:

الف) قطعه‌بندی بیرونی^۱ که عبارت است از جداسازی بخشهای مختلف نگارش مانند پاراگرافها، جملات یا کلمات.

ب) قطعه‌بندی درونی^۲ که منظور از آن، جداسازی حروف کلمات مخصوصاً در مورد کلمات سرهم^۳ نوشته شده در متون لاتین و یا رسم‌الخطهای پیوسته نظیر فارسی و عربی است.

در مواردی همچون متون فارسی که حروف بصورت سرهم نوشته می‌شوند، سه رویکرد مختلف در بازشناسی برون‌خط متون کلمات یا زیرکلمات وجود دارد [۸۲ و ۸۶ و ۶۸]:

الف) رویکرد مبتنی بر قطعه‌بندی کلمات^۴.

ب) رویکرد مبتنی بر بازشناسی کلمه به عنوان یک الگوی واحد^۵.

ج) رویکرد ترکیبی.

در رویکرد بازشناسی مبتنی بر جداسازی، ابتدا کلمه در مرحله جداسازی به حروف یا زیرحروف شکسته می‌شود؛ آنگاه قطعات جدا شده بازشناسی می‌شوند و از کنار هم قرار گرفتن آنها کلمه شناسایی خواهد شد. روشهای بکارگرفته شده در این رویکرد به دو گروه مختلف تقسیم می‌شوند:

- تقطیع کلمه به حروف.

- تقطیع کلمه به زیرحروف.

^۱ External Segmentation

^۲ Internal Segmentation

^۳ Cursive

^۴ Segmentation-based approach

^۵ Segmentation-free approach

در گروه اول کلمه به حروف جداسازی می‌شود و با شناسایی حروف جدا شده، کلمه بازشناسی می‌گردد. رویکرد بکار گرفته شده در این گروه را رویکرد مبتنی بر جداسازی و بازشناسی حروف می‌نامند. در گروه دوم کلمه به زیرحروف مثل پاره‌منحنی‌ها و ساختارهای پایه دیگر جداسازی می‌شود و با شناسایی زیرحروف و ترکیب آنها کلمه بازشناسی می‌گردد. در رویکرد بکار گرفته شده در گروه دوم که اصطلاحاً جداسازی و بازشناسی توأم نامیده می‌شود، نمی‌توان در ابتدا مرز حروف را بطور کامل مشخص کرد. بلکه حروف از ابتدا به انتهای کلمه به ترتیب بازشناسی و جداسازی می‌شوند. در هیچکدام از دو شکل رویکرد مبتنی بر جداسازی، به شکل کلی کلمه توجهی نمی‌شود و سعی بر آن است که با بازشناسی حروف یک کلمه، آن کلمه شناخته شود.

در رویکرد بازشناسی کلمه به عنوان یک الگوی واحد، تلاشی برای تقطیع کلمه به حروف و بازشناسی حروف موجود در کلمه صورت نمی‌گیرد و کلمه در قالب یک الگو بررسی می‌گردد. روشهای بکار گرفته شده در این رویکرد را می‌توان به دو گروه تقسیم کرد: گروه اول روشهایی هستند که تصویر کلمه را بصورت یک الگوی دو سطحی (باینری)^۱ یا با سطوح خاکستری^۲ در نظر گرفته و از روشهای معمول در شناسایی تصاویر که معمولاً مبتنی بر توابع فاصله می‌باشند و یا روشهایی همچون شبکه‌های عصبی و غیره استفاده می‌نمایند. گروه دوم روشهای مبتنی بر پردازش سیگنالهای تصادفی هستند. این روشها ابتدا در بازشناسی گفتار بکار گرفته شده و سپس در زمینه بازشناسی کلمات مخصوصاً کلمات دستنویس لاتین مورد استفاده قرار گرفته‌اند. نمونه‌ای از این روشها، الگوریتمهای مبتنی بر مدل مخفی مارکوف (HMM)^۳ هستند. رویکرد (ج) ترکیبی از دو رویکرد (الف) و (ب) می‌باشد.

نتیجه مطلوب مرحله قطعه‌بندی تصویری است که تنها حاوی یک کاراکتر باشد و بجز پیکسلهای پس‌زمینه، هیچ شیئی دیگری را شامل نمی‌شود. لیکن هنگامی که اشیاء چاپی در تصویر ورودی خیلی نزدیک به هم ظاهر شوند (مانند نقشه‌های هیدروگرافی)، این منظور همواره قابل حصول نخواهد بود. در چنین حالتی اغلب سایر کاراکترها یا اشیاء چاپی بطور تصادف در داخل

^۱ Binary^۲ Gray levels^۳ Hidden Markoff Model

تصویر کاراکتر قرار می‌گیرند که احتمالاً ویژگیهای استخراج شده را تحریف می‌نمایند. این مورد یکی از دلایلی است که بیان می‌دارد چرا هر سیستم بازشناسی حروف، یک انتخاب وازدگی^۱ دارد [56].

۲-۱-۲-۲- قطعه‌بندی بیرونی

این مرحله بحرانی‌ترین و حساس‌ترین قسمت در زمینه آنالیز اسناد^۲ می‌باشد و یک مرحله ضروری برای سیستمهای بازشناسی حروف برون‌خط محسوب می‌شود. گرچه مبحث آنالیز اسناد با روشها و تکنیکهای خاص خود یک حوزه تحقیقاتی تا حدی متفاوت نسبت به OCR است، لیکن تقسیم‌بندی تصویر سند به نواحی متنی و غیر متنی، یک بخش لاینفک نرم‌افزارهای OCR به حساب می‌آید. بنابراین برای افرادی که در زمینه OCR تحقیق می‌نمایند، یک دانش عمومی از تکنیکهای آنالیز اسناد ضرورت دارد.

آنالیز پیکربندی صفحات در دو مرحله انجام می‌گیرد: مرحله اول آنالیز ساختاری^۳ است که با قطعه‌بندی تصویر به بلوکهایی از تصویر سند (نظیر پاراگراف، ردیف، کلمه و غیره) مرتبط می‌باشد و مرحله دوم که آنالیز عملکردی^۴ نام دارد، با استفاده از اطلاعات مکان، اندازه و قوانین صفحه‌بندی مختلف، عملکرد هر یک از اجزاء سند نظیر عنوان، چکیده و غیره را تعیین می‌نماید.

تعدادی از روشها، نواحی همگن در تصویر سند مانند نواحی بافت‌گونه^۵ را مورد توجه قرار می‌دهند [6]. در این صورت عمل قطعه‌بندی صفحه، از طریق یافتن نواحی بافت‌گونه در تصاویر سطح خاکستری یا رنگی انجام می‌پذیرد. به عنوان مثال، Jain و همکارانش از فیلتر گابور^۶ و کانوولو کردن ماسکها استفاده می‌کنند. روش Tang و همکارانش مبتنی بر علائم فرکتالی^۷ است و الگوریتم Doerman آنالیز چند مقیاسه موجک^۸ را بکار می‌گیرد. بسیاری از روشهای قطعه‌بندی تصاویر برای تعیین نواحی همگن، توجه خود را به پردازش پیکسلهای پس‌زمینه معطوف می‌دارند و یا فضاهای

¹ Reject Option

² Document analysis

³ Structural analysis

⁴ Functional analysis

⁵ Textured regions

⁶ Gabor

⁷ Fractal signature

⁸ Wavelet multiscale analysis

سفید موجود در صفحات را مورد استفاده قرار می‌دهند. این تکنیکها شامل درخت تجزیه $X-Y$ ، پروفایل تصویرنمایی مبتنی بر پیکسل، دنبال کردن فضاهای سفید و نازک‌سازی فضاهای سفید می‌باشند [6]. می‌توان این روشها را روشهای از بالا به پایین به حساب آورد؛ زیرا بصورت بازگشتی صفحه را از طریق برش در امتداد محورهای X و Y قطعه‌بندی می‌کنند. از سوی دیگر تعدادی روش از پایین به بالا نیز وجود دارند که بر مبنای پردازش پیکسلها و اجزاء پیوسته، نواحی همگن را از اجزاء کوچک به نواحی بزرگتر رشد می‌دهند. یک مثال این گروه، روش Dostrum است که از تکنیک کلاستر بندی k نزدیکترین همسایه بهره می‌گیرد [49]. برخی از تکنیکها نیز هر دو گروه تکنیکهای بالا به پایین و پایین به بالا را ترکیب می‌نمایند. یک بررسی مختصر از کارهای انجام شده در زمینه تحلیل صفحات را می‌توان در [26] ملاحظه نمود.

۲-۲-۲- قطعه‌بندی درونی

علیرغم فعالیتهای چشمگیر دهه گذشته و تنوع تکنیکهای معرفی شده، قطعه‌بندی متون پیوسته (بخصوص متون دستنویس پیوسته) به حروف هنوز هم بصورت یک مسئله باقی مانده است. روشهای قطعه‌بندی حروف به سه دسته تقسیم می‌شوند:

الف - قطعه‌بندی صریح^۱

در این روش قطعات بر مبنای خاصیت حرف‌سان بودنشان (شباهت داشتن به یک حرف) تشخیص داده می‌شوند. فرایند قطعه‌بندی تصویر به اجزاء بامعنی یک نام خاص دارد: تشریح^۲. تشریح فرایندی است که یک تصویر را بدون بکارگیری یک کلاس بخصوص از اطلاعات شکلی تحلیل می‌نماید. معیار سنجش خوب بودن قطعه‌بندی، همخوان بودن خواص عمومی قطعات با خصوصیات مورد انتظار برای کاراکترهای معتبر است. تکنیکهای موجود مبتنی بر تشریح یک تصویر عبارتند از [6]: استفاده از فضاهای خالی و نقاط اوج، آنالیز تصویرنمایی عمودی^۳، آنالیز اجزاء

^۱ Explicit Segmentation

^۲ Dissection

^۳ Vertical projection analysis

پیوسته^۱ و علائم مشخصه^۲. علاوه بر این، می‌توان با بکارگیری قراین زبان‌شناختی، قطعه‌بندی صریح را در معرض ارزشیابی قرار داد.

ب - قطعه‌بندی ضمنی^۳

این روش قطعه‌بندی بر پایهٔ بازشناسی است و تصویر سند را برای یافتن اجزایی که با کلاسهای از قبل تعریف شده‌ای مطابقت داشته باشند، مورد جستجو قرار می‌دهد. قطعه‌بندی، بر مبنای میزان صحت نحوی یا معنایی نتایج کلی حاصله از بازیابی انجام می‌گیرد. در این رویکرد، دو گروه از روشها قابل بکارگیری می‌باشند: (۱) روشهایی که چندین روال جستجو را انجام می‌دهند و (۲) روشهایی که بردار ویژگی تصویر را قطعه‌بندی می‌نمایند.

کلاس نخست سعی در قطعه‌بندی کلمات به حروف یا سایر اجزاء بدون استفاده از الگوریتمهای تشریح مبتنی بر ویژگی دارد. به عبارت دقیقتر، در این روش بدون در نظر گرفتن محتوا، تصویر بطور اصولی به قسمت‌های همپوشان تقسیم می‌شود. از دیدگاه مفهومی، این دسته روشها از تکنیکهایی که برای بازشناسی کلمات چاپی توسعه داده شده‌اند، نشأت می‌گیرند. قائدهٔ اصلی آنها بکارگیری یک پنجرهٔ متحرک با عرض متغیر است که سلسله‌ای از قطعه‌بندیهای آزمایشی را که بایستی توسط سیستم بازشناسی حروف مورد تأیید قرار گیرند، ایجاد می‌نمایند. تکنیک دیگری، برنامه‌نویسی دینامیک و شبکه‌های عصبی را با یکدیگر تلفیق می‌کند. سرانجام در در روش رسیدگی انتخابی^۴ از شبکهٔ عصبی در مهار نمودن مشکلات قطعه‌بندی استفاده می‌گردد.

کلاس دوم از روشهای قطعه‌بندی ضمنی، تصویر را با کلاسه‌بندی زیرمجموعه‌هایی از ویژگیهای مکانی (استخراج شده از کل تصویر) بطور ضمنی قطعه‌بندی می‌نماید. این دسته از روشها را به دو گروه می‌توان تقسیم‌بندی نمود: روشهای مبتنی بر HMM و روشهای غیر متکی بر مدل مارکوف. در [14] و [42] از HMMها بمنظور سازماندهی کل فرایند بازشناسی کلمات استفاده گردیده است. روشهای غیر مارکوف از مفاهیم مورد استفاده در بینایی ماشین برای بازشناسی اشیاء

¹ Connected component analysis

² Landmarks

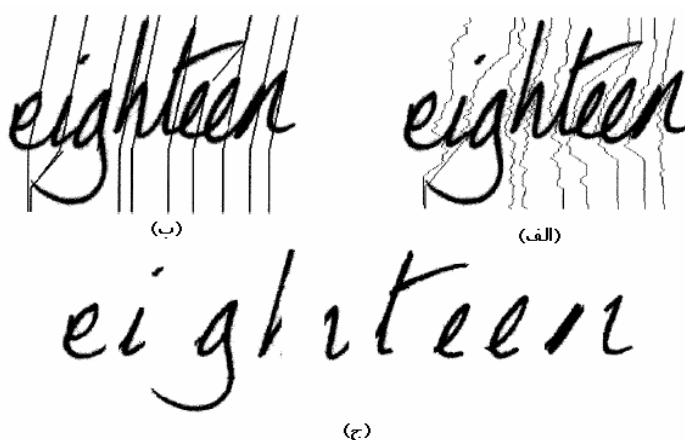
³ Implicit Segmentation

⁴ Selective attention

پنهان شده نشأت می‌گیرند. این خانواده از روشهای مبتنی بر بازشناسی، از تکنیکهای ملایم‌سازی احتمالاتی، مفهوم نقاط منتظم و نقاط تکین، و تطابق بازگشتی استفاده می‌نمایند.

ج) تکنیکهای ادغام شده

این تکنیکها، روشهای قطعه‌بندی صریح و ضمنی را در هم می‌آمیزند. یک الگوریتم تشریح با هدف قطعه‌بندی اضافی به تصویر اعمال می‌شود؛ بدین معنی که تصویر آنقدر به نواحی مختلف برش داده می‌شود تا اطمینان حاصل گردد که مرزهای قطعه‌بندی صحیح نیز در میان قطعات ایجاد شده موجود می‌باشند. پس از حصول اطمینان، قطعات ایجاد شده توسط تکنیکهای کلاسه‌بندی مورد ارزیابی قرار می‌گیرند تا از میان آنها قطعه‌بندی بهینه استخراج گردد. در [34] با مسئله قطعه‌بندی بصورت یافتن کوتاهترین مسیر از یک گراف ایجاد شده از تصویر باینری یا سطح خاکستری سند برخورد می‌شود. در [5] از احتمالات HMM - که از کاراکترهای الگوریتم تشریح بدست می‌آیند - برای ایجاد گراف استفاده می‌شود. مسیر بهینه این گراف نتیجه قطعه‌بندی حاصله از الگوریتم تشریح و بازشناسی HMM را بهبود می‌بخشد. شکل (۲-۷-الف) فواصل قطعه‌بندی اولیه که از تخمین ماکزیموم و مینیمومهای محلی به همراه اطلاعات زاویه اریب بدست آمده‌اند را نشان می‌دهد. شکل‌های (۲-۷-ب) و (۲-۷-ج) نیز بترتیب کوتاهترین مسیر برای هر فاصله قطعه‌بندی و کاراکترهای کاندید شده مورد نظر را نمایش می‌دهند. تکنیکهای ادغام شده در مقایسه با روشهای قطعه‌بندی صریح و ضمنی نتایج بهتری را به بار می‌آورند.



شکل ۲-۷- قطعه‌بندی از طریق یافتن کوتاهترین مسیر یک گراف تشکیل شده از تصویر سطح

خاکستری (الف) فواصل قطعه‌بندی (ب) مسیرهای قطعه‌بندی (ج) قطعات

تکنیکهایی که در بالا معرفی گردیدند، از توانمندیهای محدودی در قطعه‌بندی برخوردارند و بایستی مکانیزمهای آشکارسازی و تصحیح خطا را در سیستمهایی که این تکنیکها برای آنها توسعه داده شده‌اند ادغام نمود. همانطوریکه Casey و Lecolinet [12] خاطرنشان ساخته‌اند، بکارگیری هوشمندانه اطلاعات متنی و قابلیت‌های طبقه‌بندی‌کننده‌ها بر دقت سیستم خواهد افزود. «عزمی» در [۸۲] بررسی جامعی بر روی روشهای معرفی شده برای قطعه‌بندی کلمات فارسی انجام نموده است.

۲-۳- بازنمایی^۱ (استخراج ویژگیها)^۲

بازنمایی تصاویر مهمترین نقش را در سیستمهای بازشناسی حروف ایفا می‌کند. در ساده‌ترین حالت، تصاویر سطح خاکستری یا باینری به یک بازشناسی‌کننده داده می‌شوند. لیکن در بیشتر سیستمهای بازشناسی بمنظور اجتناب از پیچیدگی اضافی و افزایش دقت الگوریتمها، یک روش بازنمایی فشرده‌تر و با قدرت تمایز بیشتر مورد نیاز می‌باشد. برای این منظور مجموعه‌ای از ویژگیها برای هر کلاس استخراج می‌شوند که به تشخیص آن از سایر کلاسها کمک می‌نماید در عین حالی که نسبت به تغییرات مشخصات اعضای یک کلاس مستقل باقی می‌ماند.

ویژگیهای انتخاب شده بایستی نسبت به اعوجاجها و تغییرات احتمالی که کاراکترها در یک کاربرد خاص پیدا می‌کنند، تغییرناپذیر باشند. همچنین پدیده‌ای تحت عنوان «نفرین بعد»^۳ به ما هشدار می‌دهد که با یک مجموعه آموزشی محدود چنانچه بخواهیم از یک طبقه‌بندی‌کننده آماری استفاده کنیم، تعداد ویژگیها بطور منطقی بایستی کوچک باشد. برطبق یک قانون تجربی، تعداد الگوهای آموزشی هر کلاس بایستی ۵ تا ۱۰ برابر ابعاد بردار ویژگی انتخابی باشد. در عمل مقتضیات روشهای بازنمایی، انتخاب بهترین روش برای یک کاربرد خاص را با مشکل مواجه می‌کند. همچنین این مسئله را نیز باید مد نظر قرار داد که آیا حروف یا کلماتی که می‌بایست تشخیص داده شوند جهت و اندازه مشخصی دارند یا خیر، دستنویس یا چاپی هستند، و یا ممکن است تا چه حد بوسیله

^۱ Representation

^۲ Feature extraction

^۳ Curse of dimensionality

نویز مغشوش شده باشند. از طرف دیگر در مورد حروفی که به چندین شکل نوشته می‌شوند (مانند 'a' و 'a') ممکن است لازم باشد که بیش از یک کلاس الگو به یک کاراکتر خاص تعلق یابد [56].

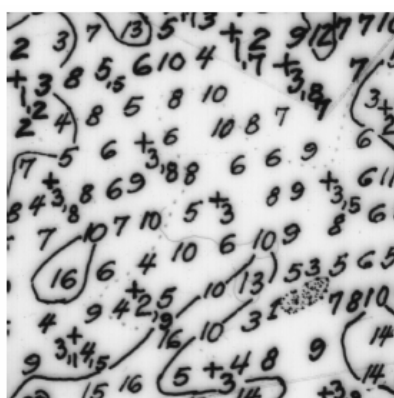
همانطور که عنوان شد، بازنمایی یک مرحله بسیار مهم در حصول راندمان مناسب برای سیستمهای بازشناسی حروف است؛ لیکن جهت دستیابی به عملکرد بهینه لازم است که سایر مراحل نیز بهینه گردند و بایستی توجه نمود که این مراحل مستقل نمی‌باشند. استخراج یک روش خاص استخراج ویژگیها، طبیعت خروجی مرحله پیش‌پردازش را به ما دیکته می‌کند و یا حداقل ما را در انتخابمان محدود می‌سازد (جدول ۱-۲). بعضی از روشهای استخراج ویژگیها بر روی تصاویر سطح خاکستری حروف منفرد عمل می‌نمایند (شکل ۲-۸-ب)؛ درحالیکه سایر روشها بر روی سمبلهای با پیوستگی مرتبه ۴ یا ۸ که از تصویر باینری اصلی جدا گردیده‌اند (شکل ۲-۸-ج)، یا بر روی سمبلهای نازک شده (اسکلت بندی شده) (شکل ۲-۸-د) و یا بر روی کانتور سمبلها (شکل ۲-۸-ه) اعمال می‌گردند.

علاوه بر این، نوع فرمت ویژگیهای استخراج شده بایستی با نیازمندیهای طبقه‌بندی‌کننده منتخب مطابقت داشته باشد. ویژگیهای نوع گرافی یا گرامری برای طبقه‌بندی‌کننده‌های ساختاری یا نحوی مناسب می‌باشند. ویژگیهای گسسته که فرضاً تنها دو یا سه مقدار مجزا دارند، برای درختهای تصمیم‌گیری ایده‌آل می‌باشند. بردارهای ویژگی با مقادیر حقیقی، مناسب طبقه‌بندی‌کننده‌های آماری هستند. همچنین چندین طبقه‌بندی‌کننده نیز ممکن است به عنوان یک روش طبقه‌بندی چند طبقه یا به عنوان طبقه‌بندی‌کننده‌های موازی (که در آن ترکیبی از نتایج طبقه‌بندی‌کننده‌های مجزا در حصول نتیجه نهایی دخالت دارند) مورد استفاده قرار گیرند. در این حالت ممکن است ویژگیهای دارای چند فرمت مختلف از حروف ورودی استخراج گردند [56].

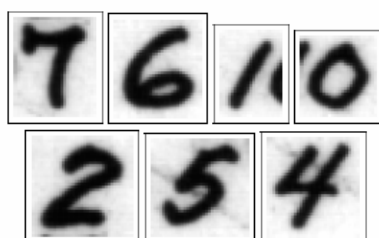
جدول ۲-۱- کاربرد برخی از روشهای استخراج ویژگیها برای شکلهای بازنمایی مختلف

(باینری، سطح خاکستری و برداری)

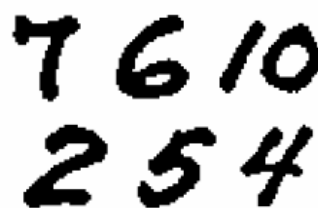
Gray scale subimage	Binary		Vector (skeleton)
	solid character	outer contour	
Template matching	Template matching		Template matching
Deformable templates			Deformable templates
Unitary Transforms	Unitary transforms		Graph description
	Projection histograms	Contour profiles	Discrete features
Zoning	Zoning	Zoning	Zoning
Geometric moments	Geometric moments	Spline curve	
Zernike moments	Zernike moments	Fourier descriptors	Fourier descriptors



(الف)



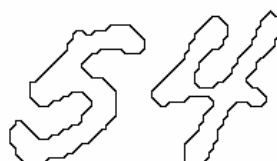
(ب)



(ج)



(د)

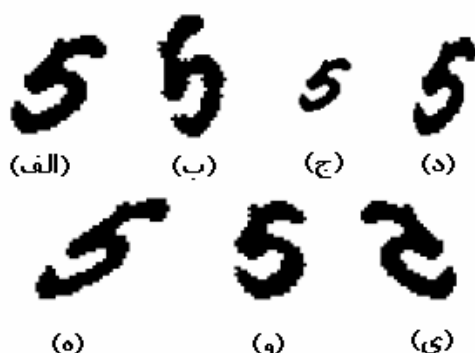


(ه)

شکل ۲-۸-الف) تصویر سطح خاکستری بخشی از یک نقشه هیدروگرافی. (ب) تصاویر سطح خاکستری حروف قطعه‌بندی شده که از بخشی از شکل "الف" استخراج گردیده‌اند. توجه شود که برای برخی ارقام، قسمتهایی از سایر اشیاء تصویری نیز در تصویر آنها موجود می‌باشد. (ج) تصویر باینری شده شکل "ب". (ج) تصویر نازک شده شکل "ج". (د) کانتورهای دو تا از ارقام شکل "ج"

- ویژگیهای پایا^۱

برای اینکه بتوانیم فرمهای تغییر یافته یک کاراکتر را نیز شناسایی کنیم، نیازمند ویژگیهایی از کاراکتر هستیم که نسبت به تبدیلات مشخصی که آن کاراکتر را دستخوش تغییر نموده‌اند، پایدار باشند. ویژگیهای پایا ویژگیهایی هستند که برای نمونه‌های تغییر یافته یک کاراکتر تقریباً مقادیر یکسانی دارند. این تغییرات شامل انتقال، تغییر اندازه، چرخش، کشیدگی، وارونگی و غیره می‌باشند (شکل ۲-۹). البته همیه تغییراتی که یک کاراکتر ممکن است با آنها مواجه گردد را نمی‌توان توسط ویژگیهای پایا مدل نمود.



شکل ۲-۹- کاراکتر '5' و نمونه‌های تغییر یافته آن. الف) تصویر اصلی ب) چرخش یافته ج) تغییر مقیاس یافته د) فشرده شده ه) اریب شده و) از حالت اریب خارج شده ی) تصویر آینه‌ای

تغییرناپذیری نسبت به اندازه و جابجایی کاراکترها به آسانی قابل دستیابی است. قطعه‌بندی به کاراکترهای مجزا، به خودی خود می‌تواند تخمینی از اندازه و محل آنها را فراهم نماید. البته ممکن است روش استخراج ویژگی ذاتاً خود نسبت به انتقال و جابجایی مستقل باشد.

تغییرناپذیری نسبت به چرخش زمانی مهم است که کاراکترهایی که می‌بایست تشخیص داده شوند بتوانند از هر راستایی برخوردار باشند. لیکن چنانچه انتظار می‌رود که همه کاراکترها جهت مشخصی داشته باشند، می‌بایست از ویژگیهای حساس به چرخش استفاده نمود تا بتوان بین کاراکترهایی نظیر "6" و "9"، "n" و "u" تفاوت قائل شد. راه حل دیگر برای این موضوع، بکارگیری ویژگیهای غیرحساس به چرخش به همراه زاویه چرخش آشکارسازی شده است. در این حالت

¹ Invariants

چنانچه زاویه چرخش محدود باشد (فرضاً بین -45° و $+45^\circ$)، تشخیص کاراکترهایی که چرخش-یافته 180° درجه‌ای یکدیگر هستند، امکان‌پذیر خواهد بود. همین موضوع نیز باید در مورد ویژگیهای مقاوم در برابر تغییر اندازه مورد توجه قرار گیرد. مثلاً اگر بخواهیم علاوه بر کاراکترها علائم نقطه-گذاری را نیز تشخیص دهیم، بایستی قادر باشیم '،' و 'o' و 'O'، و یا '،' و '9' را از هم تمیز دهیم [56].

ویژگیهای مقاوم در برابر اریب‌شدگی، در متون دستنوس که شیب کاراکترها بسته به رسم-الخط اشخاص متفاوت است و همچنین در متون چاپی که در آنها بعضی کاراکترها به فرم ایتالیک نوشته می‌شوند یا برخی فونتها شیب بیشتری نسبت به بقیه دارند، حائز اهمیت می‌باشند. عدم حساسیت نسبت به وارونگی آینه‌ای تصویر کاراکترها، در بازشناسی حروف مطلوب نمی‌باشد؛ چرا که امکان دارد تصویر آینه‌ای یک کاراکتر، سمبل غیرمجاز یا کاراکتر دیگری را نتیجه دهد.

در مورد ویژگیهایی که مستقیماً از تصاویر سطح خاکستری استخراج می‌شوند، علاوه بر موارد فوق‌الذکر، تغییرناپذیری نسبت به کنتراست بین کاراکتر و پس‌زمینه و همچنین نسبت به متوسط سطح خاکستری کاراکترها مورد نظر می‌باشد. تغییرناپذیری نسبت به متوسط سطح خاکستری بسادگی از طریق اضافه‌نمودن مقدار اختلاف بین متوسط سطح خاکستری مطلوب و متوسط سطح خاکستری واقعی به هر پیکسل قابل دستیابی است.

در حالتیکه نتوان ویژگیهای تغییرناپذیری را یافت، راه حل دیگر نرمالیزه کردن تصاویر ورودی به نحوی است که اندازه، راستا، کنتراست (پادنمایی)^۱ و سایر پارامترهای آنها یکسان گردد. لیکن این کار به تولید خطاهای گسسته‌سازی^۲ جدید منجر خواهد شد [56].

- قابلیت بازسازی مجدد^۳

در مورد برخی از روشهای استخراج ویژگیها این قابلیت وجود دارد که می‌توان تصویر کاراکتر را از ویژگیهای استخراج شده مجدداً بازسازی کرد. این خاصیت اطمینان می‌دهد که اطلاعات

^۱ Contrast

^۲ Discretization

^۳ Reconstructability

کامل درباره شکل کاراکتر، در ویژگیهای استخراج شده از آن نهفته است. گرچه برای بعضی از روشها، بازسازی کامل به تعداد زیادی از ویژگیها نیاز دارد، اما تقریبهای قابل قبولی از شکل کاراکتر اصلی را می‌توان با بکارگیری تعداد کوچکی از ویژگیهای دارای بالاترین محتوای اطلاعاتی بدست آورد. انتظار می‌رود که این ویژگیها از بالاترین قدرت تمایز بین کاراکترها برخوردار باشند.

با بازسازی مجدد تصاویر کاراکترها از ویژگیهای استخراج شده می‌توان بصورت بصری امتحان نمود که آیا تعداد کافی از ویژگیها برای نمایش اطلاعات ساختار اصلی کاراکترها استفاده گردیده است یا خیر. همچنین می‌توان از طریق بازسازی مجدد، بصورت غیر رسمی صحت عملیات پیاده‌سازی را امتحان نمود. مراجع خوب در مورد روشهای استخراج ویژگیها برای سیستمهای بازشناسی حروف لاتین، مقالات [56] و [6] و برای سیستمهای بازشناسی فارسی، پایان‌نامه [۸۲] می‌باشند.

در ادامه این بخش روشهای مختلف بازنمایی تصاویر اسناد را که به سه گروه اصلی طبقه‌بندی شده‌اند را بررسی می‌نماییم.

۲-۳-۱- تبدیل سراسری^۱ (بسط سری^۲)

یک سیگنال پیوسته معمولاً حاوی اطلاعات بیشتری از میزان مورد نیاز برای بازنمایی جهت مقاصد کلاسه‌بندی است. این موضوع می‌تواند برای تقریبهای گسسته سیگنالهای پیوسته نیز صادق باشد. یک روش برای بازنمایی یک سیگنال عبارت است از یک بسط سری بصورت ترکیب خطی از توابع ساده‌تری (دارای پیچیدگی کمتر) که به نحو مناسبی تعریف شده‌اند. در واقع، ضرایب ترکیب خطی، یک روش کدکردن فشرده‌تر سیگنال می‌باشند که به آنها تبدیل یا بسط سری سیگنال اطلاق می‌شود. در زیر، معمولترین تبدیلهای و روشهای بسط سری در حوزه سیستمهای بازشناسی حروف بررسی می‌گردند.

^۱ Global Transformation

^۲ Series Expansion

الف - تبدیلات فوریه^۱ (توصیف کننده های فوریه^۲)

تبدیل فوریه گسسته (DFT)^۳ از اهمیت ویژه ای در حوزه پردازش تصویر و سیگنال برخوردار است و در آن معمولاً از طیف دامنه بردار ویژگی در فضای اقلیدسی n بعدی به عنوان ویژگی استفاده می شود. این تبدیل از توابع پایه نمایی مختلط بهره می گیرد و چون از خاصیت فشرده کردن انرژی بالایی برخوردار است، راندمان کدینگ مطلوبی را ارائه می دهد. DFT چندین ویژگی دارد که در امر شاخص گذاری^۴ یا تطبیق الگوها مفید واقع می گردد: نخست آنکه دامنه ضرایب DFT نسبت به انتقال حساس نیست^۵ و در نتیجه با احتساب طیف دامنه به تنهایی و صرف نظر از اطلاعات فاز می توان حساسیت نسبت به تغییر مکان کاراکترها در تصویر را از بین برد. در ثانی، همبستگی حوزه زمانی را می توان بطور مؤثر با استفاده از ضرایب DFT محاسبه نمود. تبدیلات فوریه به راههای زیادی در بازشناسی حروف مورد استفاده قرار گرفته اند.

روشهای مختلف کد کردن منحنی پیرامونی در موارد زیادی برای بازشناسی حروف استفاده شده اند. در این میان توصیف کننده های فوریه نرمالیزه که نسبت به اندازه، انتقال و چرخش ثابت هستند، کاربرد بیشتری دارند. از این ویژگیها در بازشناسی حروف لاتین و کاربردهای پردازش تصویر استفاده زیادی شده است. از جمله کارهایی که با استفاده از این روش بر روی متون فارسی و عربی صورت گرفته است، عبارتند از [۸۶]:

الف - استخراج مستقیم توصیف کننده های فوریه از منحنی پیرامونی حروف.

ب - استفاده از کدهای فریمن در کد کردن منحنی پیرامونی و استخراج ضرایب فوریه از آن.

چنانچه منحنی پیرامونی حرف را به صورت یک منحنی دو بعدی با مختصات $x(n)$ و $y(n)$

در نظر بگیریم، توصیف کننده های فوریه با استفاده از تبدیل فوریه گسسته بصورت زیر تعریف می شوند:

¹ Fourier Transforms

² Fourier Descriptors

³ Discrete Fourier Transform

⁴ Indexing

⁵ Translation invariant

$$u(n) = x(n) + jy(n)$$

$$u(n) = \frac{1}{N} \sum_{k=0}^{N-1} a(k) \exp\left(\frac{j2\pi kn}{N}\right), \quad 0 \leq n \leq N-1 \quad (1-2)$$

$$a(n) = \frac{1}{N} \sum_{k=0}^{N-1} u(k) \exp\left(\frac{-j2\pi kn}{N}\right), \quad 0 \leq n \leq N-1$$

در معادله فوق $a(n)$ ضرایب فوریه را نشان می‌دهد. همانطور که گفته شد، معمولاً در کاربردهای بازشناسی تنها از اندازه این ضرایب استفاده می‌شود.

ب- موجکها^۱

اخيراً تبدیلات زیرباند و موجک گسسته (DWT)^۲ در عرصه کاربردهای کدینگ، شاخص‌گذاری تصاویر^۳ و بازشناسی حروف با مقبولیت فراوانی مواجه شده‌اند. در این تبدیلهای، یک تصویر بطور پیاپی از مجموعه‌ای از فیلترهای پایین‌گذر و بالا‌گذر عبور نموده و خروجی فیلترها یک‌چندی^۴ می‌شود تا نرخ داده‌ها ثابت باقی بماند. در DWT، تنها خروجی پایین‌گذر به صورت پیاپی فیلتر می‌گردد.

در تبدیل فوریه که نوعی از تبدیل با توابع پایه ارتوگونال (متعامد)^۵ می‌باشد، از «موج»^۶های سینوسی استفاده می‌شود که در هر دو جهت دارای دوره نامحدود هستند. بطور مشابه در حالت گسسته، بردارهای پایه تبدیل DFT نیز دارای دوره نامحدود می‌باشند و بنابراین این تبدیل دارای ویژگی «دوره-محدود»^۷ نمی‌باشد. اما در عمل به سیگنالهای گذرای برمی‌خوریم که تنها در یک دوره کوتاه زمانی غیر صفر هستند. مثلاً بسیاری از خصوصیات مهم تصاویر مانند لبه، در یک مکان مشخص قرار دارند. این ترکیبات به هیچیک از توابع پایه تبدیل فوریه شبیه نیستند و در نتیجه توابع مذکور ابزار بهینه‌ای برای آنالیز سیگنالها و تصاویری که حاوی مؤلفه‌های گذرا یا متمرکز در یک زمان یا مکان هستند، نمی‌باشند. البته می‌دانیم که تبدیل فوریه می‌تواند هر تابع تحلیلی (حتی یک

¹ Wavelets

² Discrete Wavelet Transform

³ Image Indexing

⁴ Decimate

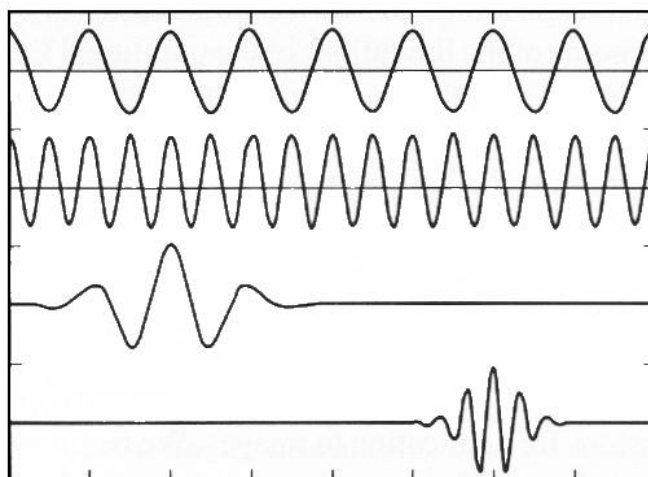
⁵ Orthogonal

⁶ wave

⁷ Compactly supported

سیگنال گذرای باریک) را بصورت مجموعی از سینوسها نمایش دهد. اما ضرایب بدست آمده از تبدیل، معیار فشرده‌ای از اطلاعات تصویر نخواهند بود.

برای رفع این نقیصه ریاضیدانان و مهندسين به کوششهایی دست زده‌اند که در آنها توابع پایه با دوره محدود بکار برده می‌شوند. به این توابع پایه «موجک» و به تبدیلاتی که بر اساس آنها هستند، «تبدیلات موجک» می‌گویند. شکل (۲-۱۰) تفاوت بین موج و موجک را نشان می‌دهد. دو منحنی بالایی، امواج سینوسی با فرکانسهای مختلف و دو منحنی پایینی، دو موجک با فرکانس و موقعیت متفاوت می‌باشند.

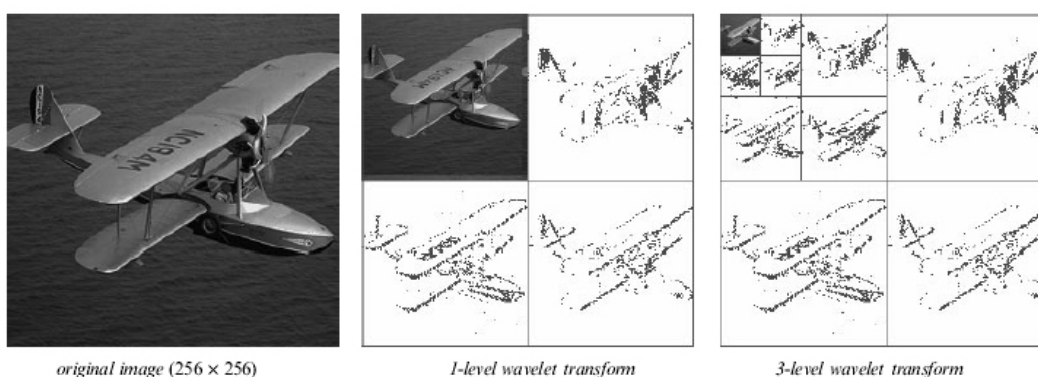


شکل ۲-۱۰- مقایسه موج و موجک

برای بسیاری از سیگنالها، تحلیل فوریه ابزاری بینهایت مفید محسوب می‌شود؛ زیرا پی بردن به محتوای فرکانسی سیگنالها از اهمیت ویژه‌ای برخوردار است. پس به چه دلیل ما نیازمند تکنیکهای دیگری همچون تحلیل موجک می‌باشیم؟ تحلیل فوریه یک نقطه‌ضعف اساسی دارد و آن اینکه در فرایند تبدیل یک سیگنال به حوزه فرکانس، اطلاعات زمانی یا مکانی آن از بین می‌رود. به عبارت دیگر، با نگاه کردن به تبدیل فوریه یک سیگنال نمی‌توان اظهار داشت که یک رویداد خاص، در چه برهه زمانی و یا در چه مکانی از تصویر بوقوع پیوسته است. چنانچه تغییرات یک سیگنال در زمان یا مکان قابل ملاحظه نباشد (سیگنال ایستادن)، این نقطه‌ضعف اهمیت چندانی ندارد. لیکن بیشتر سیگنالهای مورد توجه، در بسیاری نقاط (مانند نقاط ابتدایی و انتهایی) از ماهیت غیرایستادن یا گذرا

برخوردارند. این نقاط اغلب مهمترین قسمت سیگنال می باشند که تحلیل فوریه ابزار مناسبی برای آشکارسازی آنها محسوب نمی شود.

تبدیل موجک یک تکنیک بسط سری است که این امکان را بوجود می آورد که بتوان سیگنال را در درجات مختلف تفکیک پذیری نمایش داد. قطعات مختلف تصویر سند که ممکن است مربوط به حروف یا کلمات باشند، متناظر با سطوح مختلف تفکیک پذیری بوسیله ضرایب موجک نشان داده می شوند. بازنمایی سیگنال در فرم آنالیز چند تفکیکه (MRA)^۱ با درجه تفکیک پذیری پایین می تواند در مقایسه با MRA با درجه تفکیک پذیری بالا، تغییرات محلی حادث شده در دست نوشته ها را بنحو مؤثری به خود جذب نماید. لیکن بازشناسی در تفکیک پذیری پایین ممکن است باعث شود که جزئیات بااهمیت برای مرحله بازشناسی از دست برود. شکل (۲-۱۱) تبدیل موجک یک تصویر نمونه را نشان می دهد که در آن، ضرایب زیر یک سطح آستانه حذف گردیده اند.



شکل ۲-۱۱- تبدیل موجک یک تصویر نمونه پس از اعمال یک سطح آستانه

تبدیل موجک انواع مختلفی دارد که از توابع پایه متفاوتی برخوردارند؛ نظیر: Haar, Mexican hat, Meyer, Morlet, Symlets, Coiflets, Biorthogonal, Daubechies. آنجائیکه توابع پایه موجک Haar شکل گسسته دارند، تصور می شود که این موجک برای کار با تصاویر باینری کاراکترها، بر سایر موجکها ارجحیت داشته باشد. در اینجا از ذکر ریاضیات تبدیلات موجک بدلیل حجم نسبتاً زیاد آن صرف نظر می شود. اخیراً Mowlaei [44] و همکارانش روشی بمنظور بازشناسی کدهای پستی و اسامی شهرهای ایران که بر روی بسته های پستی بصورت دست-

^۱ Multi-Resolution Analysis

نویس نوشته می‌شوند، پیشنهاد نموده‌اند. بردارهای ویژگی این روش، از طریق بکارگیری تبدیل موجک گسسته با موجک اساسی Haar محاسبه می‌گردند. در مرحله پیش‌پردازش این الگوریتم، ابتدا محدوده هر حرف مشخص گردیده، تصویر حرف به ابعاد 64×64 پیکسل تغییر اندازه می‌یابد و در نهایت عمل نازک‌سازی انجام می‌گیرد. در مرحله استخراج ویژگیها، تبدیل موجک سه سطحی بر روی تصویر نازک شده اعمال می‌شود و ضرایب باند LL از سطح سوم استخراج می‌گردند. در انتها ماتریس ضرایب باند LL بصورت برداری در آمده، یک فیلتر پایین‌گذر بمنظور هموارسازی بیشتر ضرایب، به آن اعمال می‌شود. برای آموزش سیستم، یک شبکه عصبی MLP با قانون پس‌انتشار خطا مورد استفاده قرار می‌گیرد. راندمان این روش برای حروف و ارقام دست‌نویس فارسی بترتیب ۹۲/۳۳ و ۹۱/۸۱ درصد گزارش شده است.

ج- تبدیل گابور^۱

این تبدیل تغییر یافته تبدیل فوریه پنجره شده است که در آن پنجره مورد استفاده بجای آنکه یک اندازه گسسته داشته باشد، بوسیله یک تابع گوسی تعریف می‌گردد. به بیان دیگر، تبدیل گابور مشابه تبدیل موجک است که در آن توابع پایه طبیعت گوسی دارند و در نتیجه این تبدیل در جانمایی زمانی - فرکانسی^۲ بهینه است. موجک گابور از جهت حداقل‌سازی عدم قطعیت^۳ دو بعدی توأم بین حوزه‌های مکان و فرکانس، یک تبدیل بهینه می‌باشد. این موجک را می‌توان به عنوان آشکارسازهای جهت‌دار و تطبیق‌پذیر با مقیاس، بمنظور آشکارسازی خطوط و لبه‌ها در تصاویر مورد استفاده قرار داد. همچنین از خواص آماری این تبدیل می‌توان جهت تعیین ساختار و محتوای بصری تصاویر بهره گرفت. ویژگیهای تبدیل گابور در چندین کاربرد آنالیز تصاویر شامل کلاسه‌بندی و قطعه‌بندی بافتها، بازشناسی تصاویر، بازشناسی حروف، ثبت تصاویر و مسیریابی حرکت بکار رفته‌اند. برغم غیر متعامد بودن این خانواده موجکها، تابع گابور تنها تابعی است که می‌تواند حد تئوری دقت تفکیک توأم اطلاعات در هر دو حوزه مکان و فرکانس را حاصل نماید.

^۱ Gabor Transform

^۲ Time-frequency localization

^۳ uncertainty

یک موجک گابور دو بعدی عبارت است از یک تابع گوسی مدوله شده با یک تابع سینوسی مختلط. این تابع را می‌توان بوسیله فرکانس تابع سینوسی (W) و انحراف معیارهای σ_x و σ_y پوش تابع گوسی بصورت زیر بیان نمود:

$$\psi(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi j W x\right] \quad (2-2)$$

علیرغم این شکل کلی، هیچ تعریف استاندارد و دقیقی از تابع گابور دو بعدی وجود ندارد و اشکال مختلفی از این تابع در مقالات ظاهر شده است. بیشتر این اختلافات به اندازه‌های مختلف پوش تابع گوسی و فرکانس تابع سینوسی مربوط می‌شود. بر اساس تئوری موجک، تابع گابور که بنحو مناسبی نرمالیزه شده باشد، می‌تواند به عنوان یک موجک مادر برای تولید خانواده‌ای از موجکهای گابور غیر متعامد بکار رود.

Manjunath و Ma در [40] بمنظور شاخص‌گذاری و بازیابی تصاویر بافت‌گونه، بانکی از فیلترهای گابور نرمالیزه شده با پاسخ dc صفر را مورد استفاده قرار داده‌اند. این فیلترها بنحوی طراحی شده‌اند که نمایش آنها، درجه افزونگی پایینی را شامل شود. موجکهای گابور بکار رفته در اینجا، از طریق انبساط^۱ و چرخش تابع $\psi(x, y)$ در (۲-۲) ایجاد شده‌اند که تابع مولد آنها به فرم زیر می‌باشد:

$$\begin{aligned} \psi_{mn}(x, y) &= a^{-m} \psi(x', y'), \quad a > 1, \quad m, n = \text{Integers} \\ x' &= a^{-m} [(x - x_0) \cos(\theta) + (y - y_0) \sin(\theta)] \\ y' &= a^{-m} [-(x - x_0) \sin(\theta) + (y - y_0) \cos(\theta)] \end{aligned} \quad (3-2)$$

که در آن (x_0, y_0) مرکز فیلتر در حوزه مکان، $\theta = n\pi/K$ و S و K بترتیب تعداد کل مقیاسها^۲ و راستاهای^۳ در نظر گرفته شده می‌باشد. فاکتور a^{-m} در (۳-۲) به این دلیل ظاهر شده است که از مستقل بودن انرژی نسبت به m اطمینان حاصل شود.

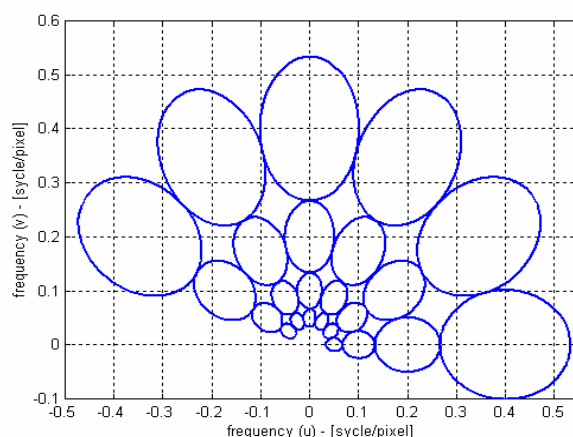
^۱ Dilation^۲ scales^۳ orientations

جهت حصول اطمینان از اینکه بانک فیلتری طراحی شده را بتوان یک خانواده مناسب از موجکهای گابور دو بعدی در نظر گرفت، فیلترهای $\psi(x,y)$ بایستی شرط پذیرش «محدودیت انرژی» را برآورده سازند که بر پایه آن، تبدیل فوریه این فیلترها باید توابع میان گذر خالص با پاسخ dc صفر باشد. این شرط با صفر نمودن پاسخ فرکانسی فیلترها در فرکانس صفر ($\Psi(0,0)=0$) و یا با اضافه نمودن یک عدد ثابت به قسمت حقیقی (زوج) فیلترها تحقق می یابد که بدین وسیله از عدم پاسخ فیلترها به نواحی با شدت ثابت اطمینان حاصل می شود. بحث فوق، روابط زیر را بمنظور محاسبه پارامترهای σ_u و σ_v فیلتر نتیجه می دهد [40]:

$$a = \left(\frac{U_h}{U_l} \right)^{\frac{1}{s-1}} \quad \text{و} \quad \sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2 \ln 2}} \quad \sigma_v = \tan\left(\frac{\pi}{2K}\right) \times \sqrt{\frac{U_h^2}{2 \ln 2} - \sigma_u^2} \quad (4-2)$$

در این مقاله طراحی جهتها و مقیاسها مبتنی بر سه اصل زیر بوده است:

- ۱- تفکیک یکنواخت جهتها: کلیه فیلترهای متعلق به یک مقیاس خاص می بایست انحراف استاندارد زاویه ای (σ_v) یکسانی داشته و اختلاف در جهت گیری آنها نسبت به یکدیگر یکسان باشد.
 - ۲- تفکیک نمایی مقیاسها: پهنای فیلترها باید با افزایش فاصله از مرکز صفحه (u,v) بصورت نمایی افزایش یابد.
 - ۳- پوشش پیوسته فضای فرکانسی: طراحی فیلترها می بایست بنحوی باشد که کانتورهای نصف توان مجاور در هر دو سمت محورهای مقیاس و جهت با یکدیگر تماس داشته باشند.
- شکل (۲-۱۲) بانک فیلترهای موجک گابور را مطابق طراحی فوق نمایش می دهد.



شکل ۲-۱۲- دامنه نصف توان ضرایب بانک فیلتری گابور

از تبدیل گابور در چند مقاله بمنظور بازشناسی حروف چینی و لاتین استفاده گردیده است. Hue و همکارانش [24] الگوریتمی بمنظور قرائت ارقام پلاک خودروها با استفاده از فیلترهای گابور معرفی نموده‌اند. ابتدا مجموعه‌ای از فیلترهای گابور با ۱۸ راستای مختلف به تصویر حروف اعمال می‌شود. سپس با استفاده از ماتریس جهت غالب گابور^۱، بردار ویژگی تصویر محاسبه می‌گردد. آزمایش نشان می‌دهد که این الگوریتم قادر به تشخیص مناسب حروف حتی در زمینه‌های بسیار نویزی است. راندمان این روش ۹۶/۶۶ درصد برآورد شده است.

GE و همکارانش [20] بتازگی یک موتور OCR برای شناسایی دستنوشته‌های زبان چینی ابداع نموده‌اند که از تکنیکهای مختلفی چون ویژگیهای گابور و مشتقات مکانی آنها به عنوان ویژگیهای خام، آنالیز مبین خطی (LDA)^۲ بمنظور استخراج ویژگیها و کاهش ابعاد فضای ویژگی، CDHMMها برای مدل کردن کاراکترهای چینی در راستای هر دو جهت افقی و عمودی، و معیار حداقل خطای کلاسه‌بندی جهت آموزش مدل بهره گرفته‌اند. در زمان تدوین این مقاله، موتور OCR آنها از یک فرهنگ لغت با ۴۶۱۶ کاراکتر (شامل ۴۵۱۶ کاراکتر چینی ساده، ۶۲ کاراکتر الفبایی - عددی^۳ و ۳۸ علامت و نماد نقطه‌گذاری) پشتیبانی می‌کند. با بکارگیری ۱,۳۸۴,۸۰۰ نمونه کاراکتر جهت آموزش سیستم بازشناسی کننده، دقت بازشناسی ۹۶/۳۴ درصد حاصل گردیده است و این در حالی است که بهترین سیستم OCR چینی موجود در بازار، روی این مجموعه از راندمان ۹۴/۰۷ درصد برخوردار بوده است.

د- ممانها (گشتاورها)^۴

ممانها بطور وسیعی در کاربردهای پردازش تصویر بمنظور توصیف شکل هندسی اشیاء تصویری گوناگون مورد استفاده قرار گرفته‌اند. این کمیتها خواص هندسی بنیادی (مانند سطح، مرکز جرم، ممان انرسی، میزان اریب بودن و درجه تیزی نقطه اوج منحنی^۵) یک توزیع را نشان می‌دهند.

^۱ Gabor Dominant Orientation Matrix

^۲ Linear Discriminant Analysis

^۳ Alphanumeric

^۴ Moments

^۵ kurtosis

همچنین ممانها را می توان بمنظور تقریب زدن تابع چگالی احتمال^۱ (هیستوگرام) تصاویر بکار برد. در اینجا خاطر نشان می شود که PDF سطوح خاکستری (شدت روشنایی) یک تصویر، همان هیستوگرام آن می باشد؛ جز اینکه با یک تغییر مقیاس، سطح کل زیر منحنی آن برابر واحد است.

ممانهای نرمالیزه ویژگیهای مناسبی برای بازشناسی حروف هستند و در موارد زیادی بکار گرفته شده اند. مزیت این ویژگیها غیر حساس بودن آنها نسبت به چرخش، تغییر مکان و اندازه حروف است. ممانهای مختلف نظیر ممانهای مرکزی^۲، ممانهای منظم^۳، ممانهای لژاندر^۴ و ممانهای شبه زرنیکه^۵ نمایش فشرده ای از تصویر سند را ارائه می دهند. ممانها بصورت بسط سریها در نظر گرفته می شوند؛ زیرا تصویر اولیه را می توان از ضرایب ممانها بطور کامل بازسازی نمود.

گشتاورهای هندسی و نرمالیزه بصورت زیر تعریف می شوند:

$$\mu_{pq} = \sum_x (x - \bar{x})^p (y - \bar{y})^q f(x, y)$$

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\frac{p+q+2}{2}}}$$

در روابط فوق، μ_{pq} گشتاور مرکزی و η_{pq} گشتاور مرکزی نرمالیزه از مرتبه p و q است.

در [7] الگوریتمی برای بازشناسی حروف دستنویس فارسی با استفاده از این ویژگیها ارائه شده است. در این الگوریتم، حروف ورودی با استفاده از خواصی مانند تعداد قطعات و وجود یا عدم وجود تک نقطه به پنج گروه طبقه بندی می شوند. در هر گروه، کلاس حرف ورودی با استفاده از طبقه بندی بیز و ویژگیهای گشتاوری تعیین می شود. در نهایت در یک مرحله پس پردازش، با بکار بردن برخی ویژگیهای ساختاری، میزان خطا در کلاسه های مشابه کاهش داده می شود. میزان بازشناسی درست برای ۳۳۰۰ نمونه از حروف مجزای دستنویس که توسط ۱۰۰ نویسنده مختلف نوشته شده حدود ۹۱ درصد گزارش شده است.

¹ Probability Density Function (PDF)

² Central moments

³ Regular moments

⁴ Legendre moments

⁵ Pseudo-Zernike

در [3] برای شناسایی حروف عربی، ابتدا بدنه آنها جدا شده و سپس هیستوگرامهای عمودی و افقی مربوط به تصویر بدنه محاسبه شده است [۸۲]. با محاسبه گشتاورهای هندسی یک بعدی برای هیستوگرامهای عمودی و افقی، مجموعه‌ای از ۹ ویژگی نرمالیزه در بازشناسی بکار گرفته شده‌اند. برای طبقه‌بندی، از یک تابع ممیز مرتبه ۲ (نوع خاصی از طبقه‌بندی‌کننده بیز) استفاده شده است. میزان بازشناسی برای سه قلم مختلف حدود ۹۸٪ گزارش شده است.

در [2] بازشناسی حروف عربی با استفاده از ویژگیهای گشتاوری و شبکه عصبی مطرح است [۸۲]. سیستم بازشناسی دارای سه مرحله مختلف است. ابتدا کلمات با استفاده از هیستوگرام عمودی به حروف شکسته می‌شوند. در مرحله بعد شش ویژگی گشتاوری مختلف برای بازشناسی حروف استخراج می‌شود. در نهایت یک شبکه عصبی پرسپترون با الگوریتم آموزشی پس انتشار خطا برای بازشناسی بکار برده شده است.

از گشتاورهای هندسی مرکزی نرمالیزه و ثابت‌های گشتاوری در [۸۰] برای بازشناسی حروف تایپ شده با یک ماشین تحریر استفاده شده است. در سیستم طراحی شده، بعد از استخراج ویژگیهای نرمالیزه از طبقه‌بندی‌کننده بیز برای تشخیص حروف و علائم استفاده شده است. تحقیق انجام شده نشان داده است که میزان بازشناسی برای یک نوع قلم در حدود ۱۰۰ درصد است و ویژگیهای استخراج شده به نوع قلم حساس هستند.

در [۷۶] برای بازشناسی حروف دستنویس (نسخ)، بعد از مرحله جداسازی ابتدا نقاط و علائم از بدنه حروف جدا شده و با استفاده از گشتاورهای لژاندار از مرتبه ۴ بازشناسی شده‌اند. برای بازشناسی بدنه حروف از گشتاورهای شبکه زرنیکه مرتبه ۴ بعنوان ویژگی و از یک شبکه عصبی پرسپترون برای طبقه‌بندی استفاده شده است. میزان بازشناسی برای ۵۰۰ کلمه دستنویس که توسط ۵ خطاط معروف ایرانی نوشته شده است ۹۵٪ بوده است. دلیل بیشتر خطاها عدم امکان جداسازی حروف در بعضی از کلمات مانند «محمد» است.

فائز و میرزاخانی در [۸۳] از گشتاورهای شبه زرنیکه و شبکه عصبی پرسپترون برای بازشناسی حروف و ارقام مجزای دستنویس نیز استفاده شده است که دقت بدست آمده برای اعداد ۹۸٪ و برای حروف ۹۹٪ گزارش شده است.

ه- بسط کارهونن لُوئو (K-L)^۱

تبدیل K-L بر اساس خواص آماری یک تصویر عمل می‌کند و در واقع یک آنالیز بردار ویژه است. توابع پایه این تبدیل، مقادیر ویژه ماتریس همبستگی تصویر می‌باشند. این تبدیل سعی در کاهش ابعاد مجموعه ویژگیها بوسیله ایجاد ویژگیهای جدیدی که ترکیبات خطی از ویژگیهای اولیه می‌باشند، دارد. تبدیل K-L حداکثر فشردگی انرژی را انجام می‌دهد و از لحاظ آماری تنها تبدیل بهینه محسوب می‌شود. از آنجائیکه توابع پایه KLT با هر دسته از تصاویر قابل تطبیق می‌باشند، با تصویرکردن^۲ تصاویر در فضای K-L و محاسبه ضرایب KLT، راندمان بالایی حاصل خواهد شد. بسط K-L در چندین زمینه شناسایی الگو از جمله بازشناسی چهره مورد استفاده قرار می‌گیرد. همچنین از این تبدیل در سیستم OCR مؤسسه ملی استانداردها و فناوری (NIST) بمنظور بازشناسی دستنویسهای فرم‌گونه بکار گرفته می‌شود. نقطه ضعف این تبدیل، حجم محاسباتی بالای آن است و بدلیل برخورداری از الگوریتمهای محاسباتی بسیار سنگین، استفاده از ویژگیهای K-L در سیستمهای OCR رایج نمی‌باشد. لیکن با افزایش قدرت محاسباتی ماشینهای محاسب، این تبدیل طی چند سال آتی به یک ویژگی محقق‌پذیر مبدل خواهد گردید. به عنوان نمونه، این تبدیل در [52] بمنظور بازشناسی حروف عربی و در [27] جهت شناسایی حروف چینی مورد استفاده واقع گردیده است.

۲-۳-۲- ویژگیهای آماری

بازنمایی یک تصویر سند بوسیله توزیع آماری آن، تا اندازه‌ای تغییرات نگارشی را مورد توجه قرار می‌دهد. گرچه این روش بازنمایی، بازسازی تصویر اولیه را امکان‌پذیر نمی‌سازد، اما از آن در جهت کاهش ابعاد مجموعه ویژگیها استفاده می‌شود که بدین وسیله سرعت بالاتر و پیچیدگی

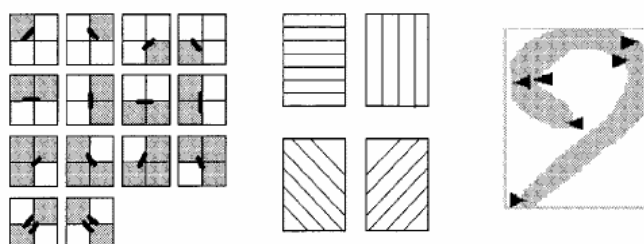
^۱ Karhunen-Loeve

^۲ projection

کمتری حاصل می‌گردد. عمده ویژگیهای آماری مورد استفاده در بازشناسی حروف بشرح زیر می‌باشند:

الف - ناحیه‌بندی^۱

در این روش، قاب در برگرفته کاراکتر به چندین ناحیه همپوشان یا غیر همپوشان تقسیم می‌شود. چگالی نقاط یا ویژگیهایی از نواحی مختلف پس از آنالیز بمنظور بازنمایی آن کاراکتر بکار می‌روند. به عنوان مثال، ویژگیهای جهتی کانتور کاراکتر، جهت کانتور کاراکتر را اندازه‌گیری می‌کند. این ویژگیها با تقسیم آرایه تصویر به نواحی مستطیلی و قطری، و محاسبه هیستوگرامهای کدهای زنجیره‌ای^۲ در این نواحی تولید می‌گردند. به عنوان مثال دیگر می‌توان به ویژگیهای نقاط خمش^۳ اشاره نمود که نشان دهنده نقاط دارای خمیدگی زیاد^۴، نقاط انتهایی^۵ و نقاط انشعاب^۶ می‌باشند. شکل (۲-۱۳) ویژگیهای جهتی کانتور و ویژگیهای نقاط خمش را نشان می‌دهد.



شکل ۲-۱۳ - ویژگیهای جهتی کانتور و نقاط خمش همراه با ناحیه‌بندی

ب - تلاقیها و فواصل^۷

یک ویژگی معمول، تعداد دفعات تلاقی یک کانتور با یک قطعه خطی در یک جهت تعیین شده است. در [4] قاب دربرگیرنده کاراکتر به مجموعه‌ای از نواحی در جهات مختلف تقسیم‌بندی می‌شود و سپس تعداد گذرهای سیاه در هر ناحیه، با توان دو کد می‌گردد. در تحقیق دیگری ([42]) مکان و تعداد دفعات گذر از پیکسلهای پس‌زمینه به پیش‌زمینه در امتداد خطوط عودی از تصویر

¹ Zoning

² Chain codes

³ Bending points

⁴ High curvature

⁵ End points

⁶ Fork points

⁷ Crossings and Distances

کلمه کد می‌شود. همچنین فاصله قطعه خطوط از یک مرز مشخص مانند قسمت بالایی و پایینی قاب کلمه را می‌توان به عنوان ویژگیهای آماری بکار برد. این ویژگیها بطور ضمنی تصریح می‌نمایند که یک سطح آستانه افقی در بالا، پایین و مرکز دست‌نوشته نرمالیزه شده وجود دارد و تعداد دفعاتی که آن کاراکتر از آن سطح آستانه عبور می‌کند، مقدار ویژگی آن کاراکتر محسوب می‌شود. هدف اصلی در اینجا تشخیص قسمت‌های بالارونده و پایین‌رونده در تصویر کاراکتر است.

ج - هیستوگرامها (تصویرنماییها)^۱

کاراکترها را می‌توان بوسیله هیستوگرام مقدار سطوح خاکستری پیکسل‌هایشان بر روی خطوط در جهت‌های متفاوت، بازنمایی کرد. این روش نمایان‌کن کاراکترها، یک سیگنال یک‌بعدی از تصویر دوبعدی ایجاد می‌کند که می‌تواند به عنوان ویژگی معرف تصویر کاراکتر مورد استفاده قرار گیرد.

۲-۳-۳- ویژگیهای هندسی^۲ و توپولوژیکی^۳

خصوصیات مختلف سراسری و محلی کاراکترها را می‌توان بوسیله ویژگیهای هندسی و توپولوژیکی که از تحمل‌پذیری بالایی نسبت به اعوجاجها و تغییرات سبک نگارشی برخوردارند، بازنمایی نمود. این روش بازنمایی ممکن است برخی دانشها در مورد ساختار شیء را نیز کد نماید و یا امکان دارد اطلاعاتی را فراهم نماید که بواسطه آن بتوان تشخیص داد که چه نوع اجزاء ساختار شیء را تشکیل می‌دهند. نمونه‌های متعدد از روشهای بازنمایی هندسی و توپولوژیکی را می‌توان در گروههای زیر جای داد:

الف - ویژگیهای توپولوژیکی

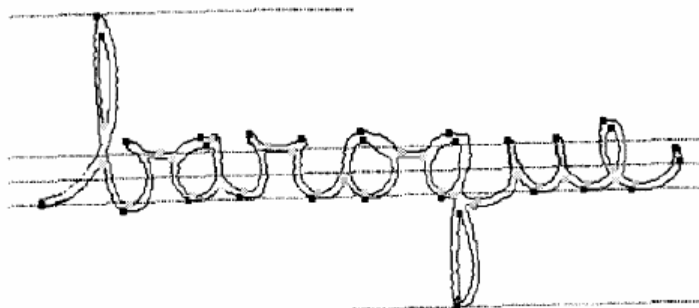
در این گروه از روشهای استخراج ویژگیها، یک ساختار از پیش تعریف شده در حرف یا کلمه مورد جستجو قرار می‌گیرد. تعداد یا موقعیت نسبی این ساختارها در کاراکتر، یک روش بازنمایی توصیفی را نتیجه می‌دهد. ساختارهای اولیه عمومی زیرحرفها می‌باشند که بدنه یک حرف

^۱ Projections

^۲ Geometrical

^۳ Topological

را تشکیل می‌دهند. این ساختارهای اولیه می‌توانند بسادگی خطوط (l) و کمانها (c) که اجزاء اصلی حروف لاتین می‌باشند و نیز به پیچیدگی منحنیها و دندانهای تشکیل دهنده حروف چینی و عربی باشند. برخی از این ویژگیها عبارتند از: نقاط حدی (کرانه)^۱، نقاط ماکزیموم و مینیموم، رئوس^۲ بالا و پایین یک حد آستانه، روزنه‌های واقع در طرف راست (c)، چپ (s)، بالا (u) و پایین (n)، نقاط تقاطع (X)، نقاط انشعاب (T)، انتهای خطی شکل (J)، حلقه‌ها (o)، جهت یک زیرحرف از یک نقطه خاص، انحنا بین دو نقطه، نقطه‌های مجزا، شکستگی بین دو نقطه، تقارن حرف، منحنیهای افقی در بالا یا پایین، پاره‌خطهای بین دو نقطه، پاره‌خطهای صعودکننده، نزول‌کننده و میانی، ارتباط بین پاره-خطوطی که یک زیرحرف را تشکیل می‌دهند و غیره. شکل (۲-۱۴) برخی از ویژگیهای توپولوژیکی را نمایش می‌دهد.



شکل ۲-۱۴- چند ویژگی توپولوژیکی: نقاط حداقل و حداکثر روی کانتورهای بیرونی و درونی، خطوط مرجع، قسمتهای فوقانی و زیرین حروف

ب- ویژگیهای هندسی

تعدادی از این ویژگیها عبارتند از: نسبت عرض و ارتفاع قاب دربردارنده کاراکتر، فاصله نسبی بین نقطه پایانی و نقطه دارای کمترین عرض، فواصل نسبی افقی و عمودی بین اولین و آخرین نقطه، فاصله بین دو نقطه، طولهای قابل تطبیق بین دو پاره‌خط، پهنای یک زیرحرف، بخشهای عمده بالا و پایین کلمات، طول کلمات و انحنا یا تغییرات انحنا.

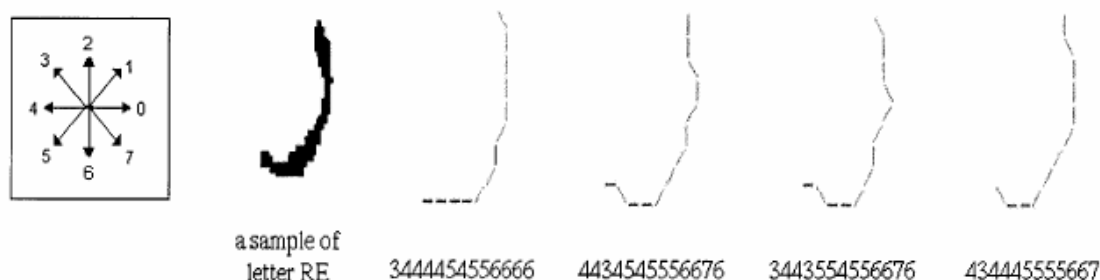
¹ Extreme points

² cusp

ویژگیهای هندسی اندازه‌گیری شده را می‌توان با یک مجموعه هندسی مناسبتر و فشرده‌تر از ویژگیها تقریب زد. از جمله این روشها، تقریب چند ضلعی کاراکترهای نازک شده است. نمونه دقیقتر روش تقریب چند ضلعی، روش بازنمایی cubic spline می‌باشد.

ج- کدگذاری

از جمله معمولترین روشهای کدگذاری، کد زنجیره‌ای «فریمن»^۱ است. این کد اساساً از طریق نگاشت زیرحرفهای یک کاراکتر به یک فضای پارامتری دوبعدی که از کدها تشکیل شده است، بدست می‌آید. شکل (۲-۱۵) روش تشکیل این کد را برای حرف «ر» نشان می‌دهد. نمونه‌های مختلفی از کدگذاری زنجیره‌ای وجود دارد به عنوان مثال در [23] قاب کاراکتر با ۵ پنجره متحرک در جهت چپ به راست پوشانده می‌شود و هر ناحیه توسط کد زنجیره‌ای کد می‌گردد.



شکل ۲-۱۵- حرف «ر» و کد زنجیره‌ای مربوط به تصویر نازک‌سازی شده آن

د- گرافها و درختها

در ابتدا کلمات یا حروف به مجموعه‌ای از قسمتهای بنیادی توپولوژیکی نظیر پاره‌خطها، حلقه‌ها نقاط تقاطع و غیره تقسیم‌بندی می‌شوند. سپس این قسمتهای بنیادی توسط گرافهای متناسب شده^۲ یا پیوندی^۳ بازنمایی می‌گردند. دو نوع روش بازنمایی تصویر بوسیله گراف وجود دارد. روش اول از مختصات شکل حروف استفاده می‌کند. در روش دوم، گره‌ها متناظر با زیرحرفها می‌باشند.

^۱ Freeman

^۲ Attributed

^۳ Relational

همچنین می توان از درختها بمنظور بازنمایی کلمات یا حروف به همراه مجموعه ای از ویژگیها که با یکدیگر ارتباط سلسله مراتبی دارند، بهره جست.

فصل سوم

بررسی پایان نامه ها و

مقالات در زمینه OCR

در این فصل به تعدادی از روشهای شناسایی حروف فارسی و عربی که در پایان‌نامه‌های کارشناسی ارشد و دکترای داخل کشور^۱ و مقالات داخلی و خارجی معرفی شده‌اند، اشاره می‌کنیم.

«عزمی» در [۸۲ و ۸۶] سیستمی جهت بازشناسی متون چاپی فارسی با استفاده از رویکرد ترکیبی (بکارگیری هر دو رویکرد مبتنی بر جداسازی کلمات و مبتنی بر بازشناسی کلمه به عنوان یک الگوی واحد) ارائه نموده است. در این روش ابتدا خطوط متن از یکدیگر جدا شده و ضخامت قلم تعیین می‌شود. سپس زیرکلمات با توجه به هیستوگرام عمودی تصویر متن جدا می‌گردند. بعد از آن نوار زمینه با محاسبه هیستوگرام افقی تعیین می‌شود. سپس با کمک یک الگوریتم اصلاحی، نوار زمینه برای هر زیر کلمه جداگانه تصحیح می‌گردد. در تعیین دقیق نوار زمینه از منحنی پیرامونی زیرکلمات استفاده می‌شود. نقاط جداسازی حروف، با استفاده از کانتور بالایی کلمات تعیین می‌شوند. نقاط جداسازی با اعمال قواعدی در قالب یک دیاگرام حالت روی منحنی پیرامونی کلمات مشخص می‌گردند. در یک آزمون روی چند صفحه متن چاپی شامل حدود ۱۱ هزار حرف با قلمهای مختلف، حدود ۹۹ درصد از حروف بدرستی جداسازی شده‌اند. برای بازشناسی حروف با قلمهای گوناگون، دو الگوریتم مختلف ارائه شده است. الگوریتم اول از کدهای فریم کانتور حروف به عنوان ویژگی و از یک اتوماتون آماری برای طبقه‌بندی استفاده می‌نماید. الگوریتم دوم از تبدیل هاف در مرحله استخراج ویژگی و از یک شبکه عصبی پس انتشار خطا در مرحله طبقه‌بندی استفاده می‌کند. الگوریتم‌های فوق روی مجموعه‌ای از حروف شامل ۱۱۵۰۰ نمونه از دو قلم مختلف آزموده شده‌اند و میزان بازشناسی به ترتیب ۹۷/۱۳ و ۹۸/۳۲ درصد بوده است. در این تحقیق برای بکارگیری شکل کلی زیر کلمات در بازشناسی آنها، از دیکشنری تصویری استفاده شده است. به این منظور، سه دیکشنری مختلف با استفاده از ویژگیهای مکان مشخصه، توصیف کننده‌های فوریه و بر چسب‌های کانتور بالایی طراحی شده و آزمایش‌های مختلفی روی آنها انجام گرفته است. در مرحله نهایی این تحقیق سیستمی برای بازشناسی زیرکلمات فارسی با یک رویکرد ترکیبی ارائه شده است. در این سیستم با جداسازی و بازشناسی حروف شاخص زیرکلمه، رویکرد مبتنی بر جداسازی و با در نظر

^۱ بیشتر این پایان‌نامه‌ها از طریق وب سایت مرکز اطلاعات و مدارک علمی ایران (<http://www.irandoc.ac.ir>) در دسترس قرار گرفته-

گرفتن یک دیکشنری تصویری برای بدنه زیرکلمات، رویکرد بازشناسی کلمه بعنوان یک الگوی واحد در نظر گرفته شده است. نوع و موقعیت نقاط و علائم نیز در بازشناسی زیر کلمات بکار گرفته شده‌اند. اطلاعات آماری درباره احتمال رخداد متوالی حروف شاخص و امتیاز حاصل از بازشناسی آنها به کمک مدل مخفی مارکوف (HMM) و یک الگوریتم ویتربی^۱ تغییر یافته برای بازشناسی زیر کلمات بکار گرفته شده‌اند. سیستم ترکیبی نهایی روی یک متن با چهار قلم مختلف آزموده شده و نتایج با جزئیات مختلف ارائه شده است.

«مسروری» در [۸۸] روشهایی برای شناسایی کلمات دستنویس فارسی در یک مجموعه محدود بدون قطعه‌بندی کلمات ارائه نموده است. این روشها در سطح کلمات یا زیر کلمات عمل می‌کنند. همچنین روشی برای تعمیم الگوریتم DTW به دوبعد معرفی شده و برای مقایسه کلمات بکار رفته است. برای آزمودن این روشها، از یک مجموعه محدود از نام ۵۰۰ شهر پرجمعیت ایران استفاده شده است. نتایج این پایان‌نامه نشان می‌دهد که از بین روشهای ارائه شده، روش DTW دوبعدی، بیشترین میزان بازشناسی را دارد. در آزمایش این روش برای مجموعه نامهای ۱۰۰، ۳۰۰ و ۵۰۰ شهر ایران، حدود ۸۸ درصد از کلمات به درستی شناسایی شده‌اند.

«مسروری» در [۸۹] روشی برای بازشناسی حروف مجزای دستنویس فارسی ارائه نموده است. وی ۳۳ کلاس برای حروف فارسی در نظر گرفته است. بازشناسی حروف در دو مرحله انجام می‌گیرد. در مرحله اول به کمک یک طبقه‌بندی کننده فازی "قاعده - پایه" که قوانین آن به وسیله مجموعه نمونه‌های آموزشی یاد گرفته می‌شوند، حروف به ۸ گروه تقسیم می‌گردند. ویژگی‌های مورد استفاده در این مرحله با روش مکانهای مشخصه در راستاهای ۹۰ و ۴۵ درجه از تصویر باینری حرف بدست می‌آید. در مرحله دوم، به وسیله طبقه‌بندی کننده‌های درخت تصمیم که برای هر گروه به طور خاص طراحی شده‌اند، بازشناسی نهایی حروف در هر گروه انجام می‌گیرد. در این مرحله، ویژگی‌های به کار رفته عموماً ساختاری هستند. مجموعه نمونه‌های جمع‌آوری شده شامل ۴۹۵۰ نمونه (۱۵۰ نمونه از هر حرف) است که توسط نویسنده‌های مختلف با سطوح تحصیلاتی متفاوت

^۱ Viterbi

نوشته شده‌اند. ۳۳۰۰ نمونه از این مجموعه (۱۰۰ نمونه از هر حرف) برای یادگیری و ۱۶۵۰ نمونه دیگر برای آزمایش به کار برده شده است که طبق نتایج گزارش شده، ۹۲ درصد از این نمونه‌های آزمایش بدرستی بازشناسی شده‌اند.

«مقسمی» در [۹۱] یک روش ساختاری جهت شناسایی حروف تایی فارسی معرفی نموده است. کد زنجیره‌ای که جهت حرکت قلم فرضی هنگام نوشتن حروف را مشخص می‌سازد، به عنوان ویژگی در نظر گرفته شده است. کد زنجیره‌ای از روی محیط مرزی حروف نازک شده بدست می‌آید. مسائل مهم در این روش عبارتند از جداسازی حروف و کاهش طول کد زنجیره‌ای. جهت جداسازی از روش مجموع فواصل نقاط تا خط پایه به همراه بررسی حرکت افقی روی خط مبنا استفاده می‌شود. جهت کاهش طول بردار ویژگی، روشی تحت عنوان روش سیستم خبره به کار گرفته شده است. الگوریتم مستقل از اندازه حروف است. همچنین این روش را بدون هیچگونه تغییری می‌توان جهت شناسایی حروف تایی انگلیسی، ارقام دستنویس فارسی و انگلیسی به کار برد. جهت طبقه‌بندی نهایی از ترکیب بانک اطلاعاتی و شبکه عصبی استفاده شده است.

«اسدی» در [۶۳] یک روش بازشناسی حروف تایی فارسی بصورت ساختاری و با پارامتری فرض کردن حروف پیشنهاد نموده است. در مرحله استخراج ویژگی‌ها از ویژگی‌های ساختاری درصد نقاط سیاه بالای خط زمینه، تعداد حفره‌های حروف، تعداد قطعات تشکیل دهنده حرف، نوع اتصال حرف به حروف راست و چپ و غیره استفاده گردیده است. این ویژگی‌ها بگونه‌ای مورد استفاده قرار می‌گیرند که بتوانند نمایانگر حروف مختلف بصورت یکتا باشند. با استفاده از جداول و درخت‌های تصمیم‌گیری، روشی جهت بازشناسی حروف تایی فارسی ارائه می‌گردد. مسائل مهم در این پایان‌نامه عبارتند از: جداسازی حروف که بر اساس موقعیت خط زمینه صورت می‌پذیرد، استخراج ویژگی‌های مناسب و بازشناسی حروف فارسی که با استفاده از بانک‌های اطلاعاتی و درخت‌های تصمیم‌گیری انجام می‌گیرد. روش دیگر بازشناسی که در این تحقیق مورد استفاده قرار گرفته است، استفاده از Padalines شبکه عصبی چند جمله‌ای‌های جداکننده است. در این روش تعدادی از ویژگی‌های استخراج شده تبدیل به ثابت‌های برای چند جمله‌ای‌های مختلف می‌گردد و برای

حالات مختلف حروف یک چند جمله‌ای ارائه می‌گردد. الگوریتم‌ها مستقل از اندازه حروف می‌باشند. روشهای ارائه شده در این پایان‌نامه بدون هیچگونه تغییری میتواند برای بازشناسی حروف تایی لاتین نیز بکار رود. برای طبقه‌بندی نهایی در هر دو روش بازشناسی نیز از بانک‌های اطلاعاتی استفاده شده است.

«دهقانی» [۷۲] از الگوریتمی مبتنی بر مدل پنهان مارکوف (HMM) در بازشناسی متون فارسی بهره برده است. در مرحله پیش‌پردازش جهت فائق آمدن بر نویزهای موجود در تصاویر ورودی، چندین عمل پیش‌پردازی از قبیل فیلتر کردن میانه‌ای، باینری کردن، اعمال فیلترهای مورفولوژیکی، نرمالیزه سازی و متمرکز کردن انجام شده است. در مرحله استخراج مشخصه‌ها، چهار نوع بردار مشخصه براساس روش RPCT از هر کاراکتر استخراج می‌شود. سیستم پیشنهادی از دو قسمت تشکیل شده است. برای هر کاراکتر در مرحله آموزش چند مدل مارکوف متناظر با بردارهای مشخصه مختلف ساخته می‌شود. در مرحله بازشناسی، نتیجه تشخیص هر کدام از طبقه‌بندی کننده‌ها جهت تشکیل نتیجه نهایی بازشناسی با یکدیگر ترکیب می‌شوند. در این تحقیق نرخ بازشناسی ۹۳ درصد برای داده‌های آموزش و ۷۱ درصد برای داده‌های آزمایش گزارش شده است.

«بحری» [۶۵] از روش عصبی فازی برای شناسایی حروف دستنویس جداگانه استفاده نموده است. ابتدا الگوریتم GLVQF که یک الگوریتم خودسازمانده می‌باشد، بکار گرفته شده است تا الگوی وزنها (که بطور نظری برابر الگوی ورودیها می‌باشد) بدست آید و سپس واریانس این ورودیها نسبت به این وزنها محاسبه شده و توأم با الگوی وزنها مناطق تعلق فازی به طور مثلی و دوزنقه‌ای را بدست می‌دهد. این مناطق در نوروهای فازی لایه دوم شبکه به خاطر سپرده شده و رویهم مقدار تعلق ورودیهای مورد شناسایی نسبت به دسته وزنها را مشخص می‌سازند. سپس این اطلاعات بدست آمده توسط لایه سوم که یک لایه پرسپترون می‌باشند فازی‌زدائی می‌گردند. در این لایه فازی‌زدائی بهینه شده است که با وجود تعداد قوانین کم فازی نتایج قابل قبولی را بدست می‌دهد.

«ثانی» [۶۹] از یک روش مبتنی بر شبکه عصبی بمنظور بازشناسی حروف دست‌نویس فارسی استفاده نموده است. ورودی سیستم فرمهایی است که شامل جداول مختلفی می‌باشند و

حروف بصورت دست‌نویس در این جداول وارد شده‌اند. ابزار تهیه فرم بصورت دلخواه نیز بعنوان پیش‌نیاز این پروژه پیاده‌سازی شده است. ورودی شبکه BPN الگوی بیتی انتخابی از تصویر حروف مورد نظر می‌باشد و خروجی سیستم یکی از ۳۲ کلاس الفبای فارسی است. سیستم پیشنهادی قادر است نزدیکی الگوی ورودی به هر یک از کلاسها را بصورت احتمالات بیان کند.

«صدوقی یزدی» در [۷۷] به مبحث دو سطحی کردن تصویر و جداسازی تصویری کلمات چاپی فارسی و لاتین پرداخته است. در بخشی از این تحقیق، پس از بررسی و آزمایش روشهای مختلف دو سطحی کردن تصاویر، الگوریتمی ارائه شده است که روش خوشه یابی با دو میانگین و روش محلی مورین را ترکیب می‌کند. این الگوریتم در مقابل نایکنواختی روشنایی و وجود لکه‌های موضعی در تصویر مقاوم است. بخش دیگر این تحقیق درباره جداسازی تصویری کلمات فارسی و لاتین است. از بین ویژگیهای مختلف هندسی، پنج ویژگی که بین کلمات فارسی و لاتین تمایز خوبی ایجاد می‌کنند، انتخاب شده‌اند. برای جداسازی کلمات، یک طبقه بندی کننده حداقل فاصله از متوسط هر کلاس بکار رفته است. پارامترهای الگوریتم طبقه بندی با استفاده از یک مجموعه تمرین شامل حدود ۲۰۰۰ کلمه فارسی و لاتین محاسبه شده است. در یک آزمایش برای یک مجموعه آزمون شامل حدود ۲۰۰۰ کلمه دیگر ۹۵ درصد از کلمات به درستی طبقه‌بندی شدند.

«شاه حسینی» [۷۵] روشی برای شناسایی حروف دست‌نویس فارسی با استفاده از شبکه‌های عصبی معرفی نموده است. شبکه عصبی مورد استفاده، یک طبقه‌بندی‌کننده سه لایه پرسپترون با قانون یادگیری پس انتشار می‌باشد. روشهای مناسبی برای بالابردن کارایی الگوریتم آموزش شبکه ارائه شده است. سه نوع گشتاور هندسی به عنوان ویژگی مورد استفاده قرار گرفته‌اند. مرحله عمده الگوریتم بازشناسی، گروه‌بندی اولیه حروف در ۱۷ گروه است. گروههای هفده‌گانه براساس نتایج حاصل از طبقه‌بندی حروف توسط یک شبکه عصبی بر روی ۳۳ کلاس تعریف شده‌اند. هر گروه شامل یک، دو، سه یا چهار حرف است که با توجه به ویژگیهای به کار رفته در این طبقه‌بندی با یکدیگر اشتباه می‌شوند. برای گروه‌بندی حروف شبکه‌ای با ۲۲ گره ورودی به تعداد ویژگیهای گشتاوری، ۳۵ گره لایه پنهان و ۱۷ گره خارجی بکار گرفته شده است. در مرحله دیگری از الگوریتم

بازشناسی حروف هر گروه با شبکه‌هایی با ابعاد کوچکتر شناسایی می‌شوند. مجموعه نمونه حروف در این تحقیق شامل ۱۶۵۰ نمونه آموزش و ۱۵۶۵ نمونه آزمون است. از این نمونه‌ها با یک روبشگر (اسکنر) دستی با دقت ۲۰۰ نقطه در اینچ تصویربرداری شده است. ۹۷/۸۲ درصد از نمونه‌های آموزش و ۹۰/۳۵ درصد از نمونه‌های آزمون به درستی دسته‌بندی شده‌اند. متوسط میزان شناسایی حروف در گروه‌های هفده‌گانه ۸۹/۴۸ درصد می‌باشد.

«تیمساری» [۶۸] روشی برای شناسایی حروف تشکیل دهنده کلمات در متون تایپ شده فارسی ارائه نموده است. روش پیشنهادی بر اساس اعمال عملگرهای مورفولوژیکی بر روی تصویر باینری کلمه و بازشناسی حروف بر مبنای تصمیم‌گیری روی این ویژگی‌ها پایه‌ریزی شده است. در این روش، اعمال جداسازی و تشخیص حروف تشکیل دهنده یک کلمه بطور همزمان صورت می‌گیرند. روند اجرای الگوریتم بدین صورت است که ابتدا با محاسبه پروجکشن (projection) افقی و سپس قائم، متن به خطوط و خطوط به زیر کلمات تفکیک می‌شوند. سپس نقطه‌های موجود در حروف سازنده زیرکلمه استخراج شده، متعاقب آن عمل نازکسازی روی تصویر باقیمانده انجام می‌پذیرد. برای انجام اینکار از الگوریتمی سریع بر مبنای عملگر مورفولوژیکی Hit/Miss استفاده شده است. استخراج ویژگی‌های لازم برای شناسایی حروف، توسط اعمال عملگر Hit/Miss با المانهای ساختاری مناسب روی اسکلت تصویر صورت می‌گیرد. تصمیم‌گیری نهایی بر روی ویژگی‌های ساختاری بدست آمده و تشخیص حروف، براساس یکسری قواعد از پیش تعیین شده انجام می‌گیرد.

«نمازی» [۹۴] الگوریتمی بر مبنای منطق فازی و ترکیب آن با شبکه عصبی بمنظور بازشناسی حروف چاپی فارسی با قلم‌های متفاوت نموده است. در این سیستم تصاویر توسط اسکنر و با دقت ۳۰۰ نقطه بر اینچ وارد کامپیوتر می‌شوند و سپس نازک می‌شوند. در این روش از گشتاورهای شبه زرنیکه به عنوان ویژگی اصلی تصویر ورودی استفاده شده است. پس از فازی کردن این گشتاورها توسط توابع عضویت مربوطه، چند شبکه عصبی با توپولوژی‌های متفاوت برای طبقه‌بندی حروف آموزش داده شده‌اند و در آخر، طبقه‌بندی کننده نهایی با استفاده از این شبکه‌ها اقدام به تشخیص الگوهای ورودی می‌نماید. آموزش و تست این سیستم با استفاده از حدود ۳۷۰۰ نمونه

چاپ شده انجام شده است که نسبت نمونه های آموزشی بیش از ۳ به یک می باشد. دقت بدست آمده در این سیستم بیش از ۹۹/۸ درصد می باشد. پراکندگی دقت سیستم در کلاسهای مختلف حدود ۲ درصد است. همچنین در این تحقیق نشان داده شده است که در شرایط مساوی، قدرت تفکیک الگوهای ورودی در محیطهای مبهم برای شبکه فازی، بالاتر از شبکه عادی می باشد.

«اردشیر بهرستانی» در [۶۲] سعی در بازشناسی حروف مجزای دستنویس فارسی که توسط افراد مختلف نوشته شده اند، داشته است. فرض بر این بوده که حروف با دقت ۲۰۰ نقطه در اینج تصویربرداری و با کد MSP ذخیره شده اند. در این تحقیق ابتدا یک مرحله پیش پردازش برای رفع بریدگی و حفره و نویز صورت می گیرد. سپس با اجرای الگوریتم برچسب زنی مؤلفه ها، نقطه و سرکش موجود، از بدنه اصلی حرف جدا و نوع آن تشخیص داده می شود. در نهایت با گروه بندی بخشهای سطرهای مختلف ماتریس تصویر حرف، زیر حروف استخراج و مورد شناسایی قرار می گیرند. برای طبقه بندی حروف از یک درخت تصمیم استفاده گردیده است. برای یک مجموعه متشکل از ۶۱ نمونه از هر حرف، میزان بازشناسی ۹۱/۵ درصد و برای مجموعه ای دیگر متشکل از ۴۰ نمونه از هر حرف، بازشناسی صحیح ۹۲/۱۲ درصد گزارش گردیده است.

«مقدم تبریزی» [۹۰] روشهایی برای شناسایی و استخراج سمبل های مربوط به نقشه های جغرافیایی و یا فنی از متن نقشه، و یا بطور کلی استخراج یک شکل و یا سمبل از تصویری که ممکن است حاوی خطوط، منحنی ها و یا نوشته باشد، پیشنهاد نموده است. هدف از این تحقیق، سعی در مکانیزه کردن روند تبدیل نقشه های دستی و تصویری به نقشه های کامپیوتری برای سیستم های GIS می باشد. برای نیل به این منظور، از سه روش مبتنی بر شبکه عصبی هاپفیلد، شبکه عصبی پرسپترون چندلایه و الگوریتم ژنتیک استفاده شده است. در روش اول کاهش انرژی شبکه منجر به رفع اغتشاشات شده و بدین طریق سمبل شناسایی می شود. در روش دوم با استفاده از سیستمی مرکب از چند شبکه MLP ابتدا منحنی های سمبل و سپس تمامی سمبل شناسایی می گردد. در روش سوم، معرفی تابع هدفی خاص و بهینه سازی آن با روش الگوریتم ژنتیک اساس شناسایی تصویر است.

نتایج حاصل از این تحقیق نشان می دهد که روش مبتنی بر شبکه عصبی پرسپترون چندلایه، بر دو روش دیگر برتری دارد.

«عباسیان» در [۸۱] الگوریتمی برای بازشناسی برخط حروف فارسی ارائه نموده است. این الگوریتم ۶۱ حرف را براساس ویژگیهای پویا و ایستای آنها و با استفاده از یک درخت تصمیم باینری که به شکل تجربی طراحی شده است، بازشناسی می کند. در این الگوریتم ابتدا با استخراج ویژگیهای زمانی و مکانی لازم، مجموعه ای از الگوهای پایه در ساختار حرف بازشناسی شده اند. این الگوهای پایه عبارتند از: پاره خطهای افقی، عمودی و مورب، منحنی بسته، زیرحرف S مانند و کمانهایی با گودی رو به چپ یا راست. در الگوریتم بازشناسی نهایی، ابتدا حروف بر اساس تعداد بخشهای آنها به چهار گروه اولیه تقسیم می شوند و سپس حروف هر گروه با استفاده از یک درخت تصمیم شناسایی می گردند. در طراحی الگوریتم از یک مجموعه آموزشی با ۲۰ نمونه از هر حرف استفاده شده است. ۹۳/۷ درصد از نویسه های یک مجموعه آزمایش شامل ۴۵ نمونه از هر حرف بدرستی شناخته شده و ۱/۹ درصد از آنها وازده شده اند. در یک آزمایش دیگر برای حروف مجزای دستنویس بر روی مجموعه ای شامل ۵۰ نمونه از هر حرف، ۹۴/۷ درصد از حروف شناخته شده و ۲/۳۶ درصد وازده شده اند. برای بازشناسی ارقام، درخت تصمیم دیگری طراحی شده است که برای تمرین آن، از یک مجموعه با ۱۰ نمونه از هر رقم استفاده شده است. با این الگوریتم ۹۴/۸ درصد از ارقام یک مجموعه آزمایش شامل ۲۵ نمونه از هر رقم، به درستی شناخته شده و ۰/۸ درصد از آنها وازده شده اند.

«رضوی» در [۷۳] روشی برای خواندن اتوماتیک فرمهای انتخاب درس ارائه نموده است. در هر فرم، حروف و ارقام بصورت جداگانه در داخل مستطیلهایی نوشته می شوند. یک الگوریتم بازشناسی حروف و ارقام که از روش N-tuple و اطلاعات پیش زمینه ای در سطح حرف (طول کد حرفی و موقعیت مکانی هر کاراکتر) استفاده می کند، طراحی شده است. برای تشخیص نهایی کد درس از روش جستجوی واژه نامه استفاده می شود. اثر تعداد N-tuple ها و چگونگی انتخاب آنها و تعداد نمونه های تمرین در میزان بازشناسی مورد بررسی شده است. در بهترین حالت (انتخاب ۸۵-3

tuple ناهمپوشان) ۱۰۰ نمونه تمرین در هر کلاس و استفاده از روش N-tuple اصلاح شده برای ۵۰ نمونه آزمایش در هر کلاس، ۹/۸۰ درصد از حروف و ۸۱ درصد از ارقام به درستی بازشناسی شدند. برای آزمایش، ۱۳۴ فرم تکمیل شده، شامل ۴۵۲ درس انتخابی، از ۶۷ دانشجو جمع آوری شد. از این فرمها با درجهء تفکیک ۲۰۰ نقطه در اینچ، به صورت باینری تصویربرداری و در کاراکتر در ابعاد ۱۶×۱۶ نرمالیزه شد. در الگوریتم نهایی خواندن فرم، میزان بازشناسی درست برای کد درس ۹۵/۱ درصد، میزان خطا ۰/۸۸ درصد و میزان وازدگی ۳/۹۸ درصد است.

«رفیعی کراچی» در [۷۴] یک روش جداسازی حروف برای متون تایپ شده فارسی پیشنهاد نموده است. این روش جداسازی مستقل از رسم الخطها می باشد و بر اساس پروفیل بالایی تصویر زیر کلمه انجام می گیرد. الگوریتم جداسازی حروف شامل سه مرحله پردازش اولیه، پیدا کردن محل اتصال حروف و پس پردازشها می باشد. پیش پردازشها دربرگیرنده جداسازی خطوط متن، محاسبه پهنای قلم، پیدا کردن نوار زمینه و جداسازی زیر کلمات می باشند. در مرحله دوم ابتدا پروفیل بالایی تصویر زیر کلمه بدست می آید. سپس با استفاده از این پروفیل و بکارگیری قواعد ساختاری محل اتصال حروف بدست می آید. در این مرحله بعضی از حروف به زیر حروف می شکنند. از پس پردازشها برای بازسازی این حروف استفاده می شود. پس پردازشها شامل قواعدی مبتنی بر رفتار بینایی انسان در جداسازی حروف و قواعد نگارشی حروف فارسی می باشند. متون نمونه بکار رفته در این تحقیق شامل ۱۳۴۴۲ حرف در ۲۶ رسم الخط مختلف هستند. این متون با دقت ۲۰۰ نقطه در اینچ تصویربرداری شده اند. در این مجموعه که ۶۷/۳۱ درصد از حروف به یکدیگر متصل هستند که ۹۹/۱۸ درصد از آنها بدرستی از یکدیگر جدا شده اند.

«امیری» [۶۴] روشی برای جداسازی متن چاپی فارسی از تصاویر گرافیکی ارائه نموده است. جدولها، روندنماها^۱ و نقشه ها از جمله تصویرهای گرافیکی هستند. این روش از سه مرحله تشکیل شده است. در مرحله اول عناصر گرافیکی شاخص با توجه به برخی ویژگیهای تصویری آنها جدا می شوند. در مرحله دوم متنهای چسبیده به عناصر گرافیکی آشکار و جدا می گردند. در مرحله سوم،

^۱ Flowchart

با استفاده از طبقه‌بندی حداقل فاصله، عناصر گرافیکی که در تصویر متن خروجی مرحله دوم باقی مانده‌اند، جدا می‌شوند. در این تحقیق از یک مجموعه تمرین شامل ۴۲ تصویر و یک مجموعه آزمایش شامل ۳۰ تصویر استفاده شده است. برای مجموعه آزمایش ۹۱/۱ درصد از عناصر متن و ۸۷/۷ درصد از عناصر گرافیک درست طبقه‌بندی شده‌اند. منابع خطا و میزان تاثیر هر یک از آنها بررسی شده‌اند و پیشنهادهایی برای کاهش آنها ارائه شده است.

«مرتضی‌پور» [۸۷] الگوریتمی برای قطعه‌بندی برخط کلمات دستنویس فارسی ارائه نموده است. منظور از قطعه‌بندی در اینجا یافتن قطعاتی است که بوسیله آنها بتوان کلیه کلمات را ایجاد کرد؛ بدین معنی که مجموعه کلمات نوشته شده توسط افراد مختلف به اجزاء ساده‌تری تقسیم می‌شوند تا برای بازشناسی مورد استفاده قرار گیرند. این الگوریتم بر اساس یافتن نقاط اصلی در یک کلمه و مقایسه شیب و زوایای پاره‌خطهای متصل‌کننده این نقاط به یکدیگر، قطعه‌بندی را انجام می‌دهد. برای جمع‌آوری کلمات دستنویس از صفحه رقومی‌کننده با قدرت تفکیک ۱۲۵ نقطه در هر اینچ و سرعت نمونه‌برداری ۳۳۳ نمونه در ثانیه استفاده گردیده است. در طراحی الگوریتم از یک مجموعه تمرین با ۲۰ کلمه و از هر کلمه ۲ نمونه دستنویس افراد مختلف استفاده شده است. در آزمایش این الگوریتم، از یک مجموعه آزمایش با ۲۰ کلمه و از هر کلمه ۲ نمونه از همان افراد استفاده شده است. در این آزمایش برای ۳ کلمه، قطعه‌بندی درست انجام نشد.

«نظام‌آبادی‌پور» در [۹۳] با اصلاح الگوریتم «عزمی» [۸۲] که مبتنی بر کانتور بالایی است، یک الگوریتم جداسازی برای متون چاپی قدیمی ارائه نموده است. برای حل مشکل نایک‌نواختی کرسی خط، روشی برای تعیین نوار زمینه پیشنهاد گردیده و با اصلاح روش برچسب‌زنی کانتور بالایی و تکمیل قواعد جداسازی، دقت الگوریتم برای متون قدیمی افزایش یافته است. نرخ جداسازی درست حروف با استفاده از این الگوریتم برای یک مجموعه آزمایش که از متون قدیمی برگزیده شده است، بدون هیچ پس‌پردازشی ۹۷ درصد گزارش شده است. برای جداسازی نقاط چسبیده به بدنه حروف نیز روشی ارائه شده است.

«نحوی» در [۹۲] برای بعضی از مراحل اتوماسیون پست و بازشناسی آدرسهای پستی فارسی روشهایی پیشنهاد نموده است. با توجه به تفاوت‌های ساختاری آدرسهای دستنویس و چاپی، سیستم‌های جداگانه‌ای برای پردازش آنها لازم منظور شده است. در این تحقیق با استفاده از ویژگیهای مناسب و روش حداقل فاصله از متوسط هر کلاس آدرسهای دستنویس از چاپی جدا و سپس بر اساس مجموعه‌ای از ویژگیهای مناسب و تصمیم‌گیریهای سلسله‌مراتبی، کدپستی از آدرس‌های چاپی و دستنویس استخراج می‌شود. ۵۰ آدرس دستنویس و ۵۰ آدرس چاپی با ۱۲ قلم گوناگون به عنوان مجموعه تمرین استفاده شده است. مجموعه آزمایش شامل ۱۵۰ آدرس دستنویس و ۱۵۰ آدرس چاپی متفاوت از مجموعه تمرین است که با درجه تفکیک ۲۰۰ نقطه در اینچ و به شکل باینری تصویربرداری شده‌اند. در تفکیک آدرسهای چاپی و دستنویس آزمایش نشان می‌دهد که ۹۸ درصد از آدرسهای مجموعه تمرین و ۶۷/۹۷ درصد از آدرسهای مجموعه آزمایش به درستی طبقه‌بندی شده‌اند. برای مجموعه تمرین، در ۹۸ درصد از آدرسهای چاپی و ۸۲ درصد از آدرسهای دستنویس، کد پستی به درستی استخراج شده است. در مجموعه آزمایش، در ۹۶/۸۸ درصد از آدرسهای چاپی و ۷۸/۶۷ درصد از آدرسهای دستنویس کدپستی به صورت صحیح استخراج شده است.

«بنی‌اسدی» در [۶۶] تشخیص حروف دست‌نویس فارسی به وسیله سیستم هایبرید نور و فازی را مورد توجه قرار داده است. نقش منطق فازی در این پایان‌نامه مدیریت عدم قطعیت موجود در اطلاعات ورودی است. در واقع مرحله استخراج خصوصیات در این پایان‌نامه به کمک منطق فازی و براساس محاسبه درجه عضویت هر حرف در هفت مجموعه فازی مختلف انجام می‌گیرد. در مرحله طبقه‌بندی نهایی شبکه‌های عصبی استفاده شده است. شبکه عصبی مورد استفاده شبکه پیش‌خور چند لایه (MLP) می‌باشد که به صورت ۴ لایه شبیه‌سازی شده است.

Faez, Mowlaei و Haghighat [44] روشی بمنظور بازشناسی کدهای پستی و اسامی شهرهای ایران که بر روی بسته‌های پستی بصورت دست‌نویس نوشته می‌شوند، پیشنهاد نموده‌اند. بردارهای ویژگی این روش، از طریق بکارگیری تبدیل موجک گسسته (DWT) با موجک اساسی

Haar محاسبه می گردند. در مرحله پیش پردازش این الگوریتم، ابتدا محدوده هر حرف مشخص گردیده، تصویر حرف به ابعاد 64×64 پیکسل تغییر اندازه می یابد و در نهایت عمل نازک سازی انجام می گیرد. در مرحله استخراج ویژگیها، تبدیل موجک سه سطحی بر روی تصویر نازک شده اعمال می شود و ضرایب باند LL از سطح سوم استخراج می گردند. در انتها ماتریس ضرایب باند LL بصورت برداری در آمده، یک فیلتر پایین گذر بمنظور هموارسازی بیشتر ضرایب، به آن اعمال می شود. برای آموزش سیستم، یک شبکه عصبی MLP با قانون پس انتشار خطا مورد استفاده قرار می گیرد. راندمان این روش برای حروف و ارقام دست نویس فارسی بترتیب $92/33$ و $91/81$ درصد گزارش شده است.

Korsheed و Clocksin [29] الگوریتمی برای بازشناسی کلمات عربی بدون تقطیع آنها^۱ طراحی نموده اند که در آن ابتدا هر کلمه به یک تصویر قطبی نرمالیزه تبدیل می شود و سپس یک تبدیل فوری دوبعدی به تصویر قطبی اعمال می گردد. طیف منتج شده نسبت به دوران، تغییر اندازه و تغییر مکان کلمات، غیر حساس می باشد. مرحله بازشناسی با بکارگیری تابع فاصله اقلیدسی نرمالیزه شده انجام می گیرد. جهت آزمایش این الگوریتم، بیش از ۱۷۰۰ کلمه در ۴ فونت مختلف مورد استفاده قرار می گیرد. کلمات در جهتهای مختلف (۰ تا ۱۸۰ درجه) و اندازه های متفاوت (۱۸ تا ۴۸pt) در نظر گرفته می شوند. نرخ خطا برای این سیستم در بهترین حالت (با بکارگیری ۱۶ باند فرکانسی پایینی) در حدود ۶ درصد برآورد شده است.

Menhaj و adab [41] الگوریتمی بمنظور تقطیع و بازشناسی حروف فارسی در فونتها و اندازه های متفاوت با استفاده از شبکه های عصبی چند لایه با تغذیه رو به جلو^۲ پیشنهاد نموده اند. فیدبک اتوماتیک بین مراحل تقطیع و بازشناسی بمیزان زیادی، خطاهای ناشی از تقطیع را حذف می نماید. مرحله بازشناسی، از شبکه های عصبی MLP با توصیف کننده های فوری بهره می گیرد. راندمان این روش برای متون copy/paste شده ۱۰۰ درصد و برای متون اسکن شده با درجه تفکیک dpi ۴۰۰، ۹۳ درصد گزارش شده است.

^۱ Segmentation free

^۲ Feedforward

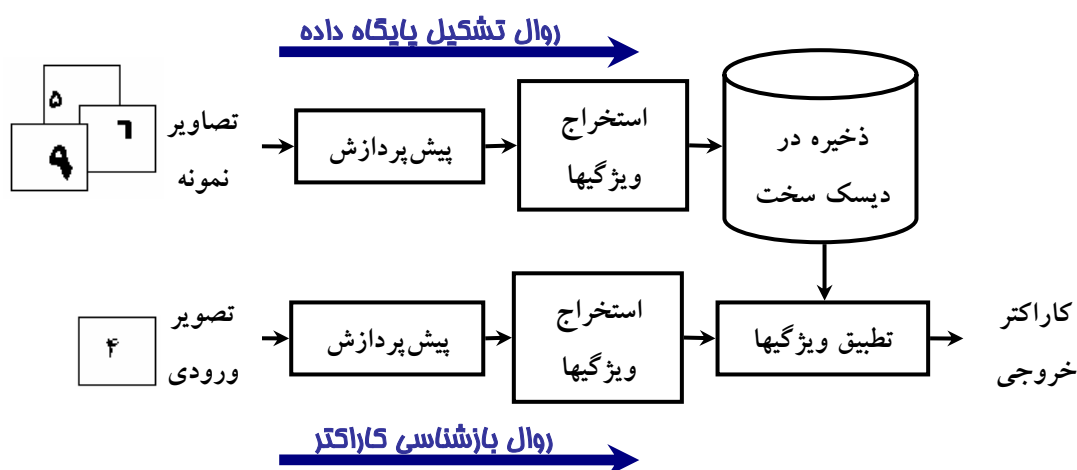
فصل چهارم

**پیاده‌سازی یک نرم‌افزار
تشخیص خودکار ارقام
فارسی**

در فصول ۲ و ۳ به برخی از روشهای معرفی شده بمنظور بازشناسی حروف فارسی و لاتین اشاره شد. در این طرح یک برنامه کامپیوتری آزمایشی جهت تشخیص خودکار ارقام فارسی پیاده‌سازی گردیده است. این برنامه به زبان برنامه‌نویسی Matlab نوشته شده است و قادر به شناسایی ارقام فارسی با دقت و سرعت بالایی می‌باشد. در ادامه به شرح این برنامه می‌پردازیم.

۴-۱- بلوک دیاگرام سیستم

بلوک دیاگرام سیستم پیشنهادی در شکل (۴-۱) نشان داده شده است.



شکل ۴-۱- بلوک دیاگرام سیستم پیشنهادی

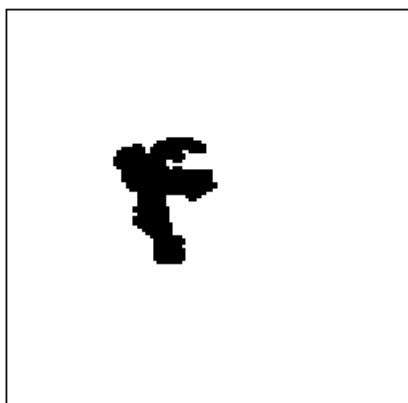
الگوریتم این سیستم متشکل از دو روال پردازشی اصلی است:

الف- روال تشکیل پایگاه داده.

ب- روال بازشناسی کاراکتر ورودی.

در این الگوریتم ابتدا در روال تشکیل پایگاه داده، تصاویر نمونه ارقام ۱ تا ۹ در فونتهای مختلف به سیستم داده می‌شوند. نام فایل هر یک از این تصاویر نمونه بگونه‌ای انتخاب شده است که برنامه متوجه می‌شود هر تصویر مربوط به چه کاراکتری می‌باشد. ویژگیهای هر کاراکتر از تصاویر مربوطه استخراج گردیده و در حافظه دیسک سخت کامپیوتر ذخیره می‌شوند. پس از خاتمه یافتن این مرحله، اکنون سیستم آماده است تا کاراکترهای ورودی را شناسایی نماید. از آنجائیکه در این

طرح تنها پیاده‌سازی آزمایشی یک نرم‌افزار تشخیص ارقام مد نظر بوده است، لذا فرض گردیده که تصاویر ورودی تنها حاوی یکی از ارقام ۱ تا ۹ فارسی می‌باشند و بنابراین قطعه‌بندی تصویر در اینجا مد نظر نیست. محل این کاراکتر در تصویر، اندازه و نوع فونت آن اختیاری بوده، همچنین ممکن است که این کاراکتر به نویز زیادی آلوده شده باشد. (شکل ۴-۲). در روال بازشناسی کاراکتر، بردار ویژگی تصویر ورودی استخراج شده و سپس با کلیه بردارهای ویژگی مربوط به تصاویر نمونه مطابقت داده می‌شود. کاراکتری که بردار ویژگی آن بیشترین میزان شباهت را به بردار ویژگی کاراکتر ورودی داشته باشد، به عنوان کاراکتر خروجی شناخته می‌شود. همانطور که پیشتر عنوان شد، سیستم با مشاهده نام کاراکتر خروجی، به موجودیت آن پی می‌برد.



شکل ۴-۲- نمونه‌ای از تصویر ورودی برنامه. محل کاراکتر در تصویر و اندازه و نوع فونت آن اختیاری است و ممکن است کاراکتر آلوده به نویز باشد.

۴-۲- الگوریتم برنامه تشخیص خودکار ارقام فارسی

در ادامه بخشهای مراحل مختلف الگوریتم برنامه تشریح می‌شوند.

۴-۲-۱- مرحله پیش‌پردازش

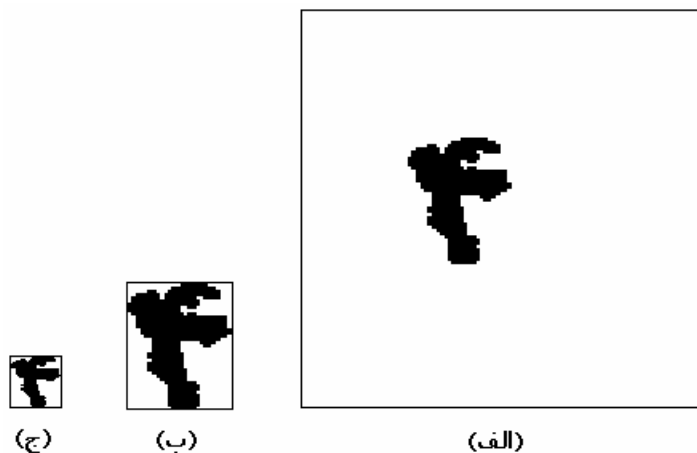
این الگوریتم دارای مراحل پیش‌پردازشی زیر است:

الف- باینری (دو سطحی) کردن تصویر ورودی.

ب- یافتن محل کاراکتر و محدود کردن تصویر ورودی به قاب دربرگیرنده کاراکتر (شکل ۴-۳-ب):

بدین ترتیب الگوریتم بازشناسی نسبت به مکان کاراکتر در تصویر غیر حساس می‌گردد.

ج- تغییر اندازه تصویر به ابعاد 32×32 پیکسل (شکل ۴-۳-ج): بدین شکل حساسیت الگوریتم بازشناسی نسبت به اندازه کاراکتر از بین می‌رود (نرمالیزه کردن اندازه).



شکل ۴-۳- اعمال مراحل پیش‌پردازشی به تصویر ورودی الف- تصویر اصلی ب- محدود کردن تصویر به قاب دربرگیرنده آن ج- نرمالیزه کردن اندازه

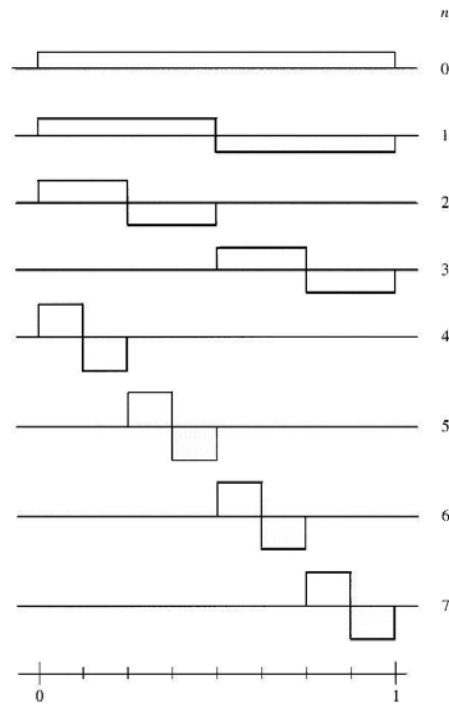
۴-۲-۲- مرحله استخراج ویژگیها

با توجه به خواص منحصر به فرد تبدیل موجک که در فصل دوم به اجمال مورد اشاره قرار گرفت، لذا این تبدیل به عنوان تابع کاندیدا جهت استخراج ویژگیها از حروف انتخاب گردید. موجکها توابعی هستند که سیگنالها را به مؤلفه‌های فرکانسی مختلف تجزیه می‌نمایند و هر مؤلفه را با یک دقت تفکیک مطابق با مقیاس آن آنالیز می‌کنند. تبدیل موجک در کاربردهای نوپرزدایی سیگنال، فشرده‌سازی تصویر، هموارسازی تصویر و آنالیز فرکتالها یک زمینه تحقیقاتی فعال بشمار می‌رود. از جمله پراستفاده‌ترین موجکها، موجک Haar می‌باشد که توابع پایه آن $h_0(x)$, $h_1(x)$, ..., $h_n(x)$... در فاصله $[0,1]$ متعامد می‌باشند (شکل ۴-۴)؛ بگونه‌ای که برای هر تابع پیوسته $f(x)$ در فاصله $[0,1]$ ، سری زیر بطور یکنواخت به $f(x)$ همگرا می‌شود:

$$\sum_{j=1}^{\infty} \langle f, h_j \rangle h_j(x)$$

در اینجا $\langle u, v \rangle$ حاصلضرب داخلی دو تابع u و v بوده و برابر است با:

$$\langle u, v \rangle = \int_0^1 u(x) \bar{v}(x) dx$$



شکل ۴-۴- توابع پایه تبدیل Haar

$\bar{v}$ در رابطه فوق مزدوج مختلط v می‌باشد.

یک نمونه از توابع پایه Haar بصورت زیر نوشته می‌شود:

$$h(x) = \begin{cases} 1 & x \in [0, 0.5) \\ -1 & x \in [0.5, 1) \\ 0 & \text{elsewhere} \end{cases} \quad h_n(x) = 2^{j/2} h(2^j x - k)$$

که در آن:

$$n = 2^j + k, \quad x \in [k2^{-j}, (k+1)2^{-j})$$

در این مرحله ابتدا با استفاده از موجک Haar، یک تبدیل موجک دوسطحی به تصویر

خروجی مرحله پیش‌پردازش اعمال می‌کنیم (شکل ۴-۵). همانگونه که در فصل دوم نیز گفته شد،

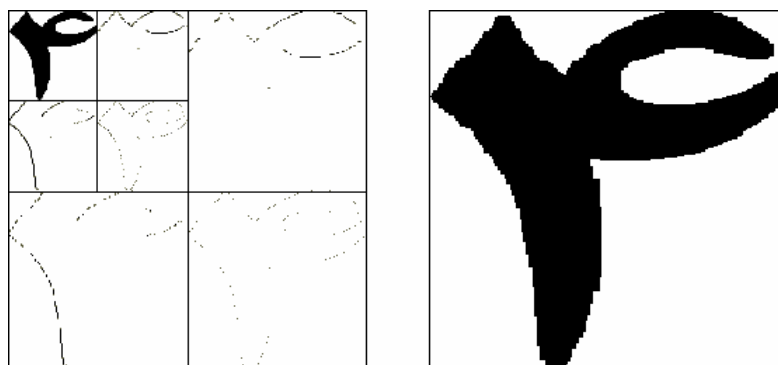
دلیل انتخاب این موجک آن است که این تبدیل به یک تصویر باینری اعمال می‌شود و در نتیجه

بدلیل آنکه توابع پایه موجک Haar ناپیوسته و دوسطحی می‌باشند، این موجک مناسبتر از سایر

موجکها بنظر می‌رسد. انجام آزمایش با موجکهای گوناگون این ادعا را تأیید نمود. پس از آن،

ضرایب باند LL پایینی (که تعداد آنها $8 \times 8 = 64$ می‌باشد) را به عنوان بردار ویژگی انتخاب می‌نماییم.

آزمایش نشان می‌دهد که انتخاب باندهای بالایی، حتی راندمان الگوریتم را کاهش خواهد داد.



شکل ۴-۵- تصویر رقم «۴» به همراه تصویر حاصل از ضرایب باندهای موجک آن. ضرایب باند LL پایینی (تصویر گوشه بالا و سمت چپ) به عنوان بردار ویژگی تصویر انتخاب می‌شود.

۴-۲-۳- مرحله تطبیق ویژگیها

در این مرحله، بردار ویژگی تصویر ورودی با کلیه بردارهای ویژگی مربوط به تصاویر نمونه

تطبیق داده می‌شوند. معیار تطبیق دهی «فاصله اقلیدسی» است که بصورت زیر تعریف می‌شود:

$$d(\bar{f}_Q, \bar{f}_D) = \sum_{k=1}^N \left(\sum_{m=1}^8 \sum_{n=1}^8 (\bar{f}_{mn}^k - \bar{f}'_{mn})^2 \right)$$

در رابطه فوق $\bar{f}_{mn}^k$ و $\bar{f}'_{mn}$ بترتیب مؤلفه‌های mn ام از بردار ویژگی k امین تصویر نمونه و

تصویر ورودی می‌باشند و N تعداد کل تصاویر نمونه است. کاراکتر متناظر با تصویر نمونه‌ای که

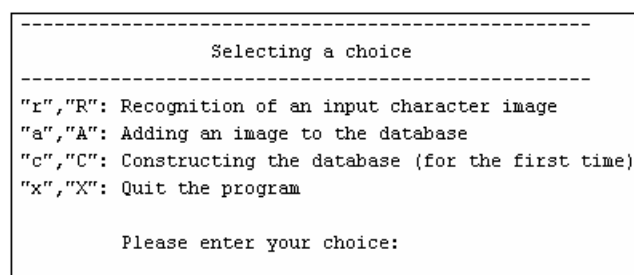
بردار ویژگی آن کمترین فاصله را با تصویر ورودی داشته باشد، به عنوان کاراکتر خروجی معرفی

می‌گردد.

۴-۳- راهنمای نرم‌افزار

این برنامه به زبان Matlab نوشته شده و در محیط Matlab 6.5 تست گردیده است. با

اجرای فایل اصلی برنامه به نام main.m منوی برنامه مطابق شکل (۴-۶) ظاهر می‌گردد.



شکل ۴-۶- منوی اصلی برنامه در محیط Matlab

این منو از کاربر می‌خواهد تا با انتخاب یکی از کاراکترهای "c"، "r"، "a" یا "x" درخواست خود را بصورت زیر به سیستم اعلام کند:

"c": اجرای روال تشکیل پایگاه داده: زمانی که کاربر برای اولین بار می‌خواهد برنامه را آزمایش کند، باید ابتدا تصاویر نمونه را به سیستم معرفی نماید تا بدین ترتیب بردارهای ویژه آنها استخراج و در حافظه ذخیره گردند. با فشردن کلید c یا C این مرحله آغاز می‌گردد و محیط اجرای برنامه، زمان پردازش هر یک از تصاویر نمونه را به کاربر گزارش می‌دهد (شکل ۴-۷). این ویژگیها تا قبل از خاتمه عملیات در حافظه Ram کامپیوتر باقی می‌مانند و در پایان، همه ویژگیها با هم در حافظه سخت دستگاه ذخیره می‌گردند.

"r": با انتخاب این گزینه، روال بازشناسی یک کاراکتر ورودی شروع می‌شود. این برنامه بصورت پیش‌فرض، تصویر "Palette.bmp" را از درایو "C:\\" دیسک سخت به عنوان ورودی می‌پذیرد. بنابراین پیش از اجرای این روال است لازم است که فایل "C:\\" "Palette.bmp" بر روی حافظه سخت دستگاه کامپیوتر موجود باشد. پس از اخذ و پردازش این فایل تصویری، پیغامی مانند شکل (۴-۸) نمایش داده می‌شود که نشان می‌دهد برنامه کاراکتر ورودی را به عنوان کدامیک از ارقام ۱ تا ۹ تشخیص داده است. با فشردن یک کلید دلخواه، برنامه به منوی اصلی برمی‌گردد. همچنین نرم‌افزار در یک پنجره دیگر بترتیب ۱۶ تصویر نمونه دارای بیشترین درجه شباهت با تصویر ورودی را به نمایش می‌گذارد (شکل ۴-۹).

"a": این روال، یک مجموعه تصاویر نمونه جدید را به مجموعه تصاویر پایگاه داده می‌افزاید؛ بدون آنکه لازم باشد روال تشکیل پایگاه داده مجدداً از اول اجرا گردد. با انتخاب این روال، کاربر بایستی دایرکتوری حاوی این تصاویر را برای برنامه مشخص سازد (مشروط بر آنکه در این دایرکتوری فایل‌های تصاویر جدید وجود نداشته باشد و نیز نام فایل‌های این تصاویر بصورتی که در source برنامه مشخص گردیده، معرف کاراکترهای موجود در آنها باشد). پس از آن برنامه، ویژگیهای این کاراکترها را استخراج کرده و آنها را به ویژگیهای قبلی می‌افزاید.

"x": با انتخاب این گزینه، اجرای برنامه خاتمه می‌یابد.

```

-----
                        Selecting a choice
-----
"r","R": Recognition of an input character image
"a","A": Adding an image to the database
"c","C": Constructing the database (for the first time)
"x","X": Quit the program

Please enter your choice: c

360 Images exist in the database

1: picture "F010.bmp" was calculated
2: picture "F011.bmp" was calculated
3: picture "F012.bmp" was calculated
4: picture "F013.bmp" was calculated
5: picture "F014.bmp" was calculated
6: picture "F015.bmp" was calculated
7: picture "F016.bmp" was calculated
8: picture "F017.bmp" was calculated

```

شکل ۴-۷- روال تشکیل پایگاه داده

```

-----
                        Selecting a choice
-----
"r","R": Recognition of an input character image
"a","A": Adding an image to the database
"c","C": Constructing the database (for the first time)
"x","X": Quit the program

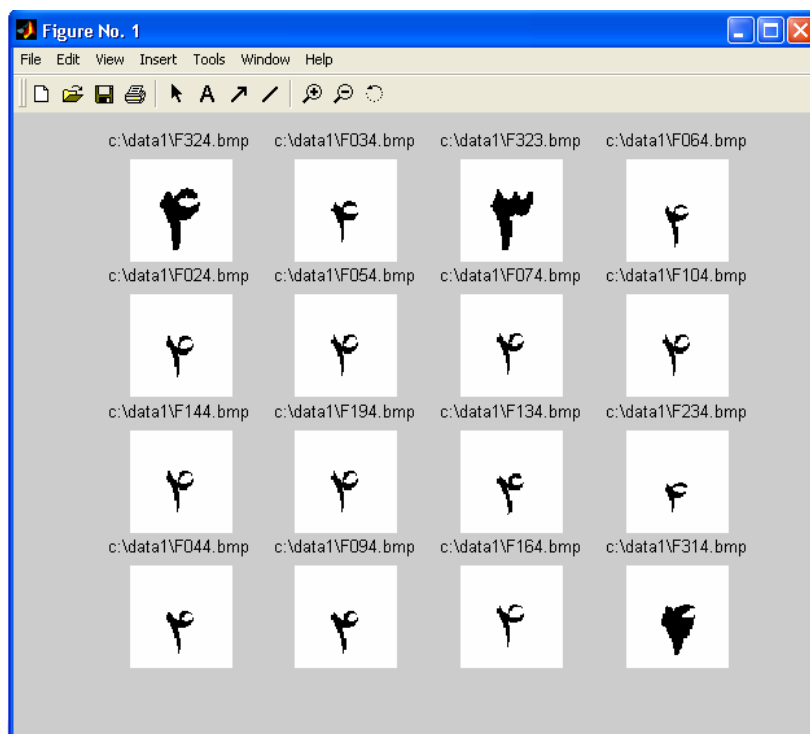
Please enter your choice: q

You have typed character "4"

*****
Search has been finished
Press any key ...
*****

```

شکل ۴-۸- روال تشخیص تصویر کاراکتر ورودی



شکل ۴-۹- تصاویر نمونه بترتیب دارای بیشترین درجه شباهت با تصویر ورودی

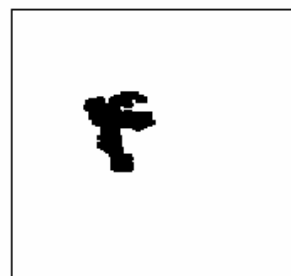
۴-۴- آزمایش برنامه

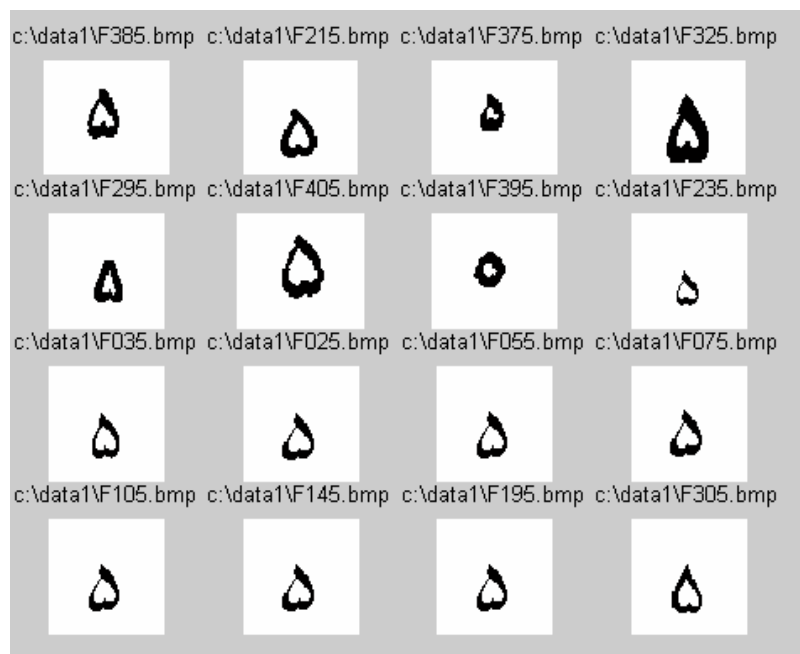
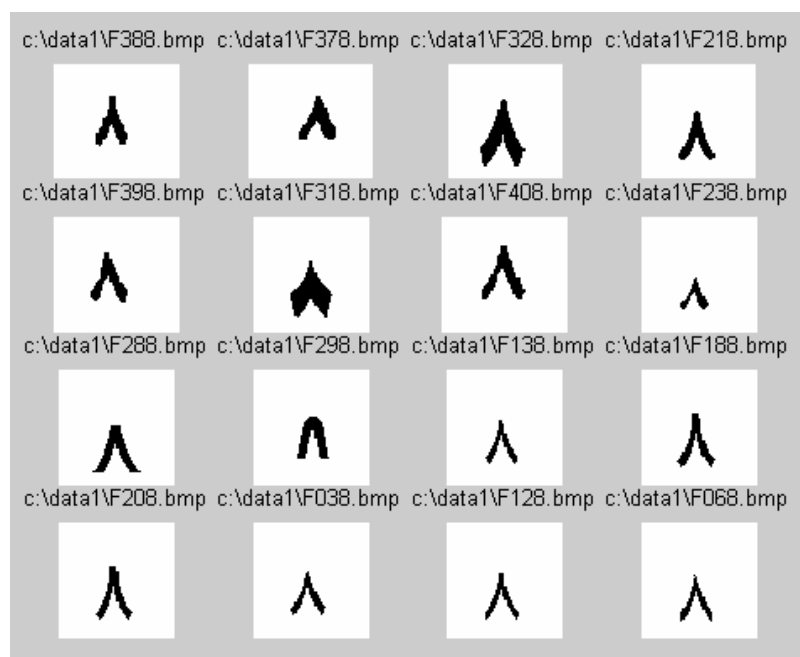
این برنامه در محیط Matlab 6.5 و بر روی یک کامپیوتر Pentium IV - 1800 MHz با سیستم عامل Windows XP تست گردیده است. در این شرایط، زمان استخراج ویژگیها برای هر تصویر نمونه حدود 50msec بوده و پردازش ۱۰۰ تصویر نمونه تنها ۵/۵ ثانیه بطول می‌انجامد که دلالت بر سرعت بسیار بالای برنامه دارد. ورودی این سیستم، ارقام شماره صفحات پایان‌نامه‌های موجود بر روی وب سایت مرکز اطلاعات و مدارک علمی ایران می‌باشد. راندمان این سیستم در حدود ۹۸٪ تعیین گردید.

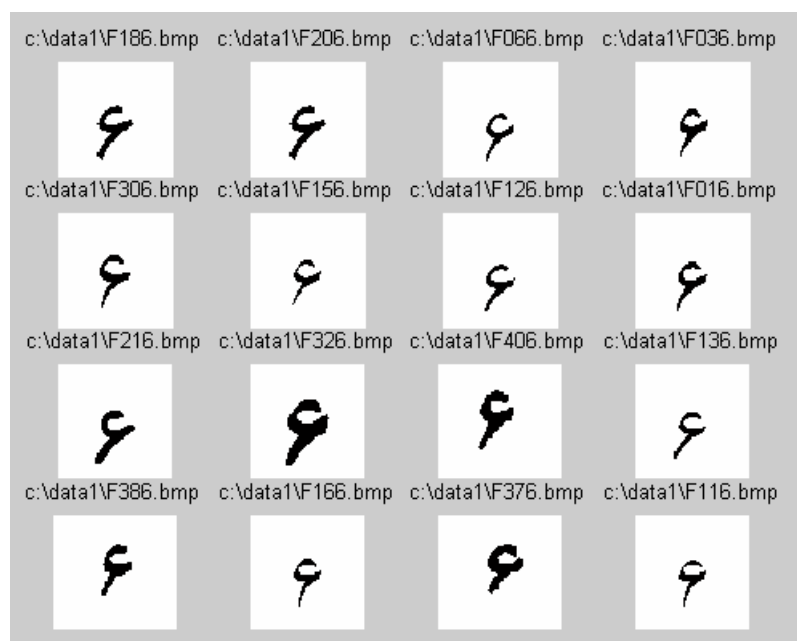
در این سیستم فرض بر آن بوده که ورودی برنامه یکی از ارقام ۱ تا ۹ می‌باشد و رقم صفر به عنوان ورودی سیستم نخواهد آمد. دلیل این فرض آن است که چون الگوریتم در مرحله پیش-پردازش اندازه کاراکترها را نرمالیزه می‌کند و تصویر نرمالیزه ارقام صفر و ۱ تقریباً به یک شکل می‌باشند (زیرا طول و عرض تصویر پس از نرمالیزه شدن یکسان می‌شود)، از اینرو ممکن است این دو کاراکتر اشتباهاً بجای یکدیگر بازنمایی گردند. یک راه حل ساده برای رفع این مشکل آن است که بجای آنکه کاراکترها را به یک اندازه مربعی نرمالیزه کنیم، تنها طول کاراکترها را نرمالیزه نموده و

عرض هر یک را با توجه به مقیاس نرمالیزاسیون تغییر اندازه دهیم. به این ترتیب نسبت طول به عرض کاراکترها قبل و بعد از نرمالیزاسیون ثابت باقی خواهد ماند. البته در مورد شماره صفحات پایان‌نامه‌ها چون رقم صفر هیچگاه به تنهایی به عنوان شماره صفحه انتخاب نمی‌شود و همیشه در کنار دیگر ارقام قرار می‌گیرد، تشخیص آن با توجه به اندازه کوچک آن براحتی میسر است.

شکلهای زیر برخی از نتایج بدست آمده از اجرای این الگوریتم را به نمایش می‌گذارد. در هر گروه، تصویر سمت چپ تصویر ورودی است و تصاویر سمت راست، تصاویر نمونه‌ای هستند که بترتیب بیشترین درجه شباهت را به تصویر ورودی دارند. اولین تصویر به عنوان تصویر خروجی شناخته می‌شود. همانطور که پیشتر نیز عنوان شد، سیستم از نام تصویر خروجی پی به ماهیت آن می‌برد.







فصل پنجم

آنالیز مستندات

در فصل ۱ مبحث آنالیز اسناد (مستندات) معرفی گردید. در این فصل به بررسی دقیقتر این حوزه می‌پردازیم.

۵-۱- مقدمه

همانگونه که شکل (۴-۱) نشان می‌دهد، دو دسته از تکنیکهای پردازشی وجود دارند که به متن موجود در سند اعمال می‌شوند [49]: دسته اول، تکنیکهای بازشناسی نوری حروف (OCR) هستند که کاراکترها و کلمات را از تصویر سند استخراج می‌نمایند (فصول ۱ و ۲). دسته دیگر، تکنیکهای آنالیز پیکربندی صفحه^۱ می‌باشند که فرمت (صفحه‌آرایی) هر بخش از متن را یافته و با توجه به آن، نقش عملکردی (عنوان، زیرعنوان، متن عادی، زیرنویس و غیره) هر یک از این بخشها را معین می‌سازند. این دو عملیات می‌توانند بصورت جداگانه بر روی متن اعمال شوند و یا اینکه نتایج حاصل از یکی، در جهت تصحیح یا کمک به دیگری عمل نماید. همانگونه که در فصل اول عنوان شد، تکنیکهای بازشناسی حروف بر حسب آنکه به متون چاپی^۲ یا دستنویس^۳ اعمال می‌شوند، از یکدیگر متمایز می‌گردند. مشابه همین تقسیم‌بندی نیز برای تکنیکهای پیکربندی صفحه وجود دارد که مطابق آن، این تکنیکها دو دسته زیر را شامل می‌شوند:

- ۱- تکنیکهایی که برای صفحات حاوی متن چاپی و فرمت شده توسعه داده می‌شوند.
- ۲- تکنیکهایی که برای فرمهایی که در آنها متن چاپی یا دستنویس درون بلوکها یا جداول نوشته می‌شوند، در نظر گرفته شده‌اند. فرمها ممکن است از پیش چاپ شده بوده و متنها بعداً در درون آنها چاپ یا نوشته شده باشند و یا اینکه فرم و متن آن با هم توسط یک نرم‌افزار (نظیر MS Word) ایجاد گردیده باشند. از آنجائیکه بیشتر تکنیکهای آنالیز پیکربندی صفحات نسبت به کجی متن سند حساس می‌باشند (مثل تکنیکهای OCR)، لازم است که قبل از شروع این مرحله، زاویه کجی متن شناسایی و تصحیح گردد.

از جنبه دیگر، تکنیکهای آنالیز پیکربندی صفحه به دو دسته زیر تقسیم می‌شوند:

^۱ Page Layout Analysis

^۲ Machine Printed

^۳ Handwritten

۱- تکنیکهای آنالیز پیکربندی ساختاری^۱ (یا فیزیکی یا هندسی).

۲- تکنیکهای آنالیز پیکربندی عملکردی^۲ (یا نحوی یا منطقی).

آنالیز پیکربندی ساختاری بمنظور قطعه‌بندی متن سند به گروههای مختلف مورد استفاده قرار می‌گیرد که بسته به فرمت سند این گروهها می‌توانند کلمات مجزا، خطوط متن یا گروههای ساختاری (گروههای خطوط متن نظیر پاراگرافهای مجزا یا هر یک از آیتمهای بخش فهرست مطالب) باشند. آنالیز پیکربندی عملکردی، از اطلاعات دامنه-محدود^۳ شامل قوانین پیکربندی یک صفحه خاص استفاده می‌کنند تا بلوکهای ساختاری سند را برچسب‌گذاری^۴ نموده و بدین ترتیب نوعی نشانه-گذاری^۵ از عملکرد بلوکها انجام می‌دهند. این نوع نشانه‌گذاری عملکردی ممکن است مستلزم بکارگیری تکنیکهای تقسیم^۶ یا ادغام^۷ بلوکهای عملکردی باشند. شکل (۵-۱) مثالی از نتیجه حاصل از اعمال آنالیزهای پیکربندی ساختاری و عملکردی را برای صفحه اول یک مقاله فنی نشان می‌دهد [28]. در این مثال، تکنیک برچسب‌گذاری عملکردی، بخشهای مختلف اعم از بلوکهای عنوان، نام نویسندگان، چکیده، کلمات کلیدی، پاراگرافهای متن عادی و نظایر آنها را مشخص نموده است. شکل (۵-۲) نیز سلسله‌مراتب ساختار منطقی یک تصویر سند را به نمایش می‌گذارد [58].

آنالیز پیکربندی ساختاری می‌تواند بصورت الگوریتمهای بالا به پایین^۸ یا پایین به بالا^۹ پیاده-سازی گردد [28، 49 و ۷۰]. در آنالیز بالا به پایین، صفحه از اجزاء بزرگ به زیرجذءهای کوچکتر قطعه‌بندی می‌شود. به عنوان مثال صفحه ممکن است به یک یا چند بلوک ستونی متنی تقسیم گردد؛ سپس هر ستون نیز امکان دارد که به بلوکهای پاراگرافی تقسیم‌بندی شود و هر پاراگراف نیز به نوبه خود به خطوط متن مجزا گردد و این روند ادامه یابد. در مورد آنالیز پایین به بالا، اجزاء پیوسته

¹ Structural Layout Analysis

² Functional Layout Analysis

³ Domain Dependant

⁴ Labeling

⁵ Indication

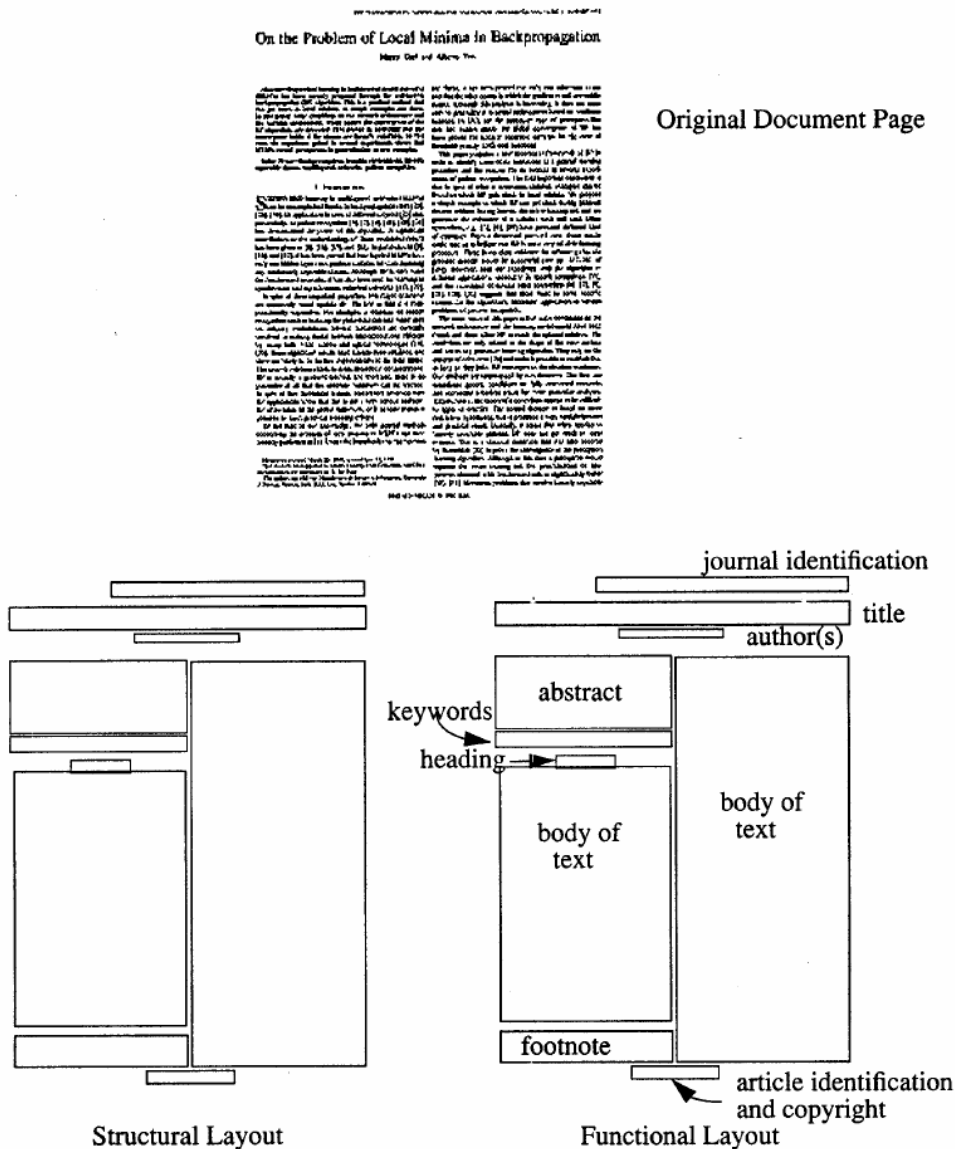
⁶ Splitting

⁷ Merging

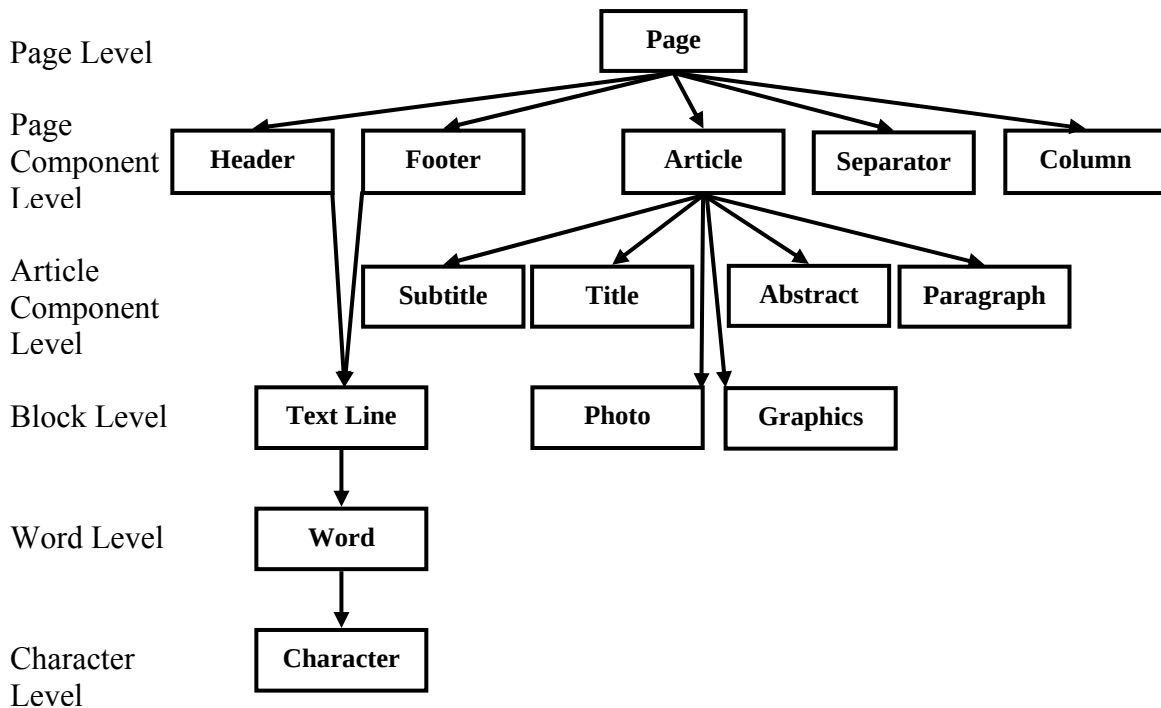
⁸ Top-Down

⁹ Bottom-Up

بترتیب به کاراکترها، کلمات، خطوط متن و اجزاء بزرگتر گسترش می‌یابند. در ادامه برخی از این تکنیکها معرفی می‌گردند.



شکل ۵-۱- صفحه اول یک مقاله فنی به همراه نتایج حاصل از پیکربندی ساختاری و عملکردی. آنالیز پیکربندی ساختاری بلوکهای را نشان می‌دهد که بر پایه فواصل موجود در تصویر اصلی مقاله قطعه‌بندی گردیده‌اند. عملیات برچسب‌گذاری در مرحله آنالیز پیکربندی عملکردی، از اطلاعات قبلی راجع به قوانین صفحه‌بندی این مقاله ژورنال (مقاله مزبور، مربوط به یکی از کنفرانسهای IEEE می‌باشد) منتج گردیده‌اند.

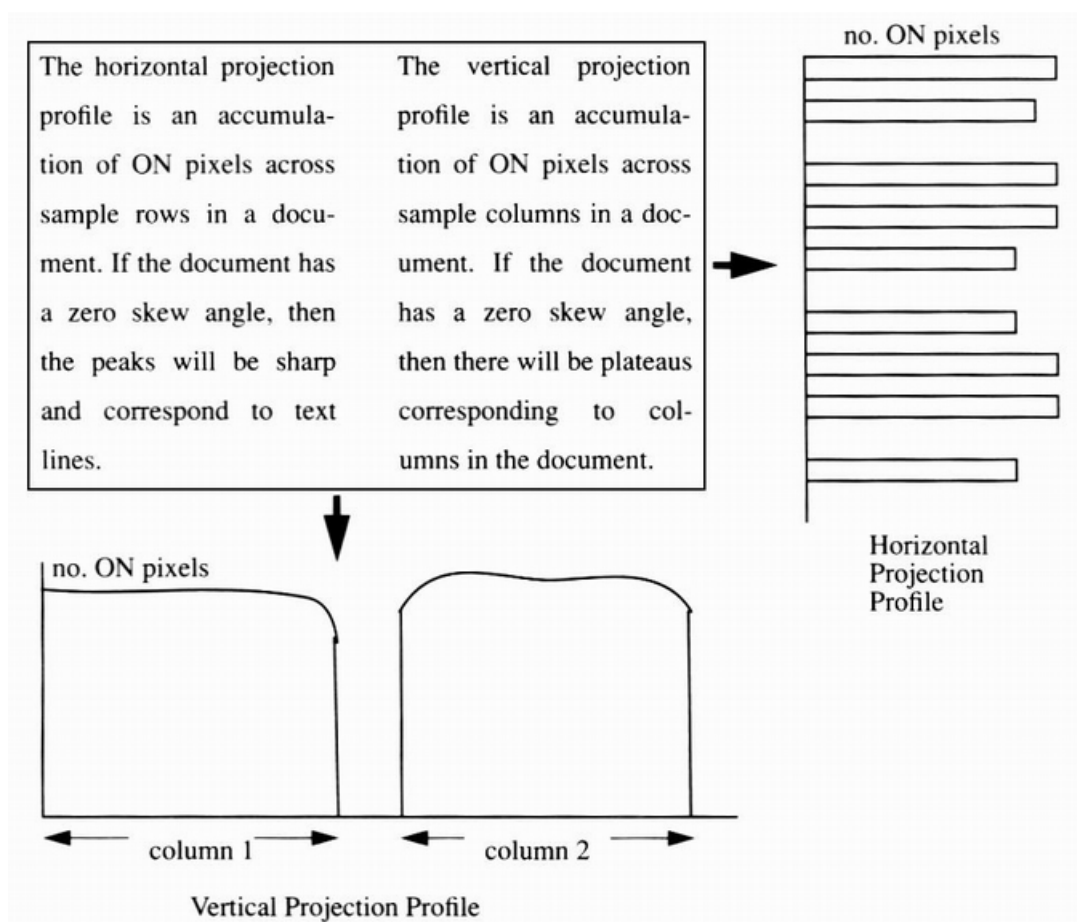


شکل ۵-۲- سلسله مراتب ساختار منطقی یک تصویر سندی

۵-۲- آنالیز بالا به پایین

هیستوگرامها (پروفایلهای تصویرنمایی)^۱ از جمله تکنیکهای پرستفاده در سیستمهای آنالیز پیکربندی اسناد می باشند. همانگونه که شکل (۵-۳) نیز نشان می دهد، بوسیله هیستوگرام عمودی می توان محل ستونها را تعیین نمود که بشکل فلاتهای قرار گرفته مابین دره ها هستند. به همین ترتیب هیستوگرام افقی، خطوط متن و پاراگرافها را در هر ستون از یکدیگر متمایز می سازد (پاراگرافها با این روش به شرطی از هم جدا می شوند که فضای خالی بین آنها، بیش از فاصله بین سطرهاى متن عادى باشد). علاوه بر این، نقش کاربردی هر یک از بلوکهای متن را می توان با این روش تعیین نمود. به عنوان مثال هیستوگرام افقی یک متن عنوان، اغلب از پهنای بیشتری نسبت به هیستوگرام متن عادى صفحه برخوردار است و یا ممکن است یک متن زیرنویس، پهنای کمتری نسبت به یک متن عادى داشته باشد. در این روش بالا به پایین، اطلاعات زیادی راجع به پیکربندی صفحه قابل استخراج است [49].

^۱ Projection Profiles



شکل ۵-۳- قسمتی از یک متن دو ستونه (قسمت بالا و سمت چپ). هیستوگرامهای افقی و عمودی تصویر متن نشان داده شده‌اند. در هیستوگرام افقی، هر میله نمایانگر یک خط متن، و در هیستوگرام عمودی، هر فلات نشان‌دهنده یک ستون می‌باشد.

معمولترین روشهای آنالیز بالا به پایین عبارتند از [49 و ۷۰]:

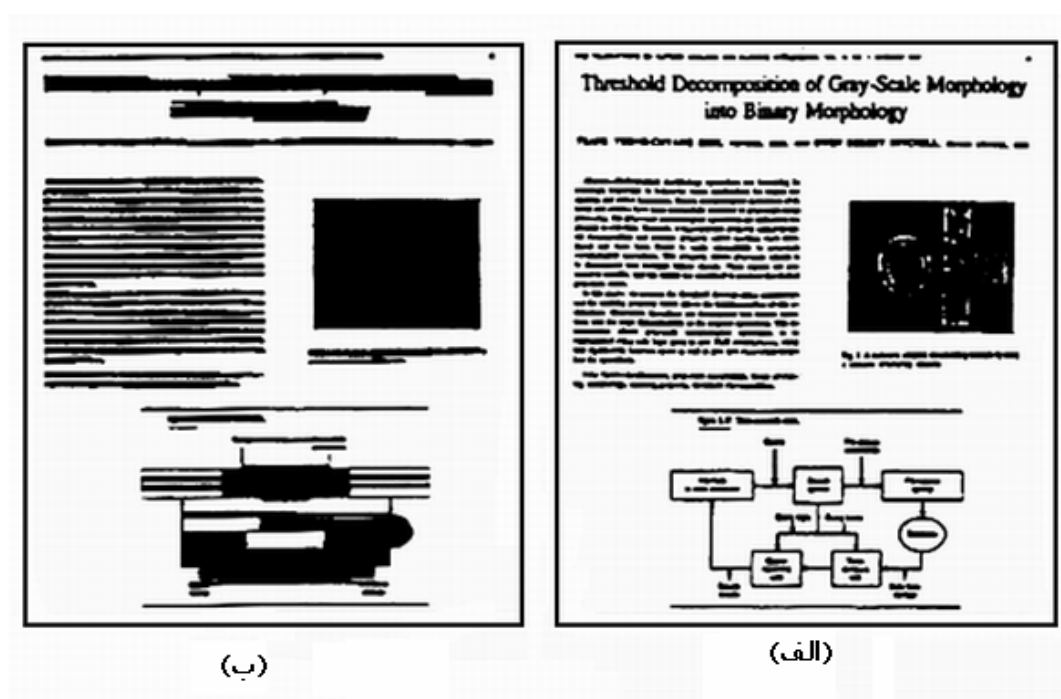
- روش آغشته‌سازی^۱.
- روش کاهش رزولوشن (درجه تفکیک) تصویر.
- روش برشهای جانبی تصویر (قطعه‌بندی و برچسب‌زنی).
- روش استفاده از فضاهای سفید صفحه.
- روش مبتنی بر تبدیل هاف.

در ادامه این روشها تشریح می‌شوند.

^۱ Run-length Smearing Algorithm

۵-۲-۱- روش آغشته‌سازی

در صورت استفاده از روش مبتنی بر هیستوگرام که در بالا عنوان شد، برای آنالیز پیکربندی، نیازی نیست که عملیات پردازش را در رزولوشن اصلی بر روی کلیه پیکسل‌های تصویر سند انجام دهیم؛ بلکه بجای آن می‌توان درجه تفکیک تصویر را کاهش داد. یک روش برای انجام این کار، هموارسازی^۱ کاراکترها به اجزاء لکه‌مانند^۲ کوچکتر و غیر قابل تشخیص (به عنوان کاراکتر) می‌باشد. بسته به میزان هموارسازی، این اجزاء می‌توانند یک کاراکتر منفرد، یک کلمه یا یک خط از متن باشند. یک روش معمول برای انجام این نوع هموارسازی، الگوریتم آغشته‌سازی است [59]. این روش حروف را به کلمات، کلمات را به خطوط متن و (گاهی) خطوط متن را به پاراگرافها یکپارچه^۳ می‌نماید. در عمل آغشته‌سازی، فضاهاى سفید بین پیکسل‌های سیاه واقع بر یک امتداد (افقی، عمودی یا هر دو) مورد بررسی قرار می‌گیرند و فضاهاىی که طول آنها کمتر از یک مقدار آستانه باشد، سیاه می‌گردند (شکل ۵-۴).



شکل ۵-۴- هموارسازی تصویر به کمک روش آغشته‌سازی الف- تصویر اصلی ب-

تصویر آغشته‌سازی شده [۷۰]

¹ Smoothing

² Blobs

³ Merge

چنانچه آغشته‌سازی در سطح حروف کلمات (یعنی تنها حروف -و نه کلمات- سیاه شوند) باشد، بایستی مقدار سطح آستانه را بزرگتر از فاصله خالی بین حروف کلمات و کوچکتر از فاصله خالی بین کلمات متن انتخاب نمود. همچنین بمنظور یکپارچه‌سازی واحدهای بزرگتر (نظیر کلمات، ... مقدار سطح آستانه بایستی به تناسب انتخاب گردد. با بکارگیری این روش علاوه بر آنکه تعداد پیکسلها کاهش می‌یابد، در مواردی همچون متنهای تُنک که هیستوگرام آنها نامناسب و حالت نویزی شکل دارد، می‌تواند نتایج بهتری به بار آورد. این روش مستلزم آن است که زاویه کجی تصویر از ابتدا تصحیح شده باشد و فضاهاى خالی بین کلمات مشخص و یکنواخت باشد [49].

۵-۲-۲- روش کاهش رزولوشن (درجه تفکیک) تصویر

در این روش فرض بر آن است که بلوکهای هم‌نوع دارای سطح خاکستری متوسط یکسانی هستند؛ مثلاً می‌توان فرض نمود که پاراگرافهای مختلف یک متن، بلوکهایی هستند که سطح خاکستری متوسط آنها به هم نزدیک است و مقدار آن با مقدار سطح خاکستری فضای خالی بین دو پاراگراف و یا با سطح خاکستری متوسط در قسمت گرافیکی متفاوت می‌باشد. در این روش با نگاشت چند نقطه مجاور هم در یک بلوک از تصویر اصلی به یک نقطه با ابعاد هندسی بزرگتر، تصویر جدیدی با درجه تفکیک کمتر ایجاد می‌شود. به عنوان مثال در یک مقاله با استفاده از یک پنجره 8×24 نقطه به یک نقطه تبدیل می‌شود. سطح خاکستری نقاط بدست آمده در تصویر جدید، میانگین سطح خاکستری نقاط متناظر در تصویر اصلی است. با انتخاب مقادیر مناسب سطح آستانه برای سطوح خاکستری، بخشهای مختلف از هم جدا می‌گردند. این عملیات در واقع نوعی آغشته‌سازی تصویر است. در تصویر جدید، اجزاء کوچکتر به یکدیگر متصل شده، اجزاء بزرگتری را بوجود می‌آورند [۷۰].

۵-۲-۳- روش برشهای جانبی تصویر (قطعه‌بندی و برچسب‌زنی)

در این روش آنالیز بالا به پایین که ساخت‌یافته‌تر بوده و مانند روشهای قبلی مبتنی بر بکارگیری هیستوگرامها می‌باشد، تصویر سند با برشهای یک در میان افقی و عمودی در امتداد

یک مزیت این روش آن است که ترکیب پردازش ساختاری و برچسب‌زنی عملکردی سبب می‌شود که روند کار هدایت شده و تصحیح شده باشد. البته این روش نیازمند دانستن ساختار اولیه بلوکها است که این نیز به میزان آگاهی از ویژگیهای بلوکها (مانند اندازه فونتها و فضای بین خطوط) بستگی دارد. همچنین بایستی توجه داشت که در این روش نیز همانند روش آغشته‌سازی، کج نبودن متن نقشی بسیار حیاتی دارد و بعلاوه در ابتدا می‌بایست نویز فلغل-نمکی^۱ (نویزی که بصورت نقطه-های ریز سیاه و سفید در تصویر دیده می‌شود) حذف گردد. این شروط، مورد نیاز بیشتری روشهای قطعه‌بندی می‌باشند.

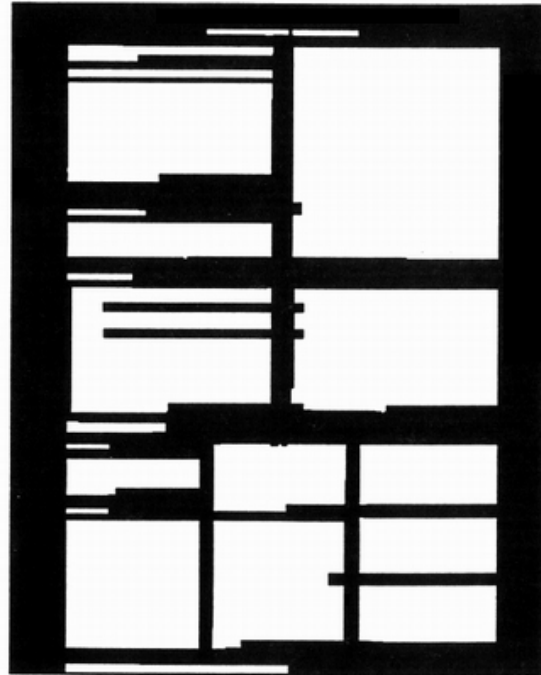
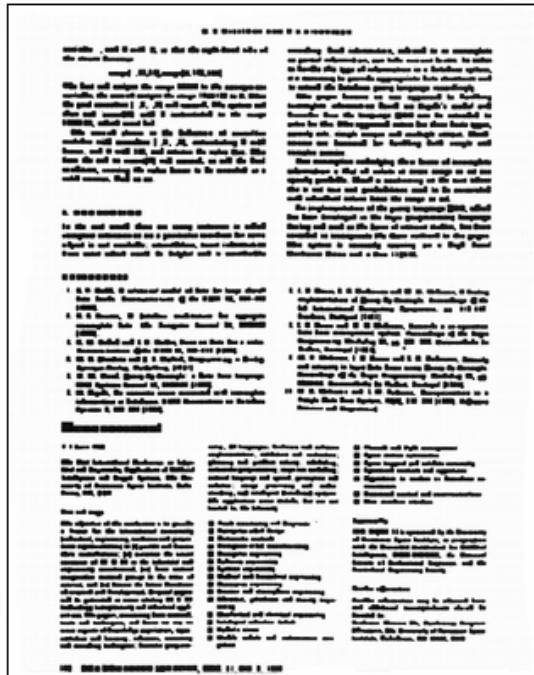
۵-۲-۴- روش استفاده از فضاهای سفید صفحه

در این روش، فضاهای خالی در صفحه (که منظور، فضای پس‌زمینه در مقابل متن پیش‌زمینه می‌باشد) بمنظور جداسازی بلوکها مورد پردازش قرار گرفته و سپس از هیستوگرامها بمنظور یافتن خطوط کمک گرفته می‌شود [9] (شکل ۵-۶). در ابتدا کلیه مستطیلهای سفید -که بصورت محلی بیشینه می‌باشند و در میان اجزاء متصل سیاه قرار گرفته‌اند- از بزرگ به کوچک فهرست می‌شوند که بزرگترین آنها، پوشش پس‌زمینه سفید روی صفحه را تشکیل می‌دهد. بدین ترتیب متن پیش‌زمینه به بلوکهای ساختاری تقسیم‌بندی می‌گردد. اندازه مستطیلهای ممکن است به نحوی انتخاب شوند که قطعه‌بندی تا سطوح پایینتر (نظیر ستونها، پاراگرافها یا حتی کاراکترهای مجزا) نیز انجام شود. صفحه می‌بایست مطابق با اصول صفحه‌بندی Manhattan نگاشته شده باشد که بر طبق آن کل صفحه می‌بایست تنها یک زاویه کجی داشته و نیز توسط برشهای افقی و عمودی، قابل تفکیک به بلوکهای مجزا باشد.

یک مزیت بکارگیری فضاهای سفید پس‌زمینه در مقابل متن پیش‌زمینه برای پیکربندی صفحه، مستقل بودن آن نسبت به نوع زبان متن است؛ زیرا در بیشتر زبانها از فضاهای خالی به نحو مشابهی در صفحه‌بندی استفاده می‌شود. مزیت دیگر آن است که پارامترهای کمتری لازم است تا

^۱ Salt-and-Pepper Noise

تعیین گردند. یک نقطه ضعف این روش، انتخاب مستطیل بیشینه است (مثلاً طولانیترین یا بزرگترین ناحیه خالی) که ممکن است بصورت شهودی قابل تعیین نباشد یا حتی نحوه تعیین آن برای سندهای دارای فرمتهای مختلف متفاوت باشد [49].



شکل ۵-۶- آنالیز پیکربندی صفحه بوسیله استخراج بلوکهای مستطیلی از فضاهای خالی

صفحه

۵-۲-۵- روش مبتنی بر تبدیل هاف

تبدیل هاف، هر خط در صفحه $X-Y$ را به یک نقطه (یا مجموعه خطوطی که یکدیگر را در یک نقطه مشترک قطع می کنند) در صفحه $\rho-\theta$ می نگارد. نحوه عمل به این صورت است که به ازای هر نقطه (x_i, y_i) در صفحه $X-Y$ ، منحنی سینوسی $\rho = x_i \cos(\theta) + y_i \sin(\theta)$ در صفحه $\rho-\theta$ بدست می آید. حال اگر تبدیل یافته نقاط مربوط به خط راست $Y = ax + b$ را بدست آوریم، دسته منحنیهای حاصل شده همگی از یک نقطه مشترک (ρ_i, θ_i) عبور می نمایند. بدین ترتیب توسط تبدیل هاف می توان یک خط راست را به یک نقطه نگاشت و نقاطی را که در امتداد یک خط راست هستند (اگر چه ناپیوسته باشند) آشکار ساخت.

تصویر یک متن شامل انواع مختلف خطوط نظیر خطوط متن، خطهای راست، جداول و فرمولها می باشد که با استفاده از این تبدیل می توان آنها را آشکار نمود و بر اساس اطلاعات بدست آمده، عملیات جداسازی بلوکها را انجام داد. همچنین بواسطه آنکه تبدیل یافته بلوکهای مختلف (متن، گرافیک، تصویر) دارای ویژگیهای متفاوتی است، می توان علاوه بر تشخیص و جداسازی بلوکها، از تبدیل هاف بمنظور طبقه بندی بلوکها نیز استفاده نمود [۷۰].

۵-۳- آنالیز پایین به بالا

یک مزیت اساسی روشهای آنالیز بالا به پایین آن است که به جهت بکارگیری ساختار کلی صفحات، سرعت می توانند عمل پیکربندی صفحه را انجام دهند. این روشها برای بیشتر سندها کارایی مناسبی دارند. اما برای صفحاتی که متن آنها مرزهای خطی نداشته باشد و یا شکلها هم در داخل و هم در بیرون متن بتوانند قرار گیرند، ممکن است نامناسب عمل کنند. به عنوان مثال بسیاری از مجلات، متنی را در پیرامون یک شکل بنحوی چاپ می کنند که بجای یک خط راست، بصورت یک شیء خاص دربیاید. همچنین بخش فهرست مطالب در مجلات و ژورنالهای علمی اغلب با شکلهایی که در داخلشان ضمیمه شده اند، همراه هستند و یا فرمت متن در آنها بصورت وسط چین (و نه چپ چین یا راست چین) یا هر فرمت مخالف با فرمت Manhattan می باشد. در این قبیل موارد، روشهای آنالیز پیکربندی پایین به بالا که در این بخش توضیح داده می شوند عملکرد بهتری دارند، اما در مقابل بار محاسباتی آنها سنگینتر است [49].

در بخش قبلی عنوان شد که روشهای آنالیز پیکربندی پایین به بالا با شروع از اجزاء کوچک صفحه، آنها را بمنظور تولید اجزاء بزرگتر بصورت پیایی در هم ادغام می نمایند تا زمانی که کلیه بلوکهای واقع در صفحه بازیافت گردند. در این بخش، تعدادی از روشهایی که در این دسته قرار می گیرند، اما از روشهای میانی کاملاً متفاوتی برای این منظور استفاده می کنند را مورد بررسی قرار می دهیم. همچنین در این بخش ایده ای از سیستمهای نرم افزاری کامل برای آنالیز پیکربندی صفحات ارائه می گردد.

Fisher و همکارانش، تعدادی از تکنیکهای معرفی شده در بخش قبل را مورد استفاده قرار داده‌اند [49]. ابتدا زاویه کجی صفحه توسط تبدیل هاف مشخص شده، سپس فاصله بین خطوط متن^۱ بوسیله مقدار پیک تبدیل فوری^۲ یک‌بعدی هیستوگرام در جهت زاویه کجی تشخیص داده شده محاسبه می‌گردد. پس از انجام هموارسازی، هیستوگرام مربوط به فواصل درون خطی^۳ (هیستوگرام فواصل بین کلمات و بین کاراکترها) تعیین می‌شود که پیک این هیستوگرام، به عنوان فاصله درون-خطی منظور خواهد شد. بعد از آن با اعمال پیاپی تکنیکهای هموارسازی در جهت زاویه کجی کلمات و خطوط متن و عمود بر جهت کجی پاراگرافها و ستونها، اجزاء متنی بمنظور تولید واحدهای بزرگتر در هم ادغام می‌شوند. نتیجه این عملیات، یکسری نقاط "روشن"^۴ است که تکنیکهای آنالیز اجزاء متصل^۵ به آنها اعمال می‌گردد. این تکنیکها شامل تعدادی عملیات آماری (نظیر تعیین محدوده ارتفاع، مساحت و طول کلمات) هستند که بمنظور جداسازی بلوکهای متن و ایجاد تمایز بین متن و زمینه بکار گرفته می‌شوند. Esposito و همکارانش [19] نیز روش مشابهی را بکار گرفته‌اند؛ متنی بمنظور کاهش محاسبات، آنها ابتدا کادر محصورکننده^۵ هر کاراکتر را یافته و سپس محاسبات را بجای پیکسلها بر روی این کادرها انجام می‌دهند.

روش Docstrum [48] تکنیک کلاستر بندی پایین به بالای k-نزدیکترین همسایه را بمنظور دسته‌بندی کاراکترها به خطوط متن و بلوکهای ساختاری بکار می‌گیرد (شکل ۵-۷). در این روش ابتدا برای هر جزء سند، اتصالهای k-نزدیکترین همسایه به اجزاء مجاور یافت می‌شود (که در اینجا k معمولاً ۴ یا ۵ است). طولها و زوایای این اتصالها بصورت هیستوگرامهایی تنظیم می‌گردند. از آنجا که بیشتر این اتصالات بین کاراکترهای واقع بر روی یک خط متن برقرار می‌شوند، زاویه پیک هیستوگرامها، به عنوان زاویه کجی و مقدار پیک آنها، به عنوان فاصله بین کاراکتری منظور می‌گردد. خطوط متن توسط این تخمینها، به عنوان گروههایی از کاراکترها و کلمات در جهت زاویه صفحه

¹ Between-line Spacing

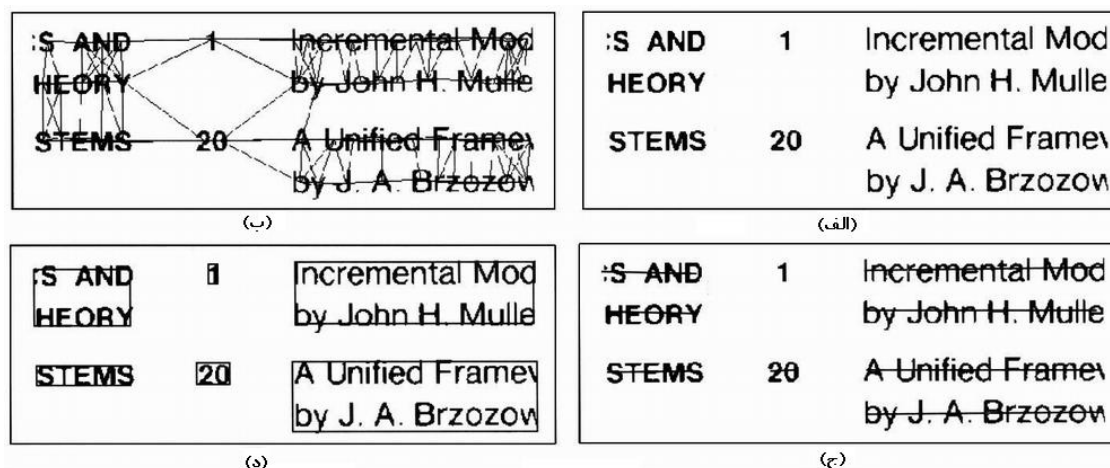
² Within-line Spacing

³ ON

⁴ Connected Component Analysis

⁵ Bounding Box

بازیافت می‌شوند. سپس با توجه به این ویژگی سند که خطوط متن مربوط به یک بلوک نسبت به بلوکهای مجاور معمولاً فاصله کمتری نسبت به یکدیگر دارند، خطوط متن دسته‌بندی خواهند شد.



شکل ۵-۷- روش آنالیز پیکربندی پایین به بالای Docstrum

الف- قسمتی از یک صفحه فهرست مطالب.

ب- نزدیکترین همسایه‌های هر یک از اجزاء سند مشخص می‌شوند.

ج- نزدیکترین همسایه‌ها نشان‌دهنده خطوط متن می‌باشند.

د- بلوکهای ساختاری به عنوان گروههای متشکل از خطوط متن بازیافت می‌گردند.

مشابه روش آنالیز بالا به پایین Nagy [45] که اطلاعات برچسب‌زنی عملکردی را همزمان با انجام آنالیز پیکربندی ساختاری صفحه مورد استفاده قرار می‌دهد، می‌توان در آنالیز پایین به بالا نیز این دو پروسه را با یکدیگر بکار گرفت. Hagita و Akiyama با استفاده از جداکننده‌های ناحیه‌ای^۱ (خطوط بین نواحی متنی) عمل قطعه‌بندی انجام می‌گیرد و سپس فضاهای خالی مشابه با سایر روشها مرزبندی می‌گردند. پس از تعیین ویژگیهای متنی و سراسری، بلوکها بازیافت گردیده و توسط ویژگیهای عمومی مستندات طبقه‌بندی می‌شوند. ویژگیهای متنی بکار رفته در اینجا، از هیستوگرامهای افقی و عمودی، تعداد گذرهای عمودی^۲ (تعداد تغییرات سیاه به سفید و سفید به سیاه در امتداد یک خط) و کادرهای محصورکننده هر کاراکتر بدست می‌آیند [49]. برخی از محققان بر این باورند که انجام پردازشهای اولیه‌ای که تا حد امکان عاری از اطلاعات مختص یک کاربرد بخصوص باشند، یک

¹ Field Separators

² Vertical Crossing Counts

مزیت به حساب می‌آید؛ چرا که این شیوه سبب ترویج تکنیکهای قدرتمند می‌گردد. از طرف دیگر، برخی نیز عقیده دارند که چنانچه اطلاعات مفیدی مختص کاربرد مورد نظر موجود و قابل استفاده باشد، می‌بایست در جهت تسهیل حل مسئله بکار گرفته شود [49]. در بخش زیر مبحث برچسب‌زنی عملکردی، مستقل از آنالیز پیکربندی ساختاری (به عنوان یک پروسه مستقل بعد از آن) مورد بررسی قرار می‌گیرد.

۴-۵- برچسب‌زنی عملکردی

همانطور که قبلاً عنوان شد، برخی از تکنیکهای آنالیز پیکربندی (مانند [38]) در حین انجام بلوک‌بندی ساختاری، برچسب‌زنی عملکردی را انجام می‌دهند. سایر روشها، این دو مرحله را بصورت پشت سر هم پردازش می‌نمایند. در هر دو حالت، برچسب‌زنی عملکردی بر مبنای قوانین خاص هر سند و با بکارگیری ویژگیهای استخراج شده از سند انجام می‌گیرد. این ویژگیها ممکن است شامل موقعیتهای نسبی و مطلق یک بلوک در صفحه، موقعیتهای نسبی و مطلق یک ناحیه در یک بلوک، قواعد عمومی فاصله‌گذاری هنگام حروف‌چینی، و اندازه فونت و شیوه نگارش متن می‌باشد. قواعد فرمت‌بندی که در جریان تولید سند بکار می‌روند، در فرایند معکوس برچسب‌زنی عملکردی مورد استفاده قرار می‌گیرند. از جمله مشکلات موجود در این زمینه، تغییرات وسیع شیوه نگارش و فرمت‌بندی در اسناد مختلف است.

بواسطه تغییرات زیاد فرمت اسناد و به جهت مشکلات عام موجود در فرایند پردازش تصاویر (مانند نویز، آشکارسازی غلط و ویژگیها و نتایج ناقص بازشناسی متون)، محدوده عملکرد اغلب تحقیقات در حوزه برچسب‌زنی عملکردی، تنها به زمینه‌های خاصی منحصر می‌شود. در حوزه شناسایی اقلام پستی، اکثر فعالیتها توجه خود را به مسئله مکان‌یابی بلوک آدرسها و فیلدهای آدرس معطوف نموده‌اند [49]. برای پست آمریکا، هدف بیشتر یافتن محل بلوک آدرسها، سپس کد Zip در خط آخر و پس از آن سایر قسمتهای بلوک آدرس می‌باشد. این مثال خوبی است که نشان می‌دهد بلوکها در مکانهای مختلف و فیلدها در خطوط متفاوت، بار معنایی متمایزی دارند. همچنین تحقیقاتی در زمینه بازشناسی فرمها صورت گرفته است که در آنها مسئله بلوک‌بندی ساختاری بوسیله خطوط

The diagram illustrates the transformation of a specific layout structure into a specific logical structure for a letter template.

Specific Layout Structure: This part shows a visual representation of a letter template. It includes a header section, a main body section, and a footer section. The layout is defined by a set of boxes and lines, with a large arrow pointing from the header to the body, and another arrow pointing from the body to the footer.

Specific Logical Structure: This part shows the logical structure of the letter template. It is represented as a tree diagram where the root node is "Letter". The "Letter" node has two children: "Sender" and "Logo". The "Sender" node has two children: "Subject" and "Date". The "Logo" node has two children: "Recipient" and "Body". The "Recipient" node has two children: "Body" and "Footer". The "Body" node has two children: "Body" and "Footer".

¹ Office Document Architecture

از آنجا که پروسه‌های آنالیز پیکربندی ساختاری و عملکردی، معکوس پروسه فرمت‌بندی اسناد می‌باشند، منطقی است که بدنبال استاندارد برای ایجاد و توصیف محتوای اسناد باشیم که بتوان آن را بمنظور انجام آنالیز پیکربندی و سنجش نتایج استفاده نمود. دو استاندارد شناخته شده در این زمینه، ODA و SGML^۱ می‌باشند [49]. گرچه هر دوی این استانداردها به توصیف اسناد می‌پردازند، اما در نحوه توصیف‌نمایی کاملاً متفاوت عمل می‌کنند.

ODA یک استاندارد تعیین ساختار سندی شیء‌گراست که هم اجزاء ساختاری و هم اجزاء عملکردی یک سند را توصیف می‌کند. این توصیف‌های ODA سلسله‌مراتبی هستند. بنابراین یک نحوه تعریف صفحه‌بندی (پیکربندی) ممکن به این صورت باشد که یک سند حاوی صفحاتی است که در آن بخشهایی نظیر ستونها، پاراگرافها و نظایر اینها موجود می‌باشد. در این استاندارد، سلسله‌مراتب عملکردی به توصیف سندی می‌پردازد که می‌تواند شامل بخشها یا فصلهایی باشد و هر فصل نیز قسمتهایی چون عناوین، زیرعناوین و غیره را در بر بگیرد. هدف اولیه ODA، ایجاد یک استاندارد مشترک جهت تبادل اسناد اداری بود.

SGML به معنی "زبان تفسیر عمومیت‌یافته استاندارد"، زبانی است برای تفسیر متن بوسیله اطلاعاتی که متن را به عنوان اجزاء عملکردی توصیف می‌نماید. این استاندارد برای اسناد منتشر شونده پدید آمده است. در فرایند نشر، این اجزاء عملکردی با یکدیگر و همراه با قواعد مختص نشر برای فرمت‌بندی سند در یک شیوه نگارش منتخب بکار گرفته می‌شوند. در مقایسه با ODA، SGML را بیشتر می‌توان به عنوان یک توصیف پایین به بالا از سند در نظر گرفت. هر دوی این استانداردها مطابق با اهدافی که از ابتدا برای آنها در نظر گرفته شده بود، توسعه یافته‌اند. مثلاً هر دو برای آنالیز پیکربندی اسناد بکار می‌روند. با این وجود، در حالی که هر دو برای نشر و آنالیز اسناد مورد استفاده قرار می‌گیرند، اما به هیچ وجه بصورت انحصاری بکار نمی‌روند. بسیاری از مراکز نشر از سیستمهایی مختص خود استفاده می‌کنند که با هیچ یک از استانداردهای ODA و SGML سازگار نمی‌باشند و

^۱ Standard generalized Markup Language

اغلب تحقیقات کنونی در زمینه بازشناسی اسناد هنوز نتایجی را که سازگار با استاندارد خاصی باشد، تولید نمی نمایند [49].

فصل ششم

**امكان سنجی پروژه
و نتیجه گیری**

در فصول ۱ و ۲، سیستمهای بازشناسی حروف مورد بررسی قرار گرفتند و همچنین یک سیستم آزمایشی بازشناسی ارقام فارسی، در فصل ۴ معرفی و پیاده‌سازی گردید. در این سیستم، ورودی برنامه فایلی است که تنها حاوی یکی از ارقام ۱ تا ۹ فارسی می‌باشد و در نتیجه مسئله یافتن مکان کاراکتر به آسانی انجام می‌شود. آزمایش این سیستم نشان داد که در صورتیکه بتوان مکان درج ارقام شماره صفحات پایان‌نامه‌ها را بدرستی شناسایی نمود، می‌توان با دقت بالایی این ارقام را تشخیص داد. بنابراین مسئله اصلی، یکی شناسایی صفحه فهرست مطالب و دیگری تشخیص مکان درج شماره صفحات در این صفحه و صفحات اصلی است. این مسائل به حوزه آنالیز اسناد مربوط می‌شوند که در فصل پنجم بررسی گردیدند. در این فصل به مسئله اصلی یعنی امکان‌سنجی یافتن صفحات TOC و MPها در پایان‌نامه‌ها جهت برقراری لینک بین آنها می‌پردازیم.

۱-۶- بلوک دیاگرام سیستم پیشنهادی

شکل (۱-۶) بلوک دیاگرام سیستم پیشنهادی برای این منظور را نشان می‌دهد که البته ممکن است در آینده بلوکهای دیگری به آن اضافه گردد). علائم اختصاری بکار رفته در این شکل عبارتند از:

TOC: صفحه فهرست مطالب.

MP^۱: صفحه اصلی.

PN^۲: شماره صفحه.

OCR^۳: بازشناسی نوری حروف.

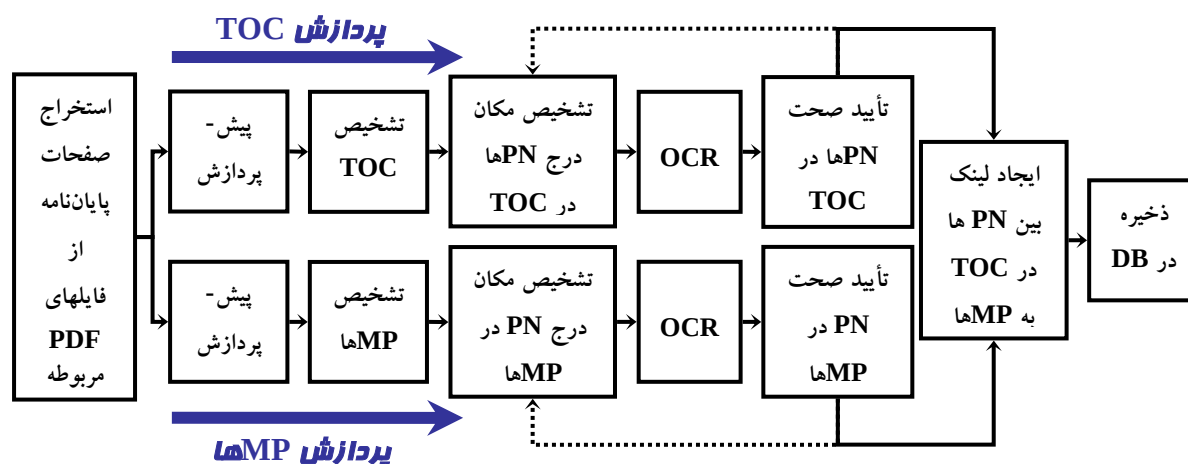
DB^۴: پایگاه داده

^۱ Main Page

^۲ Page Number

^۳ Optical Character Recognition

^۴ DataBase



شکل ۶-۱- بلوک‌دیگرام سیستم برقرارکننده لینک بین شماره صفحات پایان‌نامه‌ها در بخش فهرست مطالب به صفحات اصلی

عملکرد هر از این بلوک‌ها بصورت زیر می‌باشد:

الف- استخراج صفحات پایان‌نامه از فایل‌های PDF مربوطه: این سیستم بایستی قادر به خواندن

فرمت PDF و فرمتی که مطابق آن تصاویر در یک فایل PDF کد می‌شوند باشد.

ب- پیش‌پردازش: حداقل پردازشی که در این مرحله صورت می‌گیرد، حذف نویز فلفل-نمکی^۱ است. همچنین ممکن است لازم باشد که کج‌شدگی^۲ صفحات آشکار و تصحیح گردد (به‌خصوص در مورد صفحه TOC).

پ- تشخیص TOC: صفحه فهرست مطالب یا TOC در اغلب پایان‌نامه‌ها فرمت مشخصی دارد (شکل ۶-۲)؛ بصورتیکه که در پایان‌نامه‌های فارسی معمولاً در قسمت سمت راست هر خط، یک رشته متنی (عنوان بخشی از پایان‌نامه) و در قسمت سمت چپ آن، شماره صفحه مربوطه قرار دارد (در یک پایان‌نامه لاتین عکس این حالت اتفاق می‌افتد) و بین آنها ممکن است یک رشته نقطه‌چین، خط‌چین، نقطه-خط‌چین و یا یک خط آمده باشد. این اطلاعات می‌تواند به تشخیص TOC کمک نماید.

^۱ Salt-and-Pepper Noise

^۲ Skew

«فهرست»

I.....	چکیده
VI.....	اصطلاحات
فصل اول: معرفی سیستم‌های CBIR	
۲.....	۱-۱- مقدمه
۵.....	۱-۲- روش‌های مبتنی بر متن (TBIR)
۶.....	۱-۳- روش‌های مبتنی بر آموزش و یادگیری و مدل‌های بصری
۶.....	۱-۴- روش‌های مبتنی بر محتوا
۸.....	۱-۴-۱- بخش ارتباط با کلید
۹.....	۱-۴-۲- بخش بازیابی مبتنی بر محتوا

شکل ۶-۲- قسمتی از یک صفحه فهرست مطالب (TOC) در یک پایان‌نامه فارسی

از مقالات موجود در زمینه DIA دیده می‌شود که بیشتر تحقیقات انجام شده، درک سطح بالای صفحات TOC را مد نظر قرار داده‌اند تا بدینوسیله اطلاعات ساختاری TOC را استخراج نمایند و آن را به فرم ابرساختاری^۱ نظیر HTML یا XML نمایش دهند [38 و 39]. برای این منظور آنها فرض نموده‌اند که TOC یا قبلاً قطعه‌بندی شده است و یا برخی تکنیک‌های بازشناسی حروف/سمبول برای تعیین صفحات TOC و ساختار زیربنایی آنها مورد استفاده قرار می‌گیرند. گرچه بازشناسی سمبل بخشی از عملیات OCR می‌باشد، اما هنگامی که این نوع پردازش به مطالب ناهمسان قطعه‌بندی نشده (متن به همراه ناحیه ریاضیاتی^۲ و غیره) اعمال می‌شود، حجم محاسبات زیاد بوده، نتیجه کار غیر رضایت‌بخش خواهد بود.

Mandal و همکارانش در [38 و 39] عنوان نموده‌اند که یک راهکار مناسب‌تر برای یافتن TOC و استخراج اطلاعات آن، این است که قبل از اعمال OCR ابتدا TOC را با توجه به ویژگی‌های ساختاری آن مشخص سازیم و بدین ترتیب پردازش‌های تصویری بعدی و عملیات OCR را تنها بر روی نواحی متنی مورد نظر اعمال نماییم. لذا در این مقاله الگوریتمی بمنظور تشخیص خودکار صفحه TOC بدون بکارگیری تکنیک‌های بازشناسی حروف و با استفاده از فرم‌های مشترک متداول در

^۱ Meta-structure

^۲ Math-zone

صفحات TOC کتابها، مجلات، ژورنالها و غیره پیشنهاد گردیده است. به عبارت دیگر الگوریتم پیشنهادی تعیین می‌کند که آیا یک صفحه اسکن شده می‌تواند TOC باشد یا خیر؟ پس از این مرحله، الگوریتم بخشهای مختلف TOC اعم از عناوین فصلها یا بخشها و شماره صفحات مربوطه را شناسایی می‌نماید تا بدین ترتیب، اعمال تکنیکهای OCR با سهولت بیشتری انجام شود و نیز امکان جستجو و کاوش بخشهای مختلف سند برای کاربر فراهم آید.

در این دو مقاله با ملاحظه ۷۵ صفحه TOC مربوط به منابع مختلف لاتین چون کتابها، مجلات و غیره، ویژگیها و خصوصیات مشترک زیر در آنها گزارش گردیده است:

پ۱- TOC ممکن است یک صفحه کامل و یا بخشی از یک صفحه باشد. حالت اول بیشتر در مورد کتابها و گزارشها رخ می‌دهد. حالتها دوم نیز در برخی از ژورنالها که صفحه اول آنها شامل TOC به همراه آرم (LOGO)، عناوین و دیگر متنهاست، مشاهده می‌شود.

پ۲- در غالب مستندات، سمت راست‌ترین ستون در TOC حاوی شماره صفحات (PNها) می‌باشد و در آنها PNها بصورت راست‌چین مرتب گردیده‌اند (TOC-I - شکل ۶-۳-الف)؛ اما اسنادی نیز وجود دارند که PNها در TOC مربوط به آنها، تنها به فاصله اندکی بعد از هر عنوان بخش یا زیربخش ذکر می‌شود (TOC-II - شکل ۶-۳-ب). در این مورد معمولاً PNها به فاصله مناسبی از متن قرار دارند؛ ولی موارد خاصی نیز دیده شده که در آنها PNها تنها به فاصله اندکی بعد از متن (کمی بیشتر از فاصله معمول بین کلمات) قرار داشته‌اند.

پ۴- بین متن و PNها در هر خط ممکن است خطوط نقطه‌چین، خط‌چین و یا خط وجود داشته باشد. این "خطوط راهنما" بمنظور اجتناب از خطای دید چشم در مربوط کردن هر عنوان به شماره صفحه متناظرش بکار گرفته می‌شوند.

Contents Introduction 1 A Systematic Approach to Health Improvement 7 Healthy People 2010 Goals 8 Objectives 17 Determinants of Health 18 Health Status 21 Leading Health Indicators 24 Physical Activity 26 Overweight and Obesity 28 Tobacco Use 30 Substance Abuse 32 Responsible Sexual Behavior 34 Mental Health 36 Injury and Violence 38 Environmental Quality 40 Immunization 42 Access to Health Care 44 Bibliography 47 Appendix: Short Titles for Healthy People 2010 Objectives 53	vi WISCONSIN BLUE BOOK 2003 – 2004 TABLE OF CONTENTS State symbols Front and back endpapers Governor's letter iii Introduction iv BIOGRAPHIES 1 Alphabetical index to biographies 2 Constitutional executive officers 4 Supreme court justices 8 Wisconsin members of the 108th Congress 11 Congressional district map 17 State senate officers 18 State assembly officers 19 Members of the state senate and assembly; district maps 20 Chief clerks and sergeants at arms of the state legislature 86 Legislative district maps 87 FEATURE ARTICLE: Wisconsin's People: A Portrait of Wisconsin's Population on the Threshold of the 21st Century 99 Index to special articles in prior blue books 175 Capitol Visitor's Guide 177 WISCONSIN CONSTITUTION 179 History of constitutional amendments 222 Statewide referendum elections 228 FRAMEWORK OF WISCONSIN STATE GOVERNMENT Location of state agencies in the Madison area; map 232 The Framework of Wisconsin Government 234 Organization chart of Wisconsin state government 238 LEGISLATIVE BRANCH 241 Officers of the 2003 legislature 242 A profile of the legislative branch 243 The Wisconsin legislature 244 Session schedule, 2003-2004 247 News media correspondents accredited to the 2003 legislature 251 Legislative districts: 2000 population 252 How a bill becomes a law 253 Executive vetoes, 1993-2001 sessions 258 Legislative service 259 Political composition of the Wisconsin legislature, 1885-2003 260 Wisconsin statutes, session laws, administrative code 261 Committees of the 2003 legislature 262 Personal data on Wisconsin legislators, 1993-2003 sessions 266 Joint legislative committees 267 Administrative rules, joint committee for review of 267 Building commission, state of Wisconsin 268 Criminal penalties, joint review committee on 270 Employment relations, joint committee on 271 Finance, joint committee on 272 Information policy and technology, joint committee on 273 Legislative audit committee, joint 274 Legislative council, joint 274 Legislative organization, joint committee on 280
---	--

شکل ۶-۳- دو نوع متداول صفحه TOC اشاره شده در [38] و [39] الف- TOC-I با

ترازبندی راست PN ها ب- TOC-II دارای PN های توزیع شده

پ۵- در برخی از TOC ها ممکن است عناوین متنی چند خطی وجود داشته باشد. این حالت اکثراً در TOC ژورنالها که نام بعضی مقالات طولانی است، دیده می‌شود. در برخی از اسناد، هر یک از آیتمهای TOC با خطوط افقی از هم جدا می‌گردند.

پ۶- TOC ها برخی خواص مشابه با ساختار جدولی^۱ دارند.

پ۷- طول PN ها کوتاه بوده و معمولاً از یک تا سه کاراکتر (غالباً سه کاراکتر برای اسناد با حجم متوسط) تغییر می‌کند. واضح است که تعداد کاراکتر PN ها در TOC تغییرات اندکی دارد و بدون نوسان می‌باشد.

پ۸- PN ها تنها روند افزایشی دارند.

علاوه بر ویژگیهای فوق‌الذکر که در [38 و 39] بدانها اشاره شده است، در برخی از پایان-

نامه‌ها موارد زیر دیده می‌شود:

¹ Tabular Structure

پ ۹- در بعضی از پایان‌نامه‌ها، تمام یا برخی از PNها بصورت یک عدد نبوده، بلکه بشکل یک محدوده عددی (شکل ۶-۴-الف)، حروف الفبای فارسی یا یونانی (I, II, III, IV, ...) وحتى کاراکترهای غیر مرسوم (شکل ۶-۴-ب و ج) آمده‌اند.

پ ۱۰- TOC ممکن است در داخل یک قاب محصور باشد (شکل ۶-۴-د). همانطور که در ادامه این فصل توضیح داده خواهد شد، این امر می‌تواند مسئله شناسایی TOC را با مشکل روبرو سازد.

در [38 و 39] با توجه به ویژگیهای برشمرده شده در بالا عنوان گردیده است که علیرغم تنوع ساختارهای موجود در TOCها می‌توان آنها را به دو دسته کلی TOC-I و TOC-II تقسیم‌بندی نمود. در TOC-I (شکل ۶-۳-الف) PNها بصورت راست‌چین ترازبندی شده‌اند. در مقابل، در نوع TOC-II (شکل ۶-۳-ب) برخی PNها فاقد ترازبندی بوده و با یک فاصله پس از آیتیم مربوطه آمده‌اند. در الگوریتم ارائه شده در این مقاله برای شناسایی TOC، از کلیه ویژگیهای ذکر شده در بالا برای TOCها استفاده نگردیده است. مثلاً ویژگی "پ ۸" جز با بکارگیری تکنیکهای OCR قابل استفاده نیست و چون در اینجا بکارگیری این تکنیکها مد نظر نبوده است، این ویژگی مورد استفاده قرار نمی‌گیرد. همچنین ویژگیهای "پ ۳" و "پ ۴" نیز در طراحی الگوریتم شناسایی TOC لحاظ نگردیده‌اند.

در این مقالات الگوریتم شناسایی TOC، بر اساس یافتن PNها و عناوین بخشها/زیربخشهای مرتبط با آنهاست. PNها به عنوان کلمه، سمت راست‌ترین کلمه یک خط متن می‌باشند؛ متنها مطابق بند "پ ۵" ممکن است سمت راست‌ترین کلمه یک خط، کلمه‌ای از متن عنوان باشد. الگوریتم شناسایی TOC مطابق زیر می‌باشد:

صفحه	فهرست مندرجات*
مقدمه	*****
۱/	۱- تاریخچه سرخسك
۲/	۲- تعریف سرخسك
۳/	۳- اتیولوژی سرخسك
۵/	۴- اپیدمیولوژی سرخسك
۹/	۵- پاتولوژی سرخسك
۲۰/	۶- علائم بالینی سرخسك
۲۷/	۷- سرخسك شدید خونریزی دهنده
۲۸/	۸- سرخسك مودبقيه
۳۰/	۹- تشخیص سرخسك
۳۳/	۱۰- تشخیص افتراقی سرخسك
۳۵/	۱۱- عوارض سرخسك :
۳۵/	- اوتیت مانی
۳۷/	- ماستوئیدیت
۳۷/	- عوارض ریه‌ای
۳۸/	- لارنژیت و لارنکترانکیت

(ب)

فهرست مندرجات	مقدمه
۱-۳	خودکشی
۴-۷	عوامل خودکشی
۸-۹	علائم خودکشی قبل از ارتکاب
۱۰	جنبه‌های سهم خودکشی
۱۰	اهمیت خودکشی
۱۱	انواع حوادث
۱۱	خودکشی و طبقه‌بندی پس‌گیا تریک
۱۲-۱۷	بحران جوانی و نا هنجاریها
۱۸-۱۹	خودکشی در نوجوانان و جوانان
۲۰-۲۱	بحران جوانی
۲۱-۲۳	مداخلات اپیدمیولوژیک در شهر تهران
۲۴-۲۶	محل مطالعه و مدارک
۲۷-۳۰	جدول‌ها
۳۱-۳۴	بحث
۳۵-۴۳	نتیجه
۴۴-۴۵	خلاصه بیان مرانسه
۴۶	پیشگویی
۴۷	

(الف)

صفحه	موضوع
۱	مقدمه
۱	بخش اول
۱	کلیات
۱	فصل اول
۱	بررسی مفهوم لغوی و اصطلاحی جوهر
۷	(الف) جوهر در لغت
۸	(ب) جوهر در اصطلاح فلسفه
۱	فصل دوم
۱	مترادفات جوهر در زبان یونانی و انگلیسی
۱	فصل سوم
۱	جایگاه بحث از جوهر
۱	بخش دوم
۱	فصل اول
۱	تعریف جوهر نزد فیلسوفان متقدم بر ابن سینا
۲۲	(الف) تاریخچه کوتاهی از فلسفه یونانی
۲۴	(ب) بحث از جوهر در نزد ارسطو
۳۱	(ج) نظریات فلاسفه یونانی در زمینه جوهر

(د)

صفحه	موضوع
۱	چکیده پژوهش
۱	فصل یک - معرفی پژوهش
۳	زمینه پژوهش
۹	بیان مسئله پژوهش
۹	اهداف پژوهش
۱۵	سئوالات پژوهش
۱۱	اهمیت پژوهش
۱۳	بررسی مطالعات انجام شده
۱	فصل دوم - روش پژوهش
۲۵	نوع پژوهش
۲۵	جامعه آماری
۲۶	واحد آماری
۲۶	محبط پژوهش
۲۶	نمونه پژوهش
۲۶	حجم و نوع نمونه‌گیری
۲۹	روش گردآوری داده‌ها
۳۱	تعریف واژه‌ها
۳۳	متدولوژی پژوهش
۳۳	مدلهای جد اول عمر
۴۱	نشوری سراس (Brass)

(ج)

شکل ۶-۴-الف- PN در یک TOC که بصورت عددی یا محدوده عددی آمده‌اند. ب- وجود کاراکتر نامرسوم "/" در جلوی PN ها ج- وجود ستونی از کاراکترهای نامرسوم در جلوی PN ها د- TOC محصور در یک قاب

- آغشته‌سازی: در این مرحله، یک الگوریتم آغشته‌سازی در سطح حروف (به فصل ۵ رجوع شود) به صفحه‌اسکن شده اعمال می‌شود (شکل ۵-۶). سپس خطوط متن با محاسبه هیستوگرام افقی از یکدیگر مجزا می‌گردند.

12-1	Introduction	32
12-2	Microcomputer Organization	32
12-3	Microcomputer Organization	33
12-4	Instructions and Addressing Modes	34
12-5	Stack, Subroutines, and Interrupts	35
12-6	Memory Organization	32
12-7	Input-Output Mechanisms	37
12-8	Direct Memory Access	37
	Performance	38
	Problems	32

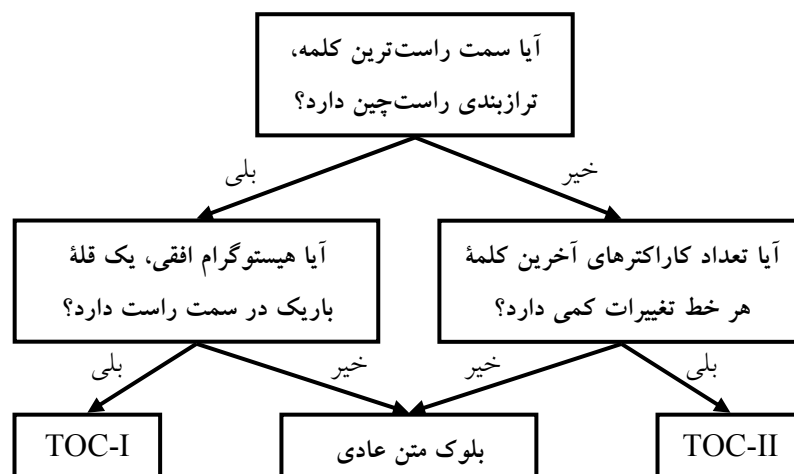
شکل ۵-۶- اعمال الگوریتم آغشته‌سازی در سطح حروف به TOC

- یافتن سمت‌راست‌ترین کلمه: برای یک خط متنی مجزای آغشته‌سازی شده، سمت راست‌ترین کلمه را می‌توان بوسیله هیستوگرام عمودی تعیین نمود. این کلمه کاندیدی برای PN می‌باشد مشروط بر آنکه به دنبال کلمه دیگری آمده باشد. سپس تعداد کاراکترهای این کلمه را محاسبه می‌نماییم.

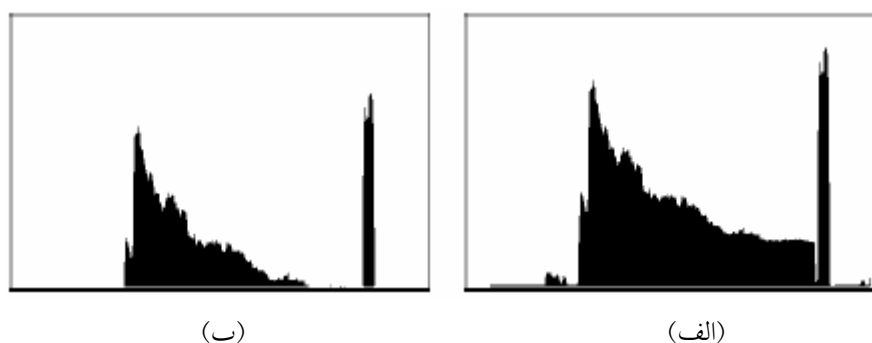
- شناسایی TOC: بمنظور شناسایی دو نوع TOC-I و TOC-II، از درخت تصمیم‌گیری^۱ مطابق شکل (۶-۶) استفاده می‌شود. منطق استفاده شده در این درخت تصمیم‌گیری برای تشخیص TOC-I و TOC-II بصورت زیر است:

TOC-I: در این مورد سمت راست‌ترین کلمه، ترازبندی راست‌چین دارد. ممکن است در هر خط فاصله زیادی بین PN و کلمه قبلی باشد. گاهی خطوط راهنما (به بند "پ ۴" مراجعه شود) نیز در این نوع TOC استفاده می‌شوند. در این نوع TOC، اثر PN‌های راست‌چین در هیستوگرام عمودی صفحه، بصورت یک قله باریک انعکاس می‌یابد (شکل ۶-۷ الف). بمنظور افزایش سهولت در آشکارسازی این قله، متوسط تعداد پیکسل‌های خطوط متن TOC به عنوان یک سطح آستانه بکار رفته است و بدین‌وسیله اثر خطوط راهنما حذف گردیده است (شکل ۶-۷ ب).

^۱ Decision Tree



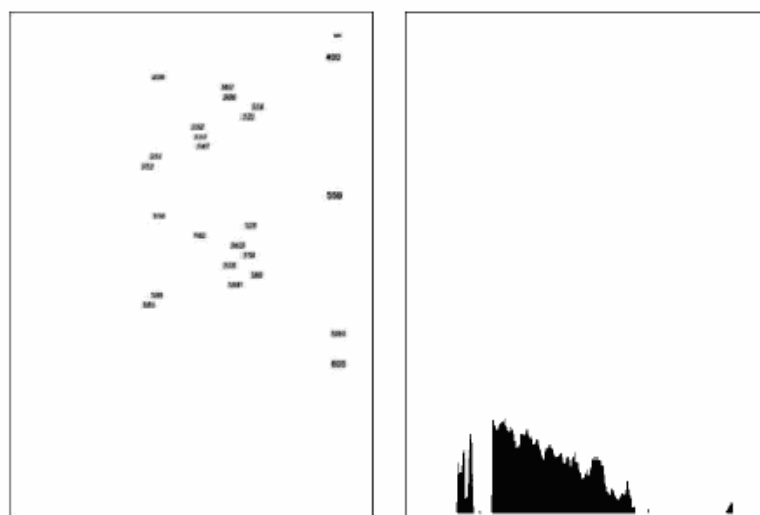
شکل ۶-۶- درخت تصمیم‌گیری برای مشخص نمودن دو نوع TOC-I و TOC-II



شکل ۶-۷- هیستوگرام عمودی صفحه الف- قبل از اعمال سطح آستانه الف- بعد از اعمال سطح آستانه

TOC-II: در این نوع TOC مانند نوع TOC-I، PN ها سمت‌راست‌ترین کلمه در تقریباً همه خطوط متن می‌باشند؛ منتهی در اینجا PN ها اغلب ترازبندی راست‌چین ندارند. در نتیجه یک قله بلند در هیستوگرام عمودی TOC-II ظاهر نمی‌گردد (شکل ۶-۸-الف). در نتیجه برای تشخیص این نوع TOC، تغییرات تعداد کاراکترهای سمت‌راست‌ترین کلمه هر خط مورد بررسی قرار می‌گیرد.

برای آزمایش این الگوریتم، از ۱۴۳ سند مختلف استفاده گردیده است که از میان آنها ۱۳۷ TOC بدرستی شناسایی شده‌اند و هیچ صفحه غیر TOC به غلط به عنوان TOC شناسایی نگردیده است [38 و 39].



(ب)

(الف)

شکل ۶-۸- ویژگیهای TOC-II الف- هیستوگرام عمودی صفحه ب- توزیع مکانی سمت راست‌ترین کلمه‌ها در هر خط

ت- تشخیص MPها: صفحات اصلی یا MPها با شماره ۱ آغاز می‌شوند. البته در برخی از پایان‌نامه‌ها در صفحات ابتدای هر فصل و برخی دیگر از صفحات اصلی، شماره صفحه چاپ نمی‌شوند؛ در نتیجه شماره این صفحات با توجه به شماره صفحه بعد یا قبل از آنها مشخص می‌گردد. اگر احتمال ناقص بودن پایان‌نامه‌ها را منتفی بدانیم (یعنی هیچ صفحه‌ای از پایان‌نامه‌ها مفقود نشده باشد)، بمنظور بالا بردن سرعت می‌توان تنها شماره چند صفحه اصلی ابتدایی را پیدا نمود و سپس شماره مابقی صفحات را با توجه به آنها محاسبه کرد.

ث- تشخیص مکان درج PNها در TOC: قبلاً گفته شد که در مورد پایان‌نامه‌های فارسی و عربی، شماره صفحات در TOC معمولاً در یک ردیف ستونی واقع در سمت چپ خطوط نوشته می‌شوند. همچنین در پایان‌نامه‌های لاتین این اعداد در یک ردیف ستونی واقع در سمت راست خطوط می‌آیند (شکل ۶-۳). بر طبق آنچه در بند "پ" ذکر شد، مکان درج PNها در TOC را می‌توان بوسیله هیستوگرام عمودی صفحه بدست آورد. از آنجا که مطابق شکل‌های (۶-۴-ب و ج و د) وجود کاراکترهای غیر رقمی یا قاب‌دربرگیرنده TOC در کنار PN محتمل است، بایستی مکان تشخیص داده شده برای PNها را توسط تکنیک‌های OCR تأیید نمود.

ج- تشخیص مکان درج PN در MPها: شماره صفحه (PN) ممکن است در حاشیه بالایی (سرصفحه^۱) یا پایینی (پاصفحه^۲) صفحات اصلی درج شده باشد (هر چند موارد نادری نیز دیده شده که PN در حاشیه سمت چپ یا راست صفحه آمده است). همچنین سه مکان چپ، راست و وسط برای درج PN در حاشیه بالایی و پایینی صفحات قابل فرض است. حاشیه صفحات ممکن است کاملاً ساده بوده و حاوی هیچ شیء خاصی بجز شماره صفحه نباشد (شکل ۶-۹). در این صورت تشخیص PNها به سبب قابل انجام است. از طرفی نیز امکان دارد اشیاء مختلفی نظیر متن، خطوط، جعبه‌ها، قاب طرحدار^۳، اشیاء خاص و حتی مهر کتابخانه IRANDOC در حاشیه صفحات موجود باشند. وجود نویزهای لکه‌مانند که گاهی به شکل یک کاراکتر رقمی می‌باشند را نیز نبایستی نادیده انگاشت. کیفیت بعضی از پایان‌نامه‌ها به حدی پایین است که تشخیص PNها حتی برای یک انسان نیز مشکل است. بعلاوه ممکن است PNها در یک جفت پرانتز یا گیومه یا در داخل یک کادر محصور باشند و یا در کنار یا زیر آنها شکلها یا خطوط خاصی موجود باشد. این موارد می‌توانند کار شناسایی PNها را بسیار پیچیده و حتی در مواردی غیر ممکن (در صورتیکه بخواهیم کلیه حالت‌های خاص را در طراحی سیستم لحاظ نماییم) سازند.

الفهرست		TABLE DES MATIÈRES:	
العنوان	الصفحة		
كلمة الطالب	۱- و	Introduction.....	1
القدمة	ز- ی	Chapitre I: Problématique du Beau.....	7
الباب الآتي		Chapitre II: Les critères du Beau chez Baudelaire.....	12
حياة الشاعر، زهد و شعره	۱- ۵۶	- Le Culte de la beauté.....	14
العمل الآتي : حياة الشاعر	۱	- La bizarrerie, principe de l'univers esthétique de Baudelaire.....	16
أ- إيمانه	۲	Chapitre III: Une destinée éternellement solitaire.....	28
ب- مولده	۳		

(ب)

(الف)

شکل ۶-۹- ستونهای درج شماره صفحات در: یک پایان‌نامه فرانسه‌زبان (شکل الف) و یک پایان‌نامه به زبان عربی (شکل ب). مرجع: وب‌سایت مرکز اطلاعات و مدارک علمی ایران.

¹ Header² Footer³ Page Border

چ- OCR: سیستمهای بازشناسی نوری حروف در فصول اول تا چهارم این طرح پژوهشی مورد بررسی قرار گرفتند. از آنجائیکه در این طرح بازشناسی ارقام ۰ تا ۹ فارسی و ۰ تا ۹ لاتین مد نظر است، با مشکل قطعه‌بندی حروف^۱ مواجه نیستیم. همچنین ممکن است در طرح آتی پیاده‌سازی سیستم، از روشهای بازشناسی بدون نیاز به قطعه‌بندی^۲ بمنظور شناسایی عناوین پایان‌نامه در بخشهای TOC و MP استفاده شود. علاوه بر این به لحاظ آنکه در مرحله "تأیید صحت" تا حدی خطاهای ناشی از بازشناسی غلط آشکار و تصحیح می‌گردد، می‌توان امیدوار بود که بتوان عمل شناسایی حروف را با دقت بالایی انجام داد.

ح- تأیید صحت PNها در TOC: همانگونه که در بند "چ" نیز عنوان گردید، مرحله تأیید صحت خطاهای ناشی از بازشناسی غلط ارقام را تا حدی تصحیح می‌نماید. بین شماره صفحات مربوط به عناوین در TOC این ارتباط منطقی وجود دارد که هر PN بایستی بزرگتر یا مساوی PN قبل از خود و کوچکتر یا مساوی PN بعد از خود باشد (شکل ۶-۱۰). در صورتیکه یک PN آشکارسازی‌شده از این ارتباط منطقی پیروی نکند، می‌توان یا با توجه به PNهای قبلی و بعدی آنرا تصحیح نمود و یا عمل لینک کردن را برای این PN انجام نداد.

فصل اول: معرفی سیستمهای CBIR	
۱-۱- مقدمه	۲
۱-۲- روشهای مبتنی بر متن (TBIR)	۵
۱-۳- روشهای مبتنی بر آموزش و یادگیری و مدل‌های بصری	۶
۱-۴- روشهای مبتنی بر محتوا	۶
۱-۴-۱- بخش ارتباط با کل‌بر	۸
۱-۴-۲- بخش بازیابی مبتنی بر محتوا	۹

شکل ۶-۱۰- ارتباط منطقی بین PNها در TOC. هر PN بزرگتر یا مساوی PN قبل از خود و کوچکتر یا مساوی PN بعد از خود است.

^۱ Character Segmentation

^۲ Segmentation-Free

خ- تأیید صحت PN در MPها: این مرحله ساده‌تر از مرحله ۸ است؛ چرا که هر PN تنها می‌تواند یکی بیشتر از PN قبلی خود و یکی کمتر از PN بعدی خود باشد. در صورتیکه PN یک صفحه چاپ نشده باشد، می‌توان عمل چک کردن با آن PN را با PNهای دیگر انجام داد.

د- ایجاد لینک بین PN ها در TOC به MPها: در این مرحله هر سطر از TOC که مربوط به یک عنوان پایان‌نامه است، به MP مربوط به آن لینک می‌شود.

ذ- ذخیره در DB: این بند نیز همانند بندهای اول و دهم، تنها یک پروسه نرم‌افزاری (و نه پردازشی) است که طی آن، لینکها در سیستم پایگاه داده مرکز ذخیره گشته و از طریق وب به کاربران سرویس-دهی می‌نماید.

۶-۲- پرسش و پاسخ

در این بخش سعی می‌شود که به برخی از پرسشهای اساسی در مورد عملکرد و راندمان این سیستم پیشنهادی اشاره شده و پاسخهای مناسبی برای آنها ذکر گردد. بعلاوه پیشنهادهایی نیز در جهت بهبود شرایط مسئله خواهد آمد.

سؤال ۱- آیا این سیستم می‌تواند تنوع وسیع فرمتها در پایان‌نامه‌ها را حداقل از لحاظ تئوری جوابگو باشد؟

پاسخ: این سوال را بدین صورت نیز می‌توان مطرح نمود که آیا می‌توان سیستمی طراحی نمود که کلیه حالت‌های محتمل در فرمت پایان‌نامه‌ها را پوشش دهد؟ برای پاسخگویی به این سؤال ابتدا لازم است که به مسئله تحت بررسی، نگاه دقیقتری داشته باشیم. یک بازبینی اجمالی پایان‌نامه‌های موجود در وبسایت IRANDOC بخوبی نشان می‌دهد که تنوع فرمتها تا چه اندازه گسترده است. به عنوان نمونه، به موارد زیر توجه نمایید:

الف- عدم یکسان بودن تعداد آیتمهای موضوعی: هر پایان‌نامه فارسی می‌تواند دارای صفحات عنوان (به فارسی)، بسم الله الرحمن الرحيم، فرم تأییدیه هیات ژوری، آیین‌نامه دانشگاه در مورد چاپ پایان‌نامه، تقدیر و تشکر، تقدیم به ...، چکیده، فهرست مطالب، فهرست اشکال، فهرست نمودارها،

فهرست جداول، فهرست نمادها و علائم اختصاری، فصول اصلی، ضمایم، منابع و مواخذ، چکیده انگلیسی، عنوان انگلیسی و غیره باشد. تعداد و ترتیب این صفحات از یک پایان‌نامه به پایان‌نامه دیگر کاملاً متفاوت است. حتی پایان‌نامه‌هایی مشاهده شده است که فاقد بخشهای اساسی نظیر فهرست مطالب یا چکیده هستند.

ب- عدم یکسان بودن عنوان هر یک از آیتمها: در زبان انگلیسی، برخی واژه‌ها برای کاربردهای خاصی اختصاصی شده‌اند و برای آنها دیگر از واژه‌های معادل استفاده نمی‌شود. مثلاً در متون علمی، واژه‌هایی مانند Title, Subject, Abstract, Contents (یا Table of Contents), Acknowledgements, Preface, Introduction, Chapter, Section, Figure, Table, Index, Annex, Appendix, References, Conclusion, Keywords, Glossary, Abbreviations و نظایر اینها کاملاً استاندارد گردیده‌اند و در نتیجه یک سیستم آنالیزکننده این اسناد براحتی می‌تواند با شناسایی چنین کلماتی، قسمت‌های مختلف یک سند را از یکدیگر تفکیک نماید. اما متأسفانه در مورد زبان فارسی چنین مطلبی به هیچ عنوان صادق نیست. مثلاً در پایان‌نامه‌ها دیده می‌شود که برای مشخص کردن صفحه TOC، عباراتی نظیر "فهرست"، "فهرست مطالب"، "فهرست موضوعات"، "فهرست عناوین"، "فهرست مباحث"، "فهرست مندرجات"، "شرح عنوانها"، "عنوان بخشها" و غیره یا حتی بدون عنوان مورد استفاده قرار گرفته است. حتی اگر سیستمی طراحی گردد که کلیه عبارتهای تاکنون بکار رفته را بشناسد، تضمینی نیست که برای پایان‌نامه‌هایی که در آینده به پایگاه اطلاعاتی اضافه می‌شوند، این سیستم بتواند همواره جوابگو باشد. از طرف دیگر، بخشهایی مانند فهرست اشکال یا جداول یا نمودارها، فرمتی کاملاً مشابه با TOC دارند و برای تمیز دادن این بخشها از یکدیگر لازم است که عنوان صفحات آنها شناسایی گردد. عدم استاندارد بودن اسامی این بخشها، در اینجا نیز مشکل‌ساز خواهد بود. بعلاوه مشکل پیوسته بودن^۱ کلمات در نگارش فارسی که کار شناسایی کلمات را در مقایسه با کلمات گسسته انگلیسی بسیار پیچیده‌تر می‌کند نیز نبایستی از نظر دور داشت.

^۱ Cursive

ج- استاندارد نبودن فرمت تدوین هر بخش از پایان‌نامه: در توضیح این مطلب، برای نمونه به مسئله شناسایی مکان شماره صفحات (PN) در صفحات اصلی (MPها) توجه فرمائید. مکان مورد انتظار برای وقوع PNها، حاشیه MPهاست. اما قابهای طرحدار حاشیه صفحات که در برخی از پایان‌نامه‌ها در مورد استفاده قرار می‌گیرد، می‌تواند شناسایی PNها را با مشکل جدی روبرو سازد. شکل (۶-۱۱) صفحه‌ای از یک پایان‌نامه را نشان می‌دهد که در آن متن اصلی و نیز شماره صفحه در داخل قابهایی محصور گردیده‌اند. همچنین در شکل (۶-۴-د) صفحه TOC از یک پایان‌نامه مشاهده گردید که بوسیله یک قاب احاطه شده بود. چنانچه حتی شکل این قابها از قبل استاندارد شده بود، امکان‌پذیر بود که برای آنها الگوریتم مناسب را طراحی کرد. لیکن از آنجا که در انتخاب این قابها کاملاً سلیقه‌ای عمل می‌شود، طراحی یک سیستم عام در این مورد ممکن است بسیار پیچیده و حتی ناشدنی باشد.



شکل ۶-۱۱- صفحه‌ای از یک پایان‌نامه که در آن متن اصلی و شماره صفحه، درون قابهای طرحدار محصور گردیده‌اند

د- کیفیت بسیار پایین برخی از پایان‌نامه‌ها: اساساً ورودیهای هر سیستم پردازشی ممکن است آمیخته به نویز باشند. هر سیستمی تا درجه مشخصی از نویز را می‌تواند تحمل نماید که برای بیشتر از آن لازم است بوسیله پیش‌پردازش مقدار نویز کاهش یابد. تکنیکهای مختلف کاهش نویز در فصل دوم مورد بررسی قرار گرفتند؛ اما بایستی توجه نمود که قدرت عمل این تکنیکها محدود می‌باشد. کیفیت برخی از پایان‌نامه‌ها در اثر گذشت زمان، کیفیت بد دستگاه کپی‌برداری یا عوامل دیگر آنچنان پایین آمده است که حتی تشخیص متن را برای انسان نیز مشکل می‌سازد (شکل ۶-۱۲). در این موارد ممکن است تکنیکهای کاهش نویز نتوانند عملکرد کاملاً قابل قبولی ارائه دهند.

۱۲

((سنگی های چهار کلمه ای))

۱- سنگی نام با درجه A چون اب با پایان سنگین بود با پایان درجه همدیگر
 مشترک به نسبت ۱- با اب مقدار رقیق شد بهر حال در محلول مابین اب و سنگ
 دانه عدد تا کف به سانتیمتر بدست ۵ دقیقه دوام پیدا کرد.

مقدار مری مابین ۱۱ و ۱۰ = ۱/۱

در یک محلول مابین صرف شده $\frac{۲۲}{۲۸+۱}$

سنگی نام اب بر حسب درجه تراشوی $۲۲/۱۷ = \frac{۲۲}{۲۸+۱} \times ۱۱$

۲- سنگی مریوط به نام اصلاح اب با یک کلمه با درجه B
 به ۵ سانتیمتر کتب اب مورد آزمایش ۵ سانتیمتر کتب محلول ۱- اصلاح اب و نیم اب و اب
 می‌کند و ۲۰ دقیقه به حال خود می‌گذارد تا اصلاح کتب اب به صورت اصلاح کتب رسوب
 کند محلول را صاف کرده به نسبت ۱- رقیق نموده به طریقی که ذکر شد سنگی آن -
 تعیین گردید.

مری مابین ۵۰ و ۱۰ = ۵/۵

سنگی B اب بر حسب درجه تراشوی $۴۲/۱ = \frac{۲۲}{۲۸+۱} \times ۵۰$

در صورتیکه کتب اب زیاد باشد باید اصلاح اب و نیم مریوط صرف شود که صاف شده با

شکل ۶-۱۲- صفحه‌ای از یک پایان‌نامه دارای کیفیت بسیار نازل

حال به سؤال مطرح شده برمی‌گردیم. همانگونه که در فصل پنجم نیز گفته شد، اکثر تحقیقات انجام شده در زمینه آنالیز اسناد توجه خود را به زمینه کاربردی خاصی محدود کرده‌اند تا با استفاده از شیوه‌های نگارش خاص آن زمینه که از قبل برای طراح سیستم معلوم است، راندمان را تا حد امکان بالا ببرند. به عنوان مثال ممکن است سیستمی طراحی شود که بتواند تنها مقاله‌های

علمی منتشر شده در یک مجله یا کنفرانس خاصی را پردازش کند. در اینجا چون طراح سیستم تنها با یک مسئله خاص روبرو است، می‌تواند از قواعد نگارشی بکار رفته در آن زمینه خاص (که از قبل معلوم است) حداکثر استفاده را بمنظور بالا بردن کارایی سیستم بنماید. اما در مورد پایان‌نامه‌های داخل کشور، بدلیل آنکه متاسفانه تا کنون هیچگونه استانداردی در رابطه با شیوه تدوین آنها وجود ندارد، با یک مسئله کاملاً عام روبرو هستیم و می‌بایست سیستم عامی طراحی نماییم که طیف بسیار وسیعی از فرمت‌ها را جوابگو باشد. بنابراین با توجه به مواردی که در بالا توضیح داده شد، امکان دارد طراحی سیستمی که بتواند کلیه فرمت‌های ممکنه را پاسخگو باشد، حتی از لحاظ تئوری نیز عملی نباشد.

سؤال ۲- چگونه می‌توان راندمان سیستم را تا حد امکان افزایش داد؟

پاسخ: راهکارهای زیر بمنظور افزایش راندمان سیستم پیشنهاد می‌شود:

الف- محدود نمودن ورودیهای سیستم: در مجموعه‌ای اینچنین وسیع از پایان‌نامه‌ها که استانداردسازی نشده‌اند، وجود هر فرمت نامتداولی انتظار می‌رود. طراحی سیستمی که بتواند یک پایان‌نامه با هر فرمت دلخواه را مورد پردازش صحیح قرار دهد، عملاً غیر ممکن می‌باشد. اما از طرف دیگر بایستی توجه داشت که فرمت بیشتر پایان‌نامه‌ها، در یک محدوده خاص تغییرات قرار دارد. بنابراین یک راهکار مناسب برای افزایش کارایی سیستم آن است که طراحی الگوریتم‌ها را تنها برای این دسته از پایان‌نامه‌ها انجام دهیم (ورودیهای مناسب). در بهترین حالت، سیستم پس از دریافت یک پایان‌نامه مشخص می‌کند که آیا پایان‌نامه یک ورودی مناسب است یا خیر (اگر سیستم خود نتواند ورودیهای مناسب را تشخیص دهد، می‌بایست با نظارت انسانی این کار را انجام داد). اگر پاسخ مثبت باشد، سیستم ورودی را قبول خواهد نمود.

ب- نیمه‌خودکار سازی سیستم پردازشی: طراحی سیستم پیشنهادی با این هدف انجام شد که کل فرایند پردازش بصورت خودکار انجام گیرد. اما بمنظور افزایش راندمان سیستم می‌توان به آن حالت

”تعاملی”^۱ بخشید. بدین صورت که سیستم پس از انجام پردازش هر مرحله (یافتن صفحه TOC، یافتن MPها، تشخیص مکان PNها و غیره)، آن را به تأیید کاربر برساند. در این حالت لازم است که برنامه یک واسطه گرافیکی (GUI)^۲ مناسب داشته باشد. به عنوان مثال فرض شود که سیستم می-خواهد صفحه TOC را پیدا کند. بدین منظور سیستم صفحه‌ای را که حدس می‌زند صفحه TOC باشد، بصورت کوچک شده^۳ در GUI برنامه نشان داده و از ناظر برنامه درخواست تأییدیه می‌نماید. چنانچه ناظر عملیات پردازشی را تأیید ننماید، سیستم فرایند را از نو آغاز می‌کند. همچنین در ساده-ترین حالت می‌توان سیستمی طراحی نمود که در آن ناظر برنامه در یکی از صفحات اصلی (MPها) با ماوس، مکان شماره صفحه (PN) را مشخص سازد و برنامه با الهام از این انتخاب ناظر، برای صفحات بعدی، خود PNها را پیدا نماید.

سؤال ۳- در مورد پایان‌نامه‌ها و گزارشهای آتی که در کشور تولید می‌شوند، چه راهبردهایی پیشنهاد می‌شود؟

پاسخ: قبلاً گفته شد که چنانچه پایان‌نامه‌ها و اسناد کشور بر طبق یک استاندارد خاص تدوین می‌شدند، یک سیستم آنالیزکننده اسناد یا یک برنامه محاسباتی براحتی می‌توانست آنها را مورد پردازش قرار دهد. به عنوان نمونه هم‌اکنون بسیاری از سازمانها و کنفرانسهای خارجی، یک فایل متنی (نظیر یک فایل Template برنامه Microsoft Word) را در اختیار کاربران قرار می‌دهند که در آن قالب گزارشی که قرار است تدوین گردد، کاملاً رعایت شده است و کاربر براحتی می‌تواند متن خود را درون این قالبهای از پیش تعیین شده قرار دهد. اقداماتی از این دست نیز می‌بایست در مورد پایان-نامه‌ها و اسناد داخل کشور اعمال گردد. ضمن اینکه دو استاندارد ODA و SGML در زمینه تولید اسناد موجود می‌باشند (به فصل ۵ مراجعه شود). با اتخاذی تدابیر مناسب در این راستا می‌توان به

^۱ Interactive

^۲ Graphic User Interface

^۳ Thumbnail

ایجاد شدن یک محیط شبکه‌ای فارغ از کاغذ^۱ در سطح تمامی کتابخانه‌ها، ادارات و سازمانهای کشور در آینده امیدوار بود.

^۱ Paperless

مراجع

- مراجع لاتین^۱

- 1 Abulhaiba S. H., Mahmood S. A., Green R. J., "Recognition of handwritten cursive Arabic characters", IEEE Trans. On PAMI, Vol. 16, No. 6, pp. 664-671, 1994.
- 2 Altuwaijir M. and Bayoumi M., "Arabic text recognition using neural networks", Proc. Int. Symp. On Circuit and Systems (ISCAS'94), pp. 415-418, 1994.
- 3 Al-Usefi H. and Udupa S.S., "Recognition of Arabic characters", IEEE PAMI, Vol. 14, No. 8, pp. 853-857, 1992.
- 4 Arica N. and Yarman-Vural F.T., "One dimensional representation of two dimensional information for HMM based handwritten recognition", Pattern Recognit. Lett., vol. 21, No. 6-7, pp. 583-592, 2000.
- 5 Arica Nafiz, Vural Fatoş Yarman, "A new scheme for off-line handwritten connected digit recognition", 14th Int. Conf. Pattern Recognit., Brisbane, Australia, pp. 1127-1131, 1998.
- 6 Arica Nazif, Yamin-Vural Fatos T., "An overview of character recognition based focused on off-line handwriting", IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, Vol. 31, No. 2, May 2001.
- 7 Azmi R., Kabir E., "A recognition algorithm for hand-printed farsi characters", Proc. Of ICT96, Istanbul, Turkey, pp. 852-855, 1996.
- 8 Badie K., Shimora M., "Machine recognition of Arabic cursive script", Pattern Recognition in Practice, Proc. of International Workshop, Amsterdam, Netherlands, pp. 315-323, 1980.
- 9 H. Baird, "Document image defect models", IAPR Workshop on Syntactic and Structural Pattern Recognition, pages 38-46, Murray Hill, NJ, June 1990.
- 10 Baird H.S., Jones S.E., and Fortune S.J., "Image segmentation by shape-directed covers", Proc. of the 10th International Conference on Pattern Recognition, Atlantic City, NJ, June, 17-21, 1990.
- 11 Bozinovic R. M. and Srihari S. N., "Off-line cursive script word recognition", IEEE Trans. Pattern Anal. Machine Intell., vol. 11, pp. 68-83, Jan. 1989.
- 12 Casey R. G. and Lecolinet E., "A survey of methods and strategies in character segmentation", IEEE Trans. Pattern Anal. Machine Intell., vol. 18, pp. 690-706, July 1996.
- 13 Casey R., Ferguson D., Mohiuddin K. and Walach E., "Intelligent Forms Processing System", Machine Vision and Applications, vol. 5, No. 3, pp. 143-155, 1992.
- 14 Chen M. Y., Kundu A., and Zhou J., "Off-line handwritten word recognition using a hidden Markov model type stochastic network", IEEE Trans. Pattern Anal. Machine Intell., vol. 16, pp. 481-496, May 1994.
- 15 Correia S.E.N., De Carvalho J.M., Sabourin, R., "Human-perception handwritten character recognition using wavelets", Proceedings XV Brazilian Symposium on Computer Graphics and Image Processing, pp.404, Oct. 2002.

^۱ مراجع بترتیب حروف الفبا مرتب شده‌اند. ابتدا نام خانوادگی و سپس نام هر نویسنده ذکر گردیده است.

- 16 Correia S.E.N., De Carvalho J.M., Sabourin, R., "On the performance of wavelets for handwritten numerals recognition", 16th International Conference on Pattern Recognition, Vol. 3, pp. 127 -130, Aug. 2002.
- 17 Dengel A., Bleisinger R., Hoch R., Fein F., Hones F., "From Paper to Office Document Standard Representation", IEEE Computer, vol. 25, No. 7, pp. 63-67, 1992.
- 18 El-Sheikh T.S., Guindi R. M., "Computer recognition of Arabic cursive script", Pattern Recognition 21, pp. 293-302, 1988.
- 19 Esposito F., Malerba D., Semeraro G., Annese E., and Scafuro G., "An experimental page layout recognition system for office document automatic classification: an integrated approach for inductive generalization", IEEE International Conference on Pattern Recognition (Atlantic Ctty, N J), pp. 557-562, 1990.
- 20 GE Yong, HUO Quiang, FENG Zhi-Dan, "Offline recognition of handwritten Chinese characters using Gabor features, CDHMM modeling and MCE training", IEEE 2002.
- 21 Gilloux E., "Hidden Markov models in handwriting recognition", in Fundamentals in Handwriting Recognition. ser. NATO ASI F, S. Impedovo, Ed. New York: Springer-Verlag, vol. 124, 1994.
- 22 Guillevic D. and Suen C. Y., "Cursive script recognition: A sentence level recognition scheme", in Proc. Int. Workshop Frontiers in Handwriting Recognit., pp. 216-223, 1994.
- 23 Guillevic D. and Suen C. Y., "HMM-KNN word recognition engine for bank cheque processing", in Proc. 14th Int. Conf. Pattern Recognit., Brisbane, Australia, 1998, pp. 1526-1529.
- 24 Hu Peifeng, Zhao Yannan, Yang Zehong, Wang Jiaqin, "Recognition of gray character using gabor filters", Proceedings of the Fifth International Conference on Information Fusion, Vol.1 , pp. 419 -424, July 2002.
- 25 Huang Lei, Huang Xiao, "Multiresolution recognition of offline handwritten Chinese characters with wavelet transform", Sixth International Conference on Document Analysis and Recognition, pp. 631 -634, Sept. 2001.
- 26 Jain A. K. and Yu B., "Document representation and its application to page decomposition", IEEE Trans. Pattern Anal. Machine Intell., vol. 20, pp. 294-308, Mar. 1998.
- 27 Jinying Chen, Yijiang Jin, Shaoping Ma, "A learning algorithm detecting the similar Chinese characters' boundary based on unequal-contraction of dimension", Proceedings of the 3rd World Congress on Intelligent Control and Automation, vol.4, pp. 2765 -2769, July 2000.
- 28 Kasturi R., O'Gorman L., and Govindaraju V., "Document Image Analysis: A Primer", Sadhana, special issue on Document Processing, accepted, in press, 2002.
- 29 Khorsheed M.S., Clocksin W.F., "Multi-font Arabic word recognition using spectral features", 15th International Conference on Pattern Recognition, vol. 6, pp. 3574 - 3577, June 2000.

- 30 Kornai András, Mohiuddin K.M., Connell Scott D., "Recognition of cursive writing on personal checks", in Proc. Int. Workshop Frontiers Handwriting Recognit., Essex, U.K., pp. 373–378, 1996.
- 31 Kunango T., Haralick R., and Phillips I., "Nonlinear local and global document degradation models", Int. J. Imaging Syst. Technol., vol. 5, no. 4, pp. 274–282, 1994.
- 32 Lam L., Lee S. W., and C. Y. Suen, "Thinning methodologies—A comprehensive survey", IEEE Trans. Pattern Anal. Machine Intell., vol. 14, pp. 869–885, Sept. 1992.
- 33 Lam Louisa, Suen Ching Y., "An evaluation of parallel thinning algorithms for character recognition", IEEE Trans. Pattern Anal. Machine Intell., vol. 17, pp. 914–919, Sept. 1995.
- 34 Lee S.W., Lee D. J., and Park H. S., "A new methodology for gray-scale character segmentation and recognition", IEEE Trans. Pattern Anal. Machine Intell., vol. 18, pp. 1045–1050, Oct. 1996.
- 35 Madhvanath S., Kim G., and Govindaraju V., "Chaincode contour processing for handwritten word recognition", IEEE Pattern. Anal. Machine Intell., vol. 21, pp. 928–932, Sept. 1999.
- 36 Mahmoud S. A., Abuhabia I., and R. J. Green, "Skeletonization of Arabic characters using clustering based skeletonization algorithm", Pattern Recognit., vol. 24, no. 5, pp. 453–464, 1991.
- 37 Mahmoud S., "Arabic Character Recognition using Fourier descriptors and character encoding", Pattern Recognition, Vol. 27, No. 6, pp. 815-824, 1994.
- 38 Mandal S., Chowdhury S.P., Das A.K. and Chanda B., "Automated Detection and Segmentation of Table of Contents Page from Document Images", 7th International Conference on Document Analysis and Recognition (ICDAR'03), pp. 398-402, Aug. 2003.
- 39 Mandal S., Chowdhury S.P., Das A.K. and Chanda B., "Automated Detection and Segmentation of Table of Contents Page and Index Pages from Document Images", 12th International Conference on Image Analysis and Processing, pp. 213-218, Sept. 2003.
- 40 Manjunath B.S. and W. Ma, "Texture features for browsing and retrieval of image data", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 8, pp. 837-842, August 1996.
- 41 Menhaj M.B., Adab M., "Simultaneous segmentation and recognition of Farsi/Latin printed texts with MLP", International Joint Neural Networks: IJCNN '02, Vol. 2, pp. 1534 -1539, May 2002.
- 42 Mohamed M. and Gader P., "Handwritten word recognition using segmentation-free hidden Markov modeling and segmentation based dynamic programming techniques", IEEE Trans. Pattern Anal. Machine Intell., vol. 18, pp. 548–554, May 1996.
- 43 Mori S., Suen C. Y., and Yamamoto K., "Historical review of OCR research and development", Proc. IEEE, vol. 80, pp. 1029–1057, July 1992.

- 44 Mowlaei A., Faez K., Haghighat A.T., "Feature extraction with wavelet transform for recognition of isolated handwritten Farsi/Arabic characters and numerals", 14th International Conference on Digital Signal Processing, Vol. 2, pp. 923-926, July 2002.
- 45 Nagy George, Seth Sharad, Viswanathan Mahesh, "A Prototype Document Image Analysis System for Technical Journals", Computer, vol 25 no 7, pp 10-22, July 1992.
- 46 Nagy, G., "Twenty years of document image analysis in PAMI", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp. 38-62, Jan 2000.
- 47 Niblack W., *An Introduction to Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1986.
- 48 O'Gorman Lawrence, "The Document Spectrum for Structural Page Layout Analysis", IEEE Trans. Pattern Analysis and Machine Intelligence, Nov., 1993.
- 49 O'Gorman Lawrence, Kasturi Rangachar, *Document Image Analysis*, IEEE Computer Society Press, Los Alamitos, 1995.
- 50 Srihari Sargur, Lam Stephen, Govindaraju Venu, Srihari Rohini and Hull Jonathan, "Document Understanding: Research Directions", Technical report CEDAR-TR-92-1, State University of New York at Buffalo: Center of Excellence for Document Analysis and Recognition (CEDAR), May 1992.
- 51 Suen C. Y., Tappert C. C., and Wakahara T., "The state of the art in online handwriting recognition", IEEE Trans. Pattern Anal. Machine Intell., vol. 12, pp. 787-808, Aug. 1990.
- 52 Tayel M., Shalaby H., Saleh, H., "Winner-take-all neural network for visual handwritten character recognition", Thirteenth National Radio Science Conference: NRSC '96, pp. 239 -249, March 1996.
- 53 Theodoridis Sergios, Koutroumbas Konstantinos, Smith Ricky, *Pattern Recognition*, Academic Press, 1st edition, January 15, 1999.
- 54 Tomio, Yamashita Akio, Itoh Nobuyasu, Kobayashi Yoshinao, Katoh Shin, Toyokawa Kazuharu and Takahashi Hiroyasu, "DRS: a workstation-based document recognition system for text entry", IEEE Computer Vol. 25, No. 7, pp. 67-71, 1992.
- 55 Trier Ø. D. and Jain A. K., "Goal directed evaluation of binarization methods," IEEE Trans. Pattern Anal. Machine Intell., vol. 17, pp. 1191-1201, Dec. 1995.
- 56 Trier O. D., Jain A. K. and Taxt T., "Feature extraction methods for character recognition - A survey", Pattern Recognition 29, pp. 641-662, 1996.
- 57 Wang Xuewen, Ding Xiaoqing, Liu Changsong, "Gray-scale character image recognition based on fuzzy DCT transform features", 15th International Conference on Pattern Recognition, Vol. 2 , pp. 235 -238, Sept 2000.
- 58 Wieser Johann and Pinz Axel, "Layout and analysis: Finding text, titles, and photos in digital images of newspaper pages", Proceedings of the 2nd ICDAR, Int. Conf. Document Analysis and Recognition, Japan, pages 774-777, IEEE press, 1993.
- 59 Wong K.Y., Casey R.G., and Wahl F.M., "Document analysis system", IBM J. Research Development, Vol. 6, No. 6, pp. 647-656, Nov 1982.

-
- 60 Yang Y. and Yan H., "An adaptive logical method for binarization of degraded document images", *Pattern Recognit.*, vol. 33, no. 5, pp. 787–807, 2000.
 - 61 Yanikoglu B. A. and Sandon P., "Recognizing off-line cursive handwriting", *Proc. Int. Conf. Comput. Vision Pattern Recognit.*, Seattle, WA, 1994, pp. 397–403.

- مراجع فارسی

۶۲ اردشیر بهرستاقی غلامرضا، "شناسایی ساختاری حروف دستنویس فارسی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: احسان الله کبیر، دانشکده مهندسی برق، دانشگاه تربیت مدرس، ۱۳۷۲.

۶۳ اسدی سعید، "بازشناسی حروف فارسی با استفاده از شبکه عصبی چند جمله‌ایهای جداکننده"، پایان‌نامه کارشناسی ارشد، به راهنمایی: فتحی محمود، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی اصفهان، ۱۳۷۷.

۶۴ امیری امیررضا، "استخراج متن چاپی از تصاویر گرافیکی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: کبیر احسان الله، دانشکده مهندسی برق، دانشگاه تربیت مدرس، ۱۳۸۱.

۶۵ بحری پیمان، "شناسایی حروف دستنویس فارسی به کمک شبکه عصبی فازی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: فتحی محمود، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، ۱۳۷۶.

۶۶ بنی اسدی امیرعلی، "تشخیص حروف دستنویس فارسی به وسیله سیستم هایپرید نور و فازی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: ساداتی ناصر، دانشکده مهندسی برق، دانشگاه صنعتی شریف، ۱۳۷۳.

۶۷ پارسی اصفهانی امیر، "بررسی روشهای ارائه اطلاعات بصورت تمام تصویر در صفحات وب"، مرکز اطلاعات و مدارک علمی ایران، بهار ۱۳۷۹.

۶۸ تیمساری بیژن، "بازشناسی حروف در کلمات تایپ شده فارسی با استفاده از روش مورفولوژی"، به راهنمایی: فهیمی حمید، دانشکده مهندسی برق، دانشگاه صنعتی اصفهان، ۱۳۷۱.

۶۹ ثانی رویا، "بازشناسی حروف دستنویس فارسی با استفاده از شبکه عصبی مصنوعی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: فهیمی مهرداد، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، ۱۳۷۳.

۷۰ رادمهر مهدی، "درک پیکربندی هندسی یک صفحه متن چاپی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: احسان الله کبیر، دانشکده مهندسی برق، دانشگاه تربیت مدرس، ۱۳۷۲.

- ۷۱ رحمتی محمد، زارعی ناصر، "شناسایی حروف تاییی فارسی با استفاده از تبدیل خودهمبستگی مورفولوژی"، سومین کنفرانس انجمن کامپیوتر ایران، ص ۱۸۵-۱۸۰، دانشگاه علم و صنعت ایران، تهران، دی ۱۳۷۶.
- ۷۲ دهقانی علیرضا، "بازشناسی حروف مجزای دستنویس فارسی با استفاده از مدل پنهان مارکف با چگالی پیوسته و ایده ترکیب چند سیستم خبره، پایان‌نامه کارشناسی ارشد، به راهنمایی: مسندی شیرازی محمدعلی، دانشکده مهندسی برق، دانشگاه شیراز، ۱۳۷۹.
- ۷۳ رضوی محمد، "خواندن اتوماتیک فرمهای انتخاب درس"، پایان‌نامه کارشناسی ارشد، به راهنمایی: کبیر احسان‌الله، دانشکده مهندسی برق، دانشگاه تربیت مدرس، ۱۳۷۵.
- ۷۴ رفیعی کراچی شعبانعلی، "شکستن کلمات تایپ شده به حروف در رسم الخطهای مختلف"، پایان‌نامه کارشناسی ارشد، به راهنمایی: کبیر احسان‌الله، دانشکده مهندسی برق، دانشگاه تربیت مدرس، ۱۳۷۳.
- ۷۵ شاه‌حسینی علی، "شناسایی حروف دستنویس فارسی با استفاده از شبکه عصبی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: کبیر احسان‌الله، دانشکده مهندسی برق، دانشگاه تربیت مدرس، ۱۳۷۴.
- ۷۶ شیرعلی شهرضا محمد حسن، "تشخیص کلمات و ارقام دستنویس فارسی به کمک شبکه‌های عصبی"، رساله دکتری، دانشکده برق، دانشگاه صنعتی امیرکبیر، ۱۳۷۴.
- ۷۷ صدوقی‌یزدی هادی، "پیش پردازش برای بازشناسی متون فارسی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: کبیر احسان‌الله، دانشکده مهندسی برق، دانشگاه تربیت مدرس، ۱۳۷۵.
- ۷۸ صفابخش رضا، آنالیز اسناد - تصحیح کج‌شدگی اسناد، طرح ملی "بازشناسی متن فارسی"، گروه (ب)، بهار ۱۳۷۸.
- ۷۹ صفابخش رضا، تصحیح کج‌شدگی اسناد، طرح ملی "بازشناسی متن فارسی"، گروه (ب)، گزارش سه‌ماهه اول، خرداد ۱۳۷۸.
- ۸۰ صفابخش رضا و دست‌پاک وحدت، "شناسایی مستقل از چرخش و اندازه حروف تاییی فارسی"، مجموعه مقالات اولین کنفرانس مهندسی برق ایران، جلد دوم، ص ۲۷۶-۲۷۷، دانشگاه صنعتی امیرکبیر، تهران، اردیبهشت ۱۳۷۲.

- ۸۱ عباسیان کریم، "بازشناسی برخط نویسه‌های فارسی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: کبیر احسان‌الله، دانشکده مهندسی برق، دانشگاه تربیت‌مدرس، ۱۳۷۶.
- ۸۲ عزمی رضا، "بازشناسی متون چاپی فارسی"، پایان‌نامه دکتری، به راهنمایی: کبیر احسان‌الله، دانشکده فنی و مهندسی، دانشگاه تربیت‌مدرس، ۱۳۷۸.
- ۸۳ فائز کریم، ختن‌زاد علیرضا و شیرعلی شهرضا محمدحسن، "تشخیص حروف و ارقام دستنویس فارسی با استفاده از گشتاورهای شبکه زرنیکه و بکمک شبکه‌های عصبی"، مجموعه مقالات سومین کنفرانس مهندسی برق ایران، ص ۲۴۰-۲۳۱، دانشگاه علم و صنعت ایران، تهران، اردیبهشت ۱۳۷۴.
- ۸۴ فهیمی حمید، حاتم احمد، "ارائه یک روش ساختاری جهت تشخیص حروف در متون تایپ شده فارسی"، مجموعه مقالات سومین کنفرانس مهندسی برق ایران (کامپیوتر)، ص ۱۹۷-۱۹۰، دانشگاه علم و صنعت ایران، اردیبهشت ۱۳۷۴.
- ۸۵ کبیر احسان‌الله، بهاری کورش، احمدزاده محمدرضا، "بازشناسی متون تایپ شده فارسی"، مجموعه مقالات اولین کنفرانس مهندسی برق ایران، جلد دوم، ص ۲۹۴-۲۸۵، دانشگاه صنعتی امیرکبیر، تهران، اردیبهشت ۱۳۷۲.
- ۸۶ کبیر احسان‌الله، عزمی رضا، طرح ملی "بازشناسی متن فارسی"، گزارش شش ماهه اول و دوم، ۱۳۷۸.
- ۸۷ مرتضی‌پور حمیدرضا، "قطعه‌بندی بر خط کلمات دستنویس فارسی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: کبیر احسان‌الله، دانشکده مهندسی برق، دانشگاه تربیت‌مدرس، ۱۳۷۸.
- ۸۸ مسروری کیوان، "شناسایی برون خط کلمات دستنویس فارسی در یک مجموعه محدود"، پایان‌نامه دکتری، به راهنمایی: کبیر احسان‌الله، دانشکده فنی و مهندسی، دانشگاه تربیت‌مدرس، ۱۳۷۹.
- ۸۹ مسروری کیوان، "بازشناسی حروف دستنویس فارسی با روش فازی، پایان‌نامه کارشناسی ارشد، به راهنمایی: کبیر احسان‌الله، دانشکده فنی و مهندسی، دانشگاه تربیت‌مدرس، ۱۳۷۳.
- ۹۰ مقدم تبریزی کاوه، "تشخیص الگوی تصویری در محیط پیچیده"، پایان‌نامه کارشناسی ارشد، به راهنمایی مشیری بهزاد، دانشکده فنی، دانشگاه تهران، ۱۳۷۵.

۹۱ مقسمی حمیدرضا، "تشخیص حروف تایی فارسی با روش ساختاری"، پایان‌نامه کارشناسی ارشد، به راهنمایی: فتحی محمود، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، ۱۳۷۶.

۹۲ نحوی منوچهر، "استخراج کدپستی از آدرسهای تایی و دستنویس"، پایان‌نامه کارشناسی ارشد، به راهنمایی: کبیر احسان‌الله، دانشکده فنی و مهندسی، دانشگاه تربیت مدرس، ۱۳۷۶.

۹۳ نظام‌آبادی‌پور حسین، "پیش‌پردازش متون چاپی فارسی برای جداسازی حروف"، پایان‌نامه کارشناسی ارشد، به راهنمایی: کبیر احسان‌الله، دانشکده فنی و مهندسی، دانشگاه تربیت مدرس، ۱۳۷۹.

۹۴ نمازی مهدی، "شناسایی حروف تایی فارسی با قلمهای متفاوت به کمک شبکه عصبی فازی"، پایان‌نامه کارشناسی ارشد، به راهنمایی: فائز کریم، دانشکده مهندسی برق، دانشگاه صنعتی امیرکبیر، ۱۳۷۳.

ضمیمہ

معرفی شرکت Sakhr

و محصولات آن

شرکت عربی Sakhr، یک شرکت پیشگام در زمینه ارائه نرم افزار و آموزش به زبان عربی می باشد که موفقیت های چشمگیری را در عرصه تکنولوژی پردازش زبان عربی حاصل نموده است. محصولات این شرکت طیف وسیعی از کاربردها نظیر آنالیز مستندات، OCR، ترجمه ماشینی، تولید صوت ماشینی، تبدیل متن به گفتار، موتور جستجوی وب، سیستم های مدیریت اطلاعات در وب و موارد دیگر را در بر می گیرد. لذا در اینجا به معرفی این شرکت و محصولات آن می پردازیم.

1. From: <http://www.sakhr.com/>

1.1. Our Vision

Sakhr is the world leader in Arabic language processing technology. Our vision is to enable Arabic IT by providing our own world class solutions and by embedding our core technologies within global IT solutions in order to adapt them to the Arabic language.

1.2. About Us

Allow me to welcome you to Sakhr Software Web Site, which serves as an ideal gateway for easy and smooth exchange of information. This easy access to information will serve to turn our seemingly boundless world into a small village, where people will exchange services within a novel framework. This is a huge step in the advancement of humankind. No doubt, the use of the Internet is a great cultural leap, which has saved immeasurable human effort and facilitated the exchange of expertise. Though it is difficult to predict what form this new world will take after the introduction of the Internet, it will undoubtedly be a world brimming with excitement and benefits.

Sakhr Software has directed its every effort over the past twenty years towards enlisting computer and communication [technologies](#) to serve the unique needs of the Arabic language, and not vice versa. The Arabic language differs tremendously in terms of its characters, [morphology](#) and [diacritization](#) from other languages, and to claim otherwise would be a mistake. Furthermore, to import solutions from these other languages would only be at the expense of the unique features of the Arabic language. This was apparent to us from the start, so we undertook the development of the Automatic Morphologizer, the Automatic Diacritizer and the electronic dictionaries. Our ardent dedication to the research and development of these tools formed the cornerstone of all future [Sakhr products](#) and Internet [solutions](#), including [Automatic Speech Recognition](#), [Machine Translation](#) and [Electronic Publishing](#) solutions. Our investment in Natural Language Processing, or [NLP](#), has yielded outstanding results, and is expected to yield even greater results as the demand for Natural Language Processing in the exchange of information grows bigger. I am confident that Sakhr has placed the Arabic language on a par with other major languages in computer and communication field, becoming not only the major provider of Arabic software, but also the leading provider of Arabic language solutions for international companies in this field.

I hope our products and solutions satisfy your corporate needs.

Mohamed Al-Sharekh
Chairman of the Board

1.2.1. Company Profile

Founded in 1982 as an affiliate of al-Alamiah Group, Sakhr Software has developed cutting-edge [technologies](#) that have left a strong mark in the Arabic IT industry.

The company pioneered a new generation of Arabic Natural Language Processing ([NLP](#)) technologies at a time when international software developers were reluctant to sponsor any serious research or development to accommodate a difficult native language such as Arabic. At the same time, local software developers lacked the technological competence and the financial resources to undertake such an ambitious endeavor.

Sakhr alone had the resources and experience, as well as the determination and vision, to invest years of intensive research in the Arabic language and information systems.

The company's tradition of innovation led to the successful introduction of the world's first Arabic home computer, which was proof of the great demand for Arabic computing.

Maintaining this innovation till today, Sakhr has managed to become the leading developer of [Arabic solutions](#) & [products](#), and has overcome the dominance of English over the Arabic IT industry.

At the same time, Sakhr has corrected the misconception that western developed solutions can be adapted to adequately meet the needs of Arab users.

In its dedication to provide Arab users with state-of-the-art intelligent [solutions](#), Sakhr faced many difficult challenges that discouraged many other players, both international and local. These challenges included the lack of any governmental funding, which led Sakhr to embark on an ambitious research and development project involving over two million man-hours, entirely funded by private investment, and fueled only by its firm belief in the value of Arabic for its individual, [corporate](#), and [government](#) clients.

Sakhr has now become the leading [Arabic technologies](#) provider in [Speech](#), [Machine Translation](#), [Search](#), [OCR](#), [Publishing](#), [Education](#), and [Knowledge Management](#). Located in Egypt, Sakhr has around 300 IT employees, with offices all over the Middle East.

1.2.2. Customers

Sakhr recognizes that the integrity of its [solutions](#) and [services](#) stems from a combination of its in-depth understanding of the unique characteristics of the Arab market it serves, as well as the core [NLP technologies](#) embedded in its offerings. Sakhr is the ideal IT solution partner for those corporations that value experience, superior performance and quality service. Sakhr's expertise in different market sectors is illustrated in the partial customer list below:

Egypt

- | | |
|-----------------------------|-------------------------------|
| • Ahram Newspaper | • Ministry of Education |
| • Al-Azhar Al-Sharif | • Ministry of Electricity |
| • Egyptian Parliament | • Ministry of Foreign Affairs |
| • Majlis Al-Shourah | • Beit El Arab |
| • Ministry of Commerce | • Vision Plus |
| • Ministry of Foreign Trade | • Aswan Governorate |

Kuwait

- | | |
|------------------------------------|------------------------|
| • Amiri Diwan | • Ministry of Planning |
| • Kuwait Fund for Arab Development | • Al Arabi Magazine |
| • Kuwait University | • Zakat House |
| • Ministry of Information | |

Lebanon

- Al Mustaqbal Newspaper
- Annahar Newspaper
- Dar Al Sayyad
- Holy Book Society
- Intracom
- Lebanese Military Council
- Lebanese Parliament
- Lebanese University
- Presidential Palace
- Time Zero

Oman

- Diwan of Royal Court
- Ministry of Foreign Affairs
- Ministry of Education
- Ministry of Palace Affairs
- Ministry of Water Resources
- Oman Telecom
- Sultan Qaboos University

Saudi Arabia

- Aramco
- Islamic Development Bank
- King Abdul Aziz City of Science Technology
- King AbdulAziz Library
- King Faisal University
- Majlis Al-Shourah
- Ministry of Defense and Aviation
- Ministry of Education
- Ministry of Foreign Affairs
- Ministry of Interior
- Ministry of Justice
- Riyadh Colleges Administration
- Saudi Airlines
- Saudi Arabian Monetary Agency
- Okaz Newspaper
- Warrior from the Desert (Mokatel Min Al-Sahra)

Syria

- General Information Center
- Cadmus Press

U.A.E.

- Al Bayan Newspaper
- Abu Dhabi Chamber of Commerce
- Cultural Foundation
- Dubai Courts
- Etisalat
- Higher Colleges of Technology
- Ministry of Foreign Affairs
- Ministry of Interior
- Ministry of Labor
- Ministry of Planning
- Sheikh Zayed University
- Emirates Heritage Club

U.S.A.

- Defense Intelligence Agency
- Georgetown University
- Harvard University
- Princeton University
- Saudi Arabia Islamic Center
- Textron Systems
- U.S. Army
- U.S. Information Services
- United Nations
- University of California

Jordan

- Ministry of Interior

Morocco

- Ministry of Finance
- National Center for Documentation

International Companies and Organizations

- Acer Incorporation
- Brothers Industries
- Creative Networks
- Epson Corporation
- Global Language Solutions
- Hewlett Packard
- Intelligent Digital Machine
- International Business Machines
- Mendez (a division of Lernout & Hauspie)

- Ricoh International
- University of Kyoto, Japan
- University of Sophia, Japan
- University of Toronto, Canada
- University of Wurzburg, Germany
- Xerox Corporation
- Lucent Technologies

1.2.3. Partners

Al-Alamiah was established in 1970 in Kuwait as a distribution company for electronic and electric products. In 1983, Al-Alamiah has made its new start in the Arabic computer field. The educational field was the company's selection, and within a relatively short period of time Al-Alamiah has managed to Arabize the MSX computer system, which was a great success with more than 500.000 sales.

In 1989 Al-Alamiah launched its software industry programs for the PC. It begins first in Kuwait, then moved its development and research center to Cairo in 1990. Since then, the group has become the pioneer in the Arabic Software industry. It provides integrated [solutions](#), [products](#), and [services](#) for [multinationals](#) and [governments](#).

The group focuses its activities on the following levels:

- Arabic Research and Natural Language Processing
- Localization and Translation Systems
- Word Processing and Desktop Publishing
- Fonts technologies
- Educational Software
- Religious Software
- Arabic Communications
- Customized Solutions for big corporations
- Hardware and Networking.

1.3. Products

Sakhr products enable businesses to manage Arabic information, quickly [locate critical content](#), and accurately identify the [key concepts](#) required for information extraction, [categorization](#), and [summarization](#).

Based on years of research on [Natural Language Processing](#), our [technology](#) products give you the ability to properly manage your Arabic information to help you save time, increase productivity, and get more value out of content.

For example, Sakhr products enable you to:

- Scan an article of an Arabic newspaper, and transform it to an image.
- [OCR](#) the resulting image to transform it to text content.
- Pass the generated text to the automatic [corrector](#) to make sure that all known Arabic mistakes are corrected.
- Use Sakhr's [keyword extractor](#) to identify important concepts in the article.
- [Classify](#) the content under one or more categories in a user defined taxonomy
- Generate a [summary](#) of the article.
- [Archive](#) the image, the generated text, and the attributes in a safe document repository.
- [Search](#) for all articles in the archive that talk about similar content.
- [Publish](#) the article content with a photo to an Internet/Intranet website.
- Allow site visitors to view the [English translation](#) of the article.

- Have a site visitor request through a [voice command](#) to listen to the article, and our TTS to [read out loud](#) the article through the speakers.

1.3.1. ArabDox

ArabDox is an Arabic/English/French Document Management System that offers enterprises an integrated solution to manage increasing amounts of non-structured information in documents.

The process begins by capturing information from all sources (papers, microfilms, faxes, e-mails, text files, HTML documents, office documents, etc.) It then describes this information with database attributes, and finally makes it readily available and searchable through a Web browser.

Paper-based documents are scanned to convert them into electronic images. Then, they are recognized using our Arabic OCR technology to extract the text into database fields. [Summary](#), and [keywords](#) are extracted from this text and stored with the original images in a compressed format. Using a Web browser, an authorized user can [search](#) this store and view documents online. This integration of Sakhr's own [OCR](#), [Keyword Extractor](#), [Summarizer](#), [Idrisi](#), Client/Server and Internet/Intranet technologies gives ArabDox its uniqueness and power, making it "The Arabic Document Management System."

We spend time creating and updating information and trying to understand it. Without information, decisions are simply random guesses, but when the right information is available to the right people at the right time, it can create opportunities for success. When people can find information quickly and easily, they are able to spend their time working effectively rather than trying to locate what they need. When people can rely on their information sources to be accurate and up to date, they are able to make decisions with confidence, rather than worrying about having to double or triple check each factor. They are also able to focus their attention on producing high quality information, rather than wasting their time worrying about distribution and security.

In short, immediate, easily secured access to up-to-date information is critical to an organization's success and its bottom line. We have a system that simplifies access management and controls all document types throughout any organization.

1.3.2. NashrNet

Sakhr content management solution, NashrNet, is a bilingual web-based content management and publishing solution. It handles the complete cycle of the content starting from the content gathering, creation, processing, to content archiving and publishing to different electronic endpoints. Being completely Web-based, NashrNet allows any authorized user from any place around the World to create, edit and distribute Web content directly via a Web browser, with no need to install any software, or to be familiar with Web authoring or Web designing/developing languages. An Internet connection, anywhere on the web, will provide the content editor access to the content he needs to update at any time.

NashrNet Administration:

The system administrator creates users and groups, he also:

- Defines the work area in which he wants to enable the users to work.
- Creates the authority for every user and group to deal with content in channels, workflows, and stages.
- Creates Channels and workflows.
- Creates search filters, and many other tasks.

NasherNet Integration:

NasherNet is customizable, so you can easily modify it to suit your preferences and special needs. You can even integrate NasherNet with other Sakhr tools and solutions for content gathering, data and documents processing and management, such as:

- [Johaina](#): for newsgathering.
- [OCR](#): for optical character recognition
- [ArabDox](#): for Arabic document management
- [Corrector](#): to automatically detect and correct common Arabic mistakes
- [Keyword Extractor](#): to extract main keywords or phrases
- [Categorizer](#): To classify content in a hierarchical tree
- [Summarizer](#): to automatically generate a summary

NasherNet handles the content throughout the following stages:

Gathering:

NasherNet provides different plug-ins to automatically receive the online dynamic content and direct it to the system repository. Content arrive from different sources, such as:

- Content contributors inside the company
- Remote authors and reporters
- News agencies
- Dynamic content providers.

Creation:

NasherNet provides a user-friendly bilingual Web-based interface for content contributors inside and outside the company to be able to create their own contents. Users can easily create and edit their articles and media using a simple HTML editor.

Storage and Management

Content is stored in a central database (repository), the system supports MS SQL and Oracle. Content is assigned editorial workflows for publishing, archiving or deletion:

- The system arranges content in channels, where any content source can be easily auto-linked to channels.
- Channel content is assigned certain editorial workflows, where the system administrator defines the workflow for the content in every channel, and assigns users to every stage in the workflow.

Distribution:

Approved content can be distributed to different electronic endpoints, such as:

- News portals.
- Websites.
- Periodicals Websites: NasherNet creates, publishes, and manages daily, weekly, and monthly newspapers and magazines websites.
- Electronic newsletters: the system arranges content and sends it automatically to registered users emails in the form of electronic newsletters.
- Electronic reports: the system arranges content and sends it automatically to certain users in the form of electronic reports.
- SMS messages: the system arranges content and sends it automatically as SMS messages.
- WAP: the system arranges content to be viewed using WAP devices.

Analysis:

- NasherNet has an analysis mechanism that does:
- System user reports: user access time, user exit time, user idle periods, number of transactions for every user, kind of transactions...etc.
- Content reports: content sources, history of every content item...etc.

Search:

To help users create new content, NasherNet provides search to the content in the repository. NasherNet also provides advanced search on the content after publishing.

- The system administrator can create advanced and complicated search templates (filters), and make the filters available for all users to search in the content without having to form the same complicated query every time.
- The user can also create custom search filters and share them with other users.
- Sakhr bilingual search engine [Idrisi](#) applies its unique and advanced free text search techniques on the published content. Moreover, NasherNet provides other useful features such as:
- Storing materials in a hierarchical tree of channels (folders) and providing an authentication definition per channel.
- Providing the ability to define the workflow per channel; i.e., the content must follow such a flow until the last stage before it is ready for publishing.
- Providing a technique to define periodical activities (deleting, archiving, etc.)
- Mapping the material to a system channel, and automatically grabbing this material into these channels and giving it an initial categorization (defined per channel).
- Giving the system user private access to his material through the ability to Check Out/In such material.
- Enabling users to modify any material or relate materials to each other.
- Providing a notification engine to transfer information among system users.
- Providing users with reminders to alert them on time.
- Providing a thesaurus system to help in decreasing data redundancy and creating the material easily.
- Having capability of adding new Add-ons to the system to help gathering materials from different news sources.

1.3.3. Idrisi

Intelligent Document Retrieval and Information Search for Internet (Idrisi) is a bilingual search engine for documents, Corporate databases (ODBC compliant), and Websites.

Idrisi search engine is the outcome of Sakhr ongoing research in the fields of Arabic [Natural Language Processing](#) and search and retrieval technologies.

Idrisi pioneers the search in Arabic language, and leads the search in English language. It adds powerful [morphological](#) search and retrieval capabilities to corporate databases, Web sites, and documents through a simple Web-based bilingual interface.

Idrisi is notable for its unrivaled easy-to-use administration modules. On the other hand, the search and retrieval processes combine power, simplicity, accuracy, and speed.

Using Idrisi with XML support, users can easily search for specific information by choosing to search within the content of a specific element. Developers can also build their own applications using the Idrisi COM Object interface.

Idrisi provides developers with a complete standard Software Development Kit (SDK) that enables the smooth and effortless integration of their applications with Idrisi.

Idrisi fully supports indexing, publishing and searching in databases regardless of their original format and platforms, through integration with ODBC drivers.

Idrisi Features:

- Integrates with INSO filters that support all Microsoft Office formats and other 225 standard formats such as HTML, TXT, RTF, and PDF.
- Supports multiple code pages, either from different platforms (such as Macintosh) or from different languages (such as French).
- Complies with Windows NT security schemes, with control of access to data.
- Capable of handling, indexing, and updating data automatically allowing end-users to get the latest and most accurate information without manually rebuilding and re-indexing collections.
- A fully bilingual search engine, supporting both Arabic and English features.
- Provides users with customizable search and result templates. These templates cover all the options that an end user may need, about the number of hits, from simple or compact results to advanced detailed information.

1.3.4. Johaina

[Johaina](#) is a news monitoring service that scans hundreds of Arabic, English and French news sites targeting the Middle East region. [Johaina](#) operates around the clock to scan its resourceful directory of the most frequently visited online newspapers, press sites, and news portals. It collects, categorizes, classifies, archives, and displays news from an extensive range of around 400 reliable sources and offers news seekers a one-stop newsstand to look for news.

Johaina Backbones:

[Johaina](#) is empowered by Sakhr's renowned search engine [Idrisi](#) and draws heavily on its powerful support for the Arabic [Categorizer](#) and [Summarizer](#), which are among other tools that give Johaina exactly what it takes to quickly monitor a host of sites, track the most recent news and classify them in an easily navigable directory.

Johaina's visitors can navigate its directory of news sources to have a feel for its main and sub-directories and reach headlines. They can, more conveniently, submit a search to directly locate the news they want. Simple and advanced levels of search are available to customize search. Also available is a conceptual search feature to enhance the search process so that Johaina offers to search for information within the larger scope of the whole subject without being restricted to search words.

Johaina Types of Service:

As well as providing -Ajeeb's- visitors with its free services, Johaina offers a four-level subscription service for those news enthusiasts who are interested in making the most of its services.

Johaina Navigation Services:

A Navigation Service that grants full access to Johaina's directory and offers an archive in which it stores cached copies of news articles for later reference; even if the article has been replaced by its host site.

Johaina Monitoring Services:

A Monitoring Service that offers subscribers the capability of creating a list of keywords, each comprising accurate search criteria. Johaina would use those keywords to make a daily search for topics and keep the results on personalized archives for subscribers. Translation of headlines and addition of extra sites to the scope of Johaina's search are among other features.

Johaina VIP Services:

A VIP Service that combines the features of both Navigation and Monitoring Services.

Johaina Channel Services:

A Channel Service that enables news display in the subscriber Web site or portal. The Channel Service is entitled for Subscriber aiming at increasing the frequency and duration of their site visits, by adding valuable and dynamic source of news to their Web Site. Johaina Channel Service provides Subscriber's employees with frequently updated news across his own Intranet.

1.3.5. OCR

OCR stands for Optical Character Recognition. When a text document is scanned, the computer recognizes this text as a graphical image. The user cannot manipulate, search, or edit the image text in its image format. An OCR program reads this scanned text, recognizes it, and then converts the figures and characters into a text document.

The OCR, backed by Sakhr's experience in Natural Language Processing [NLP](#), can solve the Arabic language unique characteristics.

To begin with, Arabic is written cursively, where several characters are connected to form what we call 'blocks of characters'. Arabic can also be written in many fonts, so that a 'block of characters' has more than one base line. Additionally, Arabic uses many types of external objects such as dots, 'Hamza' and 'Madda'.

[Diacritization](#) adds a new set of external objects. Furthermore, Arabic characters can have more than one shape according to their position inside the block of characters (initial, middle, final or standalone block of characters).

Overlapping also makes it difficult to determine the spacing between blocks of characters and words. Finally, Arabic font suppliers do not follow a common standard. Given the peculiarities of Arabic fonts and the characteristics of the Arabic language, building an OMNI OCR becomes a difficult undertaking.

Sakhr OCR combines two main technologies: Omni Technology, which depends on highly advanced research in artificial intelligence, and the second is Training Technology, which increases the accuracy of character recognition.

It can identify more than one language through Xerox Text Bridge technology; one of the most popular OCR programs. The program can also identify both Arabic and English characters on the same page. Moreover, OCR can distinguish between 26 Arabic true type fonts. It enables the user to scan large amounts of documents, save them as graphic files and classify them for later recognition in order to save time.

The program recognizes graphics and places them in their proper location on the page. The program also saves page format without any modification to tables, columns, or graphics. It can also identify diacritics and keep or remove them according to the user's preferences.

The output can be saved to disk for use in a variety of applications such as word processors. The program is compatible with many other programs, and OCR SDK (Software Development Kit) is available to allow integration with other applications. The OCR SDK is available in both DLL and ActiveX formats.

Sakhr OCR engine has been optimized using Intel tools to better run on the latest Pentium IV processors that support Hyper-Threading.

1.3.6. E-Portal

Sakhr's extensive background in the educational and [Arabic technologies](#) field had produced Sakhr's E-Portal solution that answers the FIVE main objectives of the education sector:

- Self-paced learning management system (LMS).
- School Management System (SMS).
- Internet Educational Portal (IEP).
- Virtual Classroom (Synchronous Learning).
- Content Composition tools.

Sakhr provides the educational sector with a complete E-Portal solution to help in improving education methodology especially in the following areas:

- Learning-assistant software to help students in the self-study
- Educational content with content creation tools
- Learning management software to systemize the self-paced learning process
- Follow-up and reporting system to monitor the progress of student during his self-study
- Software to facilitate the interaction between teachers and their students beyond the physical class boundaries
- Software to manage the education methodology and academic progress of students
- Software to manage the other activities inside schools
- Data mining and decision support and reporting tools to help top level executives make decisions based on the periodically collected information from bottom levels up
- Software to provide the environment for all parties in the education methodology to communicate and exchange information, and [news](#)
- Software to provide better education-related services for individuals that need to interact with the system not only through a Web-based interface where they can [search](#), and browse information, but also through the [phone](#) for those who are not Web users.

1.3.7. Books Publisher

Sakhr Books Publishing Solution is an integrated system that manages the complete cycle of entering, processing and publishing electronic and paper books on the Internet and the Intranet.

It is an easy to use system, working both on a single-user basis and multi-user basis. The system administrator can assign and manage content workflow and users access rights.

Components:

The system consists of 2 major components:

- Data processing: a cycle that handles all the data processing steps from data entry to publishing on the Internet. This cycle consists of the following workstations:
 - a. Scanning
 - b. OCRing
 - c. Proofreading and auto-correction
 - d. Indexing and BookML tagging
 - e. Release of entered materials.
- Books online administration: Web-based application that takes input from the data processing cycle and controls books publishing and administration. It also remotely controls the display and administration of the electronic publishing services that are included in the system. Using this Web-based application, the system administrator can do the following tasks remotely:

- a. Create and modify online libraries.
- b. Create and modify categories, books, parts and chapters.
- c. Add publishing houses, authors, and hotspots.
- d. Publish books to different hosting servers using a collection of ready-made templates.
- e. Index books contents for [search](#).
- f. Modify published pages.
- g. Control online publishing services.

Main Features:

- Books contents follow a unified workflow from entering to release
- User authentication required for every stage in the workflow
- Easy to manage remotely (Web-based system)
- No need for Web development (ready-made templates for publishing)
- No need for programming know-how (Easy Interface)
- Saves time and efforts (fast data processing & publishing)
- Publishing in BookML standard format
- Special tagging for places, persons, events, Quran verses, Hadith...etc
- Serves multiple-libraries with multiple branches
- Advanced search capabilities
- Accompanied with a big set of [Electronic Publishing services](#).

Publishing Services:

The system provides a collection of electronic publishing services:

- Website design: in addition to the library's homepage, the system provides a collection of ready-made templates to allow the site administrator select from for publishing different books.
- Navigation: provides access to published books using different ways such as categories tree, books names, publishing houses, authors and book cover thumbnails.
- Search: Sakhr search engine "[Idrisi XML](#)" empowers the system with advanced morphological search capabilities.
- Informative services: pages to display information about publishing houses, books, authors, translators, editors, and distributors.
- Interactive services: discussion forums, voting, books rating, print page, send by email and other services.

1.3.8. SET

Sakhr Enterprise Translation (SET) system is a web-based [Machine Translation](#) solution from English to Arabic and vice versa. This system targets large companies and institutions that have a continuous flow of translation work.

You can use SET to automatically translate documents and web pages. Then you can edit the translated sentences before saving them in a database for future reuse. Moreover, you can create glossaries containing your own terminology, hence saving time and effort, ensuring consistency, and enhancing the translation quality throughout the process. SET Features:

- Automatic translation from English<=>Arabic.

- Ability to save your translated sentences in a database called translation memory for future reuse. This will enhance the translation quality especially when you translate files in the same domain.
- Ability to add your own glossaries to enhance and customize translation output.
- High-speed translation for common file formats: html, rtf, and txt.
- Web pages translation.
- Lookup for meanings of words in a Bi-directional Dictionary, Arabic<->English.
- Integrated bilingual spell checker to [correct spelling](#) mistakes before translation.
- Administrator has full control on the account and can define access rights for each user.
- Ability to submit any document for high-quality human translation. Our professionals translate this document in a timely manner at a competitive rate.
- Easy access from anywhere, just type your login name and password.
- Simple, user-friendly web interface with a detailed help section.
- Benefit from daily updates and ongoing developments on our engines and glossaries.

Keeps you updated with up-to-the-minute information and knowledge. It helps your organization to streamline its translation process by utilizing state-of-the-art technologies and tools.

1.3.9. Ibsar

Ibsar (Vision in English) is an integrated solution for visually impaired and blind users. It is based on Sakhr Arabic/English Text to Speech [TTS](#) and Sakhr Optical Character Recognition [OCR](#) engines. Ibsar enables blind users to read printed books and documents as well as electronic files on their own, without any help. It also enables them to write texts in both Arabic and English, in addition to saving and printing these texts in Braille.

This integrated solution enables the blind user to access the Internet and read Arabic and English Web sites using a friendly easy-to-use method designed for visually impaired and blind people. Moreover, Ibsar allows users to send, receive, write, read, and manage E-mail messages, helping users reach out for a wider world. Ibsar provides a tutorial system to train novice users to use the normal keyboard. It guides them step-by-step till they master the use of the keyboard to fully manage Ibsar.

At the end, Ibsar is a powerful Screen Reader that converts the output of the screen into high quality human voice so that the users can professionally use most of the software programs in Arabic and English. For example, users can navigate their mobile phones, send and receive SMS messages, have them read, and add or delete data to the phonebook in Arabic and English.

Ibsar system benefits:

- Allows the users to perform self-training.
- Maintains the users independence and privacy.
- Provides assistance for blind and visually impaired users in terms of education, culture, and career development.
- Supports any standard computer so it can be used at home, university, office as it runs on any Windows operating system.

1.3.10. TTS

Sakhr Text-To-Speech (TTS) engine converts any Arabic/English text into a human voice. Sakhr has been focusing in the last 5 years on creating an Arabic TTS engine that can match in its quality the human voice. This technology gives businesses a competitive edge by allowing them to provide their customers with the latest static and dynamic information anytime, anywhere using normal telephones and mobiles.

How a TTS can help businesses?

The huge quantity of dynamic data that people need to check, on the move, about their daily activities as news, sports, weather, stock options, bank accounts, E-mails etc., is on the rise. It thus needs a powerful TTS engine that converts them into human voice in real time, and at the same time to be able to handle the tremendous load of telephone calls using the latest techniques in load balancing and streaming.

How a TTS technology can save money and time?

The TTS technology is becoming inevitable in some businesses that need to feed their customers with the latest and vital information in real time. These businesses usually use IVR (Interactive Voice Response) systems and call centers to communicate this information to their customers and prospects. Converting vital data stored in Web sites, databases, and files into human voice using the traditional expensive and time-consuming human recordings in studios is a hard and long process, since the information is usually dynamic. In some cases, it would be impossible to track these changes using human recordings methods. Only a powerful TTS engine that supports Arabic and English languages can be the right solution for the information stored in IVR systems and call centers.

Why an Arabic TTS is different?

Arabic is a difficult language and it is not like languages as English, French, or Spanish. Those languages, written in Latin alphabet, have vowels while the Arabic language has special characters called "[diacritics](#)". These diacritics give the Arabic words the correct meaning inside a sentence. For example, two Arabic words that have different meanings are written the same and only the diacritics help the reader to distinguish between them. Examples of diacritics are: For this reason, Sakhr has developed the [Diacritizer](#) engine.

his engine can insert the diacritics needed in Arabic texts automatically. It is the main component of Arabic TTS. Without the Diacritizer, the output quality of TTS engine would be inaccurate and not clear. Since Arabic native speakers write Arabic text without diacritics, the TTS engine should handle the non-Diacritized text. The [Diacritizer](#) will convert non-Diacritized text into a Diacritized text and then the TTS engine will convert it to a clear human Arabic voice.

1.3.11. ASR

The market's globalization, the wide use of the mobile phones, and the complexity of the non-Latin languages such as Arabic, make the ASR (Automatic Speech Recognition) technology an important component of any telephony application or a call center.

Speaker independence is the most critical issue of the Arabic ASR engine. In addition to handling speech recognition in noisy environments, Sakhr's Arabic ASR engine supports keywords spotting allowing it to understand specific words within a spoken speech. This feature is very useful when dealing with people who are new to the system and don't know the right utterance to say within a context. The engine supports many Arabic accents including the Levant, Gulf, and Egyptian accents.

Sakhr has developed a Large Automatic Speech Recognition (LASR) engine to use in Directory Assistance DA applications, and Yellow/White pages services. The LASR handles huge databases with millions of Arabic full names distributed over blocks with a hundred thousand names in each single block.

1.4. Services

Sakhr provides [translation](#) and [localization](#) services that enable [multinational](#) companies to adapt their products for the Arabic market and integrate these products with existing local environments. That helps multinational corporations and contractors implement large-scale solutions that target the Arab world.

For local [governmental](#) offices and [publishing](#) houses, Sakhr provides consulting and management services in the field of Internet/Intranet [content gathering](#), workflow and diffusion.

1.4.1. Translation

Sakhr established its Translation Unit in the heart of the Middle East to help international companies reengineer and adapt their products to the Arabic-speaking communities. Backed by Sakhr's outstanding position in the Arab IT industry for over 18 years, only Sakhr Translation and [Localization](#) Services can take your business into the world's fastest growing markets.

At Sakhr Translation Unit powered by our premier translation technologies, the commitment is to make sure that your Arabized products will communicate in Arabic as fluently as we do. At Sakhr Translation Unit, we employ successful translation management model that adapts to the specific needs of our clients. Our project management team uses the translation process that takes care of your project's intricate specifications, timeframe and cost estimations.

In contrast to Latin languages that read from left to right, Arabic text reads from right to left. This peculiarity renders Arabization a complex process that requires professional staff to complete. Our detail-conscious translators and language editors will make sure that your software, web page or documentation are in the proper linguistic and cultural format for the Arabic market.

We have successfully completed tens of translation projects as we provided the translation our clients need to successfully develop and market their products anywhere in the Arab world. Sakhr Translation Unit has proven to be professional, reliable, and successful in bulky translation projects for Sony Gulf, Intel, Lucent, Mendez, Canon, Ricoh, Lexmark, Raytheon, Compaq, Open Text, and others.

1.4.2. Localization

Localization is the process of adapting hardware, operating systems or software applications for a particular market or region. In other words, it is not mere translation, It also involves the production of a localized version of the product for a target audience to speak their language and meet their needs. This process implies the localization of user interface, user's guides, and online help.

As the world becomes a global community, any company that tends to expand its business to new markets must know how to properly approach these markets. Sakhr helps you introduce your product to the Arab market, making your products and services truly effective, market-tailored and successful. Sakhr relies on its long experience in Arabic IT market, efficient [translation](#), and localization tools, in addition to the highly qualified team of translators and localizers.

Sakhr has a very long and successful track record in this area starting with the first Arabic enabled PC running under DOS, then the first Arabic Windows OS, then the first Arabic web browser Sindbad, which is a complete localization of the Netscape Communicator package. Mitsubishi, LG, Ricoh, and Hand Spring are currently relying on Sakhr's localization

experience to offer their office equipment and mobile devices to a huge Arabic speaking market segment that demands Arabic interface.

1.4.3. Publishing

With our unique art and design concepts, we design distinctive Internet sites that will make your customers come back. Your page design is based on your company's image and in line with your corporate specifications. It can be as simple or as sophisticated, as you require, it can be also monolingual, bilingual, or multilingual.

Web Development services at Sakhr includes studying, designing and building Websites for customers. They also include graphic design, structural design, data entry (processing and [correction](#)), informative services pages, [dynamic content display](#) and interactive services, in addition to the advanced morphological free search capabilities of Sakhr [Idrisi](#) search engine.

Our Web development services may include also some additional optional services such as [translation](#), domain registration, Website hosting, administration, maintenance and updates in addition to [Johaina](#) news channels.

Electronic Publishing Services (CD Production):

Sakhr uses its own Electronic Publisher solution for publishing information such as HTML pages stored in either documents or databases to CDs, which allows the information to be searched and displayed using any Internet browser.

The program's capabilities are not limited to the publishing process, but it also includes additional features such as search capabilities through Sakhr's [search engine](#), provided with each copy of the published product.

The Electronic Publisher offers full support for indexing and retrieving information from databases through ODBC drivers. It is a Web-based application that utilizes the client-server strategy, which allows the user to administer the Electronic Publisher remotely and locally at the same time. It provides users with a variety of bilingual search capabilities allowing them to quickly retrieve precise information found on database tables or documents.

[Click here to view some Web Development projects](#)

[Click here for more information about Sakhr Web Development Services](#)

1.5. Solutions

Companies can employ Sakhr modular [products](#) and [services](#) to create custom solutions for unique business problems, help introduce new services, and capitalize on opportunities. In today's competitive global economy, the optimization of your organization's resources is critical. Whether you are:

- An [Arabic Governmental Organization](#) looking to automate your internal processes,
 - An [Arabic Educational Institute](#) trying to make the move to the e-learning environment,
 - A [Multinational Company](#) expanding in the Arab market,
 - An [Internet Service Provider](#) competing in a difficult market segment, and looking for a special extra service to provide to your subscribers,
 - A [System Integrator](#) providing Arabic enabled Solutions to your customers,
 - A [Call Center](#) Manager requiring the advanced Speech capabilities to cut down on your operational cost,
 - Or a [Publisher](#) who is offering web-based services,
- Sakhr has the solution for all your business problems.

1.5.1. Government

Sakhr provides Arabic E-Gov portals with the following 4 features that are considered a MUST for a government to better service its citizens, businesses, and employees:

1. **Search**: One of the most requested features in a portal is the search, especially in a Governmental portal that acts as the MAIN Gateway to ALL governmental services. The homepage should display a small text box in a very visible area, where the visitor enters his/her query and a Search button to submit his query.

Almost 85% of all information needed by a citizen is found in non-structured information repositories like web pages, office documents (Word, Excel, Power Point, and Emails), text files, PDF files, etc.

Accessing this information through browsing only or through database search techniques takes a lot of impractical time. So, a government portal should employ a Free Text Search engine to answer queries in less than three seconds. Yet, since this information is bilingual (Arabic/English) and mostly Arabic, in Arab Governments, you need a Search Engine that understands the language, and can relate different forms of the same word to each other for example:

أحمد و احمد، ساره و ساره، الحریری و حریری، علم و عالم، عبدالله و عبد الله، مستثمر و إستثمار

As Arabic is a highly inflectional language, exact match searches do not produce accurate results.

2. **Speech (Phone Access)**: Not all Citizens are web citizens capable of using a web browser or even have Internet connection, especially in an Arab environment. So, Arab Governments that offer their information and services to all their citizens must provide not only a web interface, but also a phone interface for the same information and services, so that citizens can access these services and information through a land or mobile line.

Government phone services should be automated to increase availability, and serve more citizens concurrently. This is done by employing an IVR (Interactive Voice Response) system, allowing a machine (computer server) to interact with a caller, by automatically receiving a call, playing recorded messages, taking input by asking the caller to press some phone buttons, and replying to the caller through other recorder messages. Yet, in real time, not all information given to a caller is in static form, and recorded before hand. So, a TTS (Text To Speech) engine should be employed to read the caller dynamic information on the phone (both Arabic and English text depending on the caller-selected language).

An ASR (Automatic Speech Recognition) engine should be employed to recognize caller voice commands. The ASR is mandatory when the choices are more than 4 or 5, as you can't ask the caller, for example, to press 1 for Jeddah, 2 for Cairo, 3 for Beirut... and 215 for Kuwait. But you can ask him to just say the name of the city, and the ASR would recognize this input from a long list of cities.

3. **MT (Machine Translation)**: Citizens or residents come from different backgrounds and speak different languages mainly Arabic and English, but not all information found in government offices is bilingual, so you need to be able to provide this information to both Arabic and English speaking citizens in their preferred language, no matter what the information source language is.

Yet, due to the size of this information, its manual translation becomes an overwhelming and a very costly task, especially when the nature of this information is dynamic.

An [MT](#) (Machine Translation) Engine should be employed to translate on the fly this information from English to Arabic or vice-versa. The accuracy of the translation is not perfect, but it can be improved tremendously when incorporated or humanly assisted with a specific list of glossaries of terms or translation memories.

4. [OCR \(Optical Character Recognition\)](#): For the above 3 features to be offered to citizens, the information itself must be in electronic format, so that it can be processed. However, most of the information is in paper format, and should be scanned through hi-resolution scanners to be converted into electronic images that can be stored on computers.

Yet, even if an image contains text, for the computer, it is just an image.

An [OCR](#) (Optical Character Recognition) Engine should be employed to optically analyze and recognize characters in this image, converting it into text. This engine should recognize not only English text, where each character is separated from the other, but also Arabic text with linked and diacritized letters as follows:

نَصٌّ عَرَبِيٌّ مُشَكَّلٌ بِأَخْرُفٍ مُتَّصِلَةٍ

The OCR saves the government the cost of storing image files, that is at least 10 times the cost of storing text files, and enables it to process this text by [translation](#), and/or [Conversion to speech](#) (none of which can be applied on images).

1.5.2. Education

Historical Background:

During the 1980's and 1990's, Sakhr had set its mark in the computer-based Arabic education sector very strongly. This was through four factors:

- Providing many edutainment titles on CDs, both developed in-house, and licensed from international companies and localized in Sakhr.
- Developing educational material for various educational subjects, levels, and years. Sakhr produced its courseware on CDs for use on home computers.
- Developing computer software to manage all administrative, financial, and academic activities inside a school. Sakhr has designed the system to manage individual schools.
- Producing many software titles that help students in the educational process such as: electronic encyclopedias, electronic dictionaries, and tools for equations and maps drawing.

New Sakhr Strategy Online with the Newest Industry Trends:

The trend Sakhr took in the last two decades has started to decline during the last few years because of many reasons:

- The market of educational titles, courseware, and management software has become very competitive, and piracy has affected the market drastically.
- The Internet has changed the way people think and work daily with computers, including the educational activities. Therefore, using CDs for educational software is not anymore the most convenient way.

- Education authorities in the Arab region have changed, and became more oriented towards collective learning rather than individual learning, incorporating in the learning process advanced technologies, and tools such as: Learning Management Systems, Internet Based Courseware, and Virtual Classrooms.
- The objective of using school management systems has changed from managing schools individually, to managing a group of schools collectively, to support top executives in decision making and planning, in addition to managing the daily activities inside each school.

Sakhr E-Portal:

Sakhr's extensive background in the educational and [Arabic technologies](#) fields has produced Sakhr's [E-Portal](#) solution that answers the main objectives of the education sector.

1.5.3. Multinationals

Multinational companies scramble just to keep up with the explosive pace of innovation and take advantage of the international demand for their products. Factors driving this evolution include merging technologies, short product life cycles, and competition in this global market. So, building long-term profitability is a key, and successful multinational companies must seek out ways to control costs and rationalize processes, while matching the speed of their competitors.

Sakhr is Your Globalization Partner:

By outsourcing your globalization efforts to Sakhr, we can help transform these challenges into competitive advantages. Putting our expertise in product, [localization](#) solves the problems facing your company.

Sakhr has years of experience serving major [market players](#) in [translation](#), and localization processes. Our services include the following features:

- **Planning and Analysis:** We identify and analyze local market requirements, and develop a comprehensive project plan that includes [localization](#), [translation](#), and [integration services](#). This plan is then executed for the life of the project till we deliver the completed product, and supporting material.
- **Technical Writing:** We work with you as an integral part of your product development program to ensure accurate, on-time delivery of your documentation.
- **Testing:** We recognize the importance of global quality assurance, so we provide our customers with a complete QA service designed to ensure product consistency.

1.5.4. Publishers

Sakhr designs and implements solutions to help publishers and periodicals efficiently manage and publish their information. Sakhr builds solutions that enable our customers to author, store, and manage information in a central data repository, publish this information in multiple formats, share content across publications, and import and export data to and from internal business systems.

We begin by developing a clear understanding of your goals, requirements, and business issues, and then recommend, design, and implement an effective solution for your needs. Our solutions include:

- Custom, [component-based](#) development.
- Integration of best-of-breed [software applications](#).

- Value-added data conversion and integration [services](#).

The result is a flexible, scalable, open-ended customized solution that effectively meets your needs, wholly integrated within your business, and produces significant return on investment.

Sakhr provides publishing solutions in two main categories:

- [Content Management Solutions](#): Solutions for publishers of newspapers, magazines, newsletters, or corporate information.
- [Books Publishing Solutions](#): Solutions for books publishers who need to automate their publishing, editorial, and production processes.

Our extensive resources, experience, and expertise, combined with our research in Arabic language [technologies](#), make Sakhr a powerful ally in your quest to gain a competitive advantage in today's business world.

1.5.5. Call Centers

Sakhr offers the most pleasant Arabic Text-to-Speech ([TTS](#)) solution in the market. Our TTS is able to read, understand, and convert non-[Diacritized](#) Arabic text into the most human-sounding male and female voices available. The high quality and intelligibility of Sakhr TTS enable businesses to deploy highly-interactive and user-friendly applications, enhancing the user's experience and improving customer relationships. Sakhr TTS together with Sakhr Automatic Speech Recognition ([ASR](#)) engine are used to speech-enable a wide range of call center applications, including unified messaging systems, interactive voice response (IVR) applications, and voice portals.

Sakhr also offers its Large Automatic Speech Recognition ([LASR](#)) engine that handles huge databases of millions of Arabic full names in Directory Assistance (DA) applications, and Yellow/White pages services.

1.5.6. ISP's

Sakhr offers ISPs a complete starter kit designed to grant their Arabic subscribers easy access to the Internet. The ISP kit handles the user's Internet configuration details. It checks the system for the components needed to connect to the Internet as Dial-up Adapter, Dial-up Networking, modem and TCP/IP protocol, Proxy server, and security settings, and prepares them automatically. It also helps users set up their e-mail accounts for major email packages, using a simple bilingual Arabic/English interface.

This includes an On-Point multilingual Dictionary (Arabic, English, French) that provides words definitions while surfing the Web, [Machine Translation](#) button connecting to a bi-directional translation site that translates Arabic/English Web pages, and Arabic Search button connecting to a search engine to retrieve Arabic content on the Web.

Sakhr can customize the ISP kit in accordance with the ISP requirements, allowing distribution of the kit through CDs or through online download from the ISP web site.

ISPs can work with Sakhr to offer their subscribers online services like: [Johaina](#), [Talking browsers](#), and web site [translations](#).

1.5.7. System Integrators

Sakhr offers System Integrators the possibility to enable their application users to transfer images from their Desktop to a text document, by simply using Sakhr's Optical Character Recognition ([OCR](#)) Software Development Kit (SDK).

System integrators can use Sakhr [Corrector](#) SDK to eliminate all common Arabic Mistakes from their text documents before doing further processing like: [Keyword Extraction](#), [Classification](#), [Summarization](#), and/or [Translation](#), using the corresponding Sakhr SDKs in each of these text analysis [NLP](#) areas.

To search for Arabic text in huge repositories, Sakhr provides system integrators with [Idrisi](#) free text search engine SDK, that is currently serving the full archive of millions and millions of news articles in [Johaina](#) online news collection service.

For Speech Applications developers, Sakhr offers both an Automatic Speech Recognition ([ASR](#)) SDK, and a Text To Speech ([TTS](#)) SDK, providing highly intelligible and superior voice quality, fully customizable commands, volume, speech-rate, spelling prosody, vocabulary, and integration capabilities, thanks to its standard Application Programming Interface (API).

1.6. Technology

The Arabic language is unique in some of its features like the extensive use of [diacritics](#) replacing most vowel letters, the always-cursive way of its written text, and its highly inflectional nature.

Sakhr has done basic research on the core linguistic characteristics of the Arabic language, and [natural language processing](#), employing more than 100 IT professionals for more than 10 years dedicated to produce Arabic Linguistic Components that perfectly understand the language. These components work on all language levels from the character level, to the document level passing through the word, sentence, and paragraph levels.

[Machine Translator](#) from Arabic to English & vice-versa, [Optical Character Recognizer](#), Automatic [Diacritizer](#), [Summarizer](#), [Categorizer](#), [Corrector](#), [Speech Recognizer](#), and [Text To Speech](#) Synthesizer are some of the patented (or patent pending) technologies at Sakhr that power all of our products and solutions.

1.6.1. NLP

With the information revolution in full swing, electronic documents have become a principal medium of information. Businesses and governments alike are producing electronic documents at an increasingly rapid pace.

While numerical information was the dominant data type in the 1980's, representing almost 66% of the whole information bulk, textual information now constitutes around 70% of that bulk. Consequently, as the flow of information increases, linguistics is set to become the basic ingredient for software development.

Sakhr recognized this fact early on, as it also recognized the fundamental linguistic differences between the Arabic and English languages at the character, word, sentence, and continuous text levels.

The most striking difference between Arabic and other languages is that Arabic text is usually presented without vowels. Vowels, when used, are presented by diacritical marks, placed above or below the character. [Diacritization](#) defines the sense of each word, and how it will be pronounced. However, the use of diacritics has lapsed in modern Arabic writing.

For Example, consider the English words "some", "sum" and "same". Non-Diacritized Arabic would reduce these words to "sm" leaving it to the individual reader to decipher the word and add its missing vowels depending on the context.

For the average Arabic reader, this poses no problems. But, for the machine, it's an extremely difficult task.

Realizing the complications inherent at all linguistic levels, but determined to address this long-standing problem, Sakhr has launched a multi-phase, large-scale research project for the automatic processing of written Arabic. This resulted in the development of its [core technology](#) with a set of [products](#) that use computational linguistic algorithms to transform language structure into binary systems.

1.6.2. Morphology

Sakhr's MMMP is a morphological analyzer-synthesizer that provides basic analysis for a single Arabic word covering the whole range of modern and classical Arabic.

The analyzer part identifies all possible stem forms of a word, i.e. extracting its basic form stripped from the affixes, the morphological data of the word such as root, the Morphological Pattern (MP), and its part of speech.

The synthesizer works in a reverse mode to regenerate the word from its morphological forms (stem, root, morphological pattern, part of speech and/or affixes).

Sakhr has designed the Morphological Processor to produce word level analysis though regeneration and comparison. In other words, we consider the analysis correct, when the Morphological Processor is able to regenerate and compare the output to the original word.

1.6.3. Correction

Sakhr Corrector automatically detects and corrects Arabic spelling mistakes and common Arabic mistakes, as well as grammatical mistakes.

This process is very important in order to carry out detailed text analysis tasks, where the input text must be correct before attempting to extract its [keywords](#), [categorize](#), [summarize](#), [translate](#), [search](#), or do any further processing.

1.6.4. Keywords Extraction

Sakhr Keywords Extractor analyzes an Arabic document, and identifies its individual text phrases and key data items automatically.

It helps companies transform key business processes by identifying important information within high volumes of content, then intelligently extracting the most valuable details for use throughout the enterprise.

Keywords help classify the document in a user defined subject tree for easy browsing and access to information. It helps cross-link documents to each other in a dynamic way, and reduces the time needed to search for its content.

The Sakhr Arabic Keyword Extractor engine gives you the ability to quickly access individual features and phrases for use in creation of custom text summaries, information classification, or other text mining activities.

1.6.5. Categorization

Sakhr Categorization Engine accurately and efficiently organizes valuable information into a logical topic tree or taxonomy of enterprise content.

By providing clear visibility into every stage of the categorization process, from taxonomy creation, to content population, to access, and display management, you are able to build and maintain the most accurate and flexible information resource for end user retrieval.

The Categorization engine employs the [Sakhr Corrector](#) to automatically correct the input Arabic text from common Arabic mistakes, and the [Keyword Extractor](#) to identify a prioritized list of keywords used to accurately map the document to the user defined subject tree.

1.6.6. Sumarization

The Sakhr Arabic Summarizer engine identifies the most relevant sentences within a text and displays them in the form of a short text summary. The Summarizer makes it easy to scan just the important sentences within documents, greatly reducing the time needed to read and process items manually.

It provides companies with short text summary for each item using a prioritized list of key sentences and gives the ability to select a specific level of summarization.

The Summarization engine employs the [Sakhr Corrector](#) to automatically correct the input Arabic text from common Arabic mistakes, and the [Keywords Extractor](#) to identify a prioritized list of keywords to accurately identify the important sentences.

1.6.7. Machine translation

Machine translation was debated long before computers actually executed it. In fact, it was one of the earliest applications suggested for digital computers. However, these early visions of a machine that would translate any text with the click of a button without the intervention of a human translator have proved to be somewhat naive. Reality has proven to be more difficult and, in many ways, more challenging than earlier envisioned.

Sakhr Software took up the challenge of developing English to Arabic, and Arabic to English Machine Translation engines relying on its experience and extensive technology arsenal in the field of [Natural Language Processing \(NLP\)](#).

The main objective of machine translation is to translate text from one natural language to another. The increased demand for translation due to world globalization, and the growing importance of communication has led to the rise in demand for machine translation, which is fast becoming a cornerstone of translation projects around the world.

Sakhr Software has used its MT Engine to power www.tarjim.com, a Web-based online translation service that translates, on-the-fly, any English Web page into Arabic and vice versa, while preserving original Web page layout and adjusting language direction. It also allows for the translation of text.

1.6.8. Diacritization

Since written Arabic does not include vowels (Diacritics), Sakhr has developed an Automatic Diacritizer to generate these missing vowels on the fly to enable Arabic [Text-To-Speech](#)(TTS) rendering.

Diacritization disambiguates word meanings relying on an extensive basic research on the area of [Natural Language Processing \(NLP\)](#).

Automatic Diacritization depends on various levels for language processing and analysis, starting from [morphological level](#) and ending with disambiguation of word meanings relying on an extensive basic research in the area of [Natural Language Processing \(NLP\)](#), and large linguistic databases that Sakhr has developed over long years.

The Diacritizer engine outputs Diacritized Arabic text instantly with an accuracy of 98%, allowing you to select the required Diacritization level: full, mandatory, or case ending diacritics (Syntax).

1.6.9. Speech

According to Frost & Sullivan in their U.S. IVR systems market report, 48.5% of IVR ports sold in the US will be speech-enabled by 2004-2005. Also according to Gartner Group, 92% of all customer interactions happen by the phone, and 80% of all unresolved issues on the Web transition to the phone.

Speech is the most natural interface between man and machine. Speech-enabled applications free agents from routine tasks and orient them to more revenue generating activities. Sakhr developed both Arabic Text to Speech [TTS](#) and Automatic Speech Recognition [ASR](#) engines. The TTS engine converts any Arabic computer readable text into a human sounding synthetic voice. The ASR engine recognizes Arabic utterances and commands from different speakers and different accents.

These speech technologies can be used and integrated with hundreds of different applications such as IVR (Interactive Voice Response), Call centers, Unified Messaging Systems, Talking Emails, Directory Assistance Systems, Voice Dialing Systems to name a few. These best-of-breed technologies supports the latest standards emerged on the market as SAPI (Speech Application Programming Interface), VoiceXML, SALT (Speech Application Language Tags) and telephony development tools as Intel CTADE (Computer Telephony Application Development Environment), .NET Speech SDK, ENVOX etc.

Why businesses must switch to speech technologies?

Saving Money:

- Operating costs of a speech-enabled system are about 10% of a comparable human-powered call center and half to 1/3 the cost of touch-tone or web system.
- Speech-enabled system is 24x7 available and automatically expands and contracts to handle peak loads and off-hours.
- Shortening call time, which means lower call charges

Improving Customer's experience:

- Never busy, never on "hold" even during peak hours.

- No numerical menus, the customer just asks for the product or service they need.
- Some studies show that 80% of the callers encountering DTMF systems will quit or select to transfer to an operator. This rate drops to 20% using speech-enabled systems.
- Ultra fast service for experienced users.

1.7. Press

Welcome to Sakhr's Press Section. This section covers all performed activities at the Company. Rich in information and inspirational for writers, the section displays press interviews with Mr. Mohammad Al Sharekh, the Chairman of the Board of al-Alamiah Group/Sakhr.

It also demonstrates the press releases written about Sakhr's news along with its sister companies of al-Alamiah Group, including products, new versions, projects, contracts, and the Company's granted awards, in addition to the VIP visits, events, exhibitions, seminars, and conferences participation.

Published press articles in different newspapers, magazines, periodicals, and websites about Sakhr are collected and uploaded, besides, the monthly newsletters, which contain Sakhr Software and al-Alamiah Group's news, in addition to fixed features such as: the distinguished character, the latest world technologies, the caricature, and finally the list of awards granted to Sakhr Software.

1.7.1. Awards

As a natural result of its leadership in Arabic Software industry, Sakhr has got many prizes and awards that testify to its top-notch solutions. Following is a complete list for all the awards Sakhr has got.

1. Arabian Business e-Achievement Award E-Visionary of the Year (2002):



Mohammed Al Sharekh, the owner and chairman of Sakhr Software, won the 'e-Visionary of the Year' award at the Arabian Business e-Achievement Awards, instituted by ITP and hosted by Dubai Internet City and Dubai World Trade Center.

The award, presented by HH General Sheikh Mohammed Bin Rashid Al Maktoum, Crown Prince of Dubai and UAE Defense Minister at Dubai Internet City, on the sidelines of GITEX Dubai 2002, recognized Al Sharekh's role as the leader of Arabic software.

2. Internet Shopper (T4S International Group):



- Sakhr Site: Best Arabic Language Supported Site - Gitex Cairo 2000

- Sakhr Site: Best Arabic Portal - presenting translation services from English Sites to Arabic Language - Gitex Cairo 2000

3. Riao2000 Conference:



- Machine Translation service for internet sites: creative application award

Damtex Middle East Group:



- Sakhr: Best Arabic Application Development Company – 1999

DIT (Dabbagh for Information Technology):



- Arab developers achievements in the IT field Dr. Nabil Aly, Vice Chairman of the Board - Gitex 1999.
- 2000 D.I.T.(2000 Awards of Excellence) -ArabDox- 2.0.
- Arabic applications developer of the year - Sakhr Software Co. - Gitex 1999.
- Arabic Agenda/productivity application, Arabic MMMP software - Gitex 1999.
- Best Arabic site in the Middle East, Sakhr site - Gitex 1998.
- Best Company contributing in Arabic software development, Sakhr Software Co. - Gitex 1998.
- Best Arabic company for Internet products, Sakhr Software Co. - Gitex 1998.
- Best Arabic site in the Middle East, Sakhr site - Gitex 1998.
- Best Company contributing in Arabic software development, Sakhr Software Co. - Gitex 1998.
- Best Arabic company for Internet products, Sakhr Software Co. - Gitex 1998

ITP Group:

- Best Business applications award - Gitex 1998.
- Best Arabic software for 1998, OCR Gitex 1998

PC Magazine - Arabic Edition:



- Automatic Corrector 1999 - Gitex Dubai 1999.
- Editors choice, Sindbad Arabic browser - Gitex 1998.
- October 1998 PC magazine (Best of Termium) IDRISI.
- Middle East IT man of the year - Gitex 1997, Mr. Mohamed Al Sharekh - el- Alamiah Group President and Sakhr Software Co. Chairman of the Board).
- Best Arabic software Developing company - Sakhr software - Gitex 1997.
- Best Arabic Internet technologies (for Sindbad, NasherNet, Idrisi and Al-Daleel) - Gitex 1997.
- Second best Internet Arabic site - Sakhr site - Gitex 1997.
- Editors choice, Sindbad best Arabic browser - May 1997.
- June 1997 PC magazine (Editors' choice) Sindbad-Netscape 3.0.
- May 1997 PC magazine (Editors' choice) Sindbad-Netscape 3.0.
- Editors choice, Sakhr MMMP - Gitex 1996.
- Editors choice, Al Hadeeth Al Shareef - Gitex 1996.
- Editors choice, Al Qur'an Al Kareem - Gitex 1996.
- Editors choice, OCR (Optical Character Recognition) - Gitex 1996.
- Editors choice, Typing Tutor - Gitex 1996.
- Editors choice, Ozzie's World - Gitex 1996
- Editors choice, KC series for teaching English - Gitex 1996.
- Editors choice, The Dictionary built-in with Al Ostaz application - 1995.
- Best Arabic CD, Al Qur'an Al Kareem - 1995.
- April 1995 PC magazine (Editors' choice) Sakhr Dictionary.
- May 1995 PC magazine (Presented by D.I.T) Sakhr Dictionary

Byte Magazine - Middle East:



- Best Internet Technologies, Sindbad - Gitex 1997.
- Best Arabic Internet software, Sindbad - Gitex 1996.
- Best Arabic Internet technology, Idrisi - Gitex 1996.
- Best Arabic technology, Automatic Diacritizer - Gitex 1996.
- Best Arabic software, Teaching Arabic as a foreign language - Gitex 1996.
- Editors choice, Electronic Publishing, Gitex 1996.
- Best Arabic technology for Arabic texts retrieval system - 1996.
- Best Arabic application for El Ostaz word processor - 1996.
- One of best 20 Arabic applications, the Corrector - 1995.
- Best Arabic software, Al Qur'an Al Kareem - 1995.
- Editors choice, for the best Arabic product, The Dictionary - 1995

Internet World:



- Sakhr receives the Readers' Award as the best Arabic Islamic site on the Internet October 1997

Windows:



- Ajeeb - best Arabic Web Portal - Gitex Dubai 2001.
- ArabDox - best Arabic Business Software award - Gitex Dubai 2001.
- Tarjim product of the year - Readers' choice - 2000.
- Sakhr CAT best Arabic software Gitex Cairo 1999.
- Sakhr Company of the year by Arab readers' choice 2000.
- Windows Magazine (Top 60 products of 2000) ArabDox.

Other Award:

- Best Products Award for 1998 - Comdex Egypt 1998.
- 1998 Comdex (Best of show) NasherNet 2.0.
- 1998 Comdex Sindbad Communicator.
- 1998 Comdex IDRISI.
- Certificate of Estimation from the Human Resources Investment Company (Kuwait) - May 1996.
- Certificate of Estimation from the Ministry of Interior (Kuwait) - January 1996.
- Certificate of Estimation from the Teachers Committee in Kuwait - April 1993.
- Certificate of Estimation from the Saudi Computer Committee.
- Certificate of Estimation from the Ministry of Education in the Kingdom of Saudi Arabia.
- Certificate of Estimation from the King Saud University in the Kingdom of Saudi Arabia.

2. From: <http://www.aramedia.com/>

2.1. What is an OCR technology?

OCR stands for Optical Character Recognition. When you scan a text document, the computer can only see this text as graphical figures on a page. That text is not usable, searchable, or editable. It's considered as an image.

OCR programs are designed to read this text and recognize it. It is a software that converts the shapes (in a scanned text) into a text document. However, there may be some errors especially if the font is small or the scanning unclear. But after all, it is considered as a faster and more time-saving way than typing the text.

2.2. Arabic OCR

With the Arabic language having its unique characteristics, it was more difficult for Arabic OCR technology to exist. There're many challenges in Arabic text recognition. First Arabic is written cursively so that more than one character are written connected to form what we call 'Block of Characters- BC'. It can also be written in many fonts so that a BC has more than one base line. Then Arabic uses many types of external objects such as dots, 'Hamza' and 'Madda'. Also Arabic may be written diacritized which adds a new set of external objects. Besides, Arabic characters can have more than one shape according to their position inside the BC (initial, middle, final, or standalone BC). There's also Overlapping, which makes the problem of determination of spacing between BCs and words difficult. Moreover, Arabic fonts suppliers do not follow a common standard and therefore, Arabic fonts are very different and characters characteristics are not reliable to build an OMNI OCR.

Depending on its researches in computational linguistics, Sakhr Software was able to get over these problems and produce its Arabic OCR program.

It combines two main technologies: Omni technology which depends on highly advanced research in Artificial Intelligence, And the Training technology which increases accuracy of character recognition. Sakhr OCR can identify more than one language through Xerox TextBridge technology, one of the most popular OCR programs. The program can also identify both Arabic and English characters on the same page.

The OCR can distinguish between 26 true types of Arabic fonts. The Advanced version of the program is designed to enable the user scan large amounts of documents, save them as graphic files and classify them to be recognized later to save time. The program recognizes graphics and puts them in their proper place on the page. The program also saves page format without any modification to tables, columns or graphics. It can also identify diacritics and keep or remove them according to the user's choice.

The output can be saved to disk for use in a variety of applications such as word processing. The program is compatible with many programs such as DDE or OLE automation.

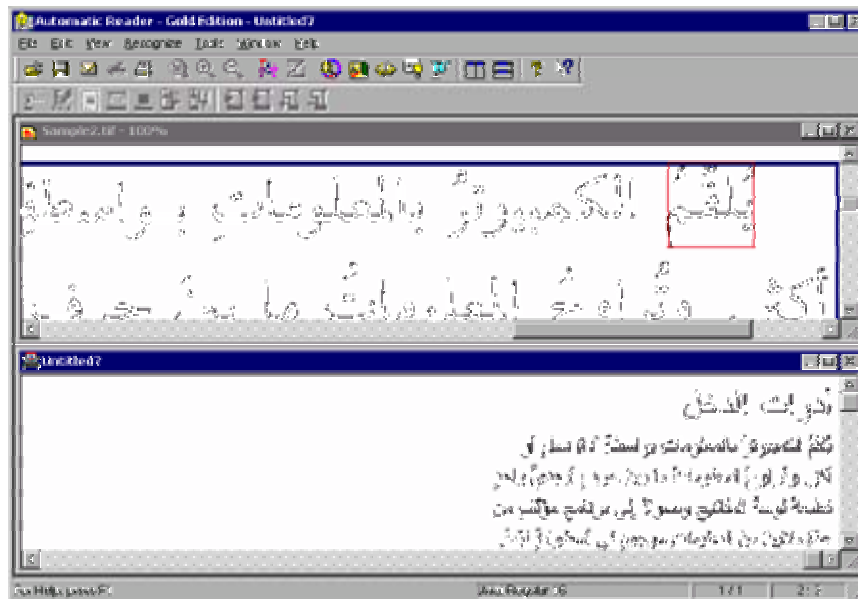
The Advanced Version and the Office Version of the OCR are compatible with the Arabic or any European edition of Microsoft Office 95.

2.3. Overview of Sakhr, The Best Arabic OCR Technology

The OCR, backed by Sakhr's experience in Natural Language Processing [NLP](#), can solve the Arabic language unique characteristics.

To begin with, Arabic is written cursively, where several characters are connected to form what we call 'blocks of characters'. Arabic can also be written in many fonts, so that a 'block of

characters' has more than one base line. Additionally, Arabic uses many types of external objects such as dots, 'Hamza' and 'Madda'.



[Diacritization](#) adds a new set of external objects. Furthermore, Arabic characters can have more than one shape according to their position inside the block of characters (initial, middle, final or standalone block of characters).

Overlapping also makes it difficult to determine the spacing between blocks of characters and words. Finally, Arabic font suppliers do not follow a common standard. Given the peculiarities of Arabic fonts and the characteristics of the Arabic language, building an OMNI OCR becomes a difficult undertaking.

Sakhr OCR combines two main technologies: Omni Technology, which depends on highly advanced research in artificial intelligence, and the second is Training Technology, which increases the accuracy of character recognition.

It can identify more than one language through Xerox Text Bridge technology; one of the most popular OCR programs. The program can also identify both Arabic and English characters on the same page. Moreover, OCR can distinguish between 26 Arabic true type fonts. It enables the user to scan large amounts of documents, save them as graphic files and classify them for later recognition in order to save time.

The program recognizes graphics and places them in their proper location on the page. The program also saves page format without any modification to tables, columns, or graphics. It can also identify diacritics and keep or remove them according to the user's preferences.

The output can be saved to disk for use in a variety of applications such as word processors. The program is compatible with many other programs, and OCR SDK (Software Development Kit) is available to allow integration with other applications. The OCR SDK is available in both DLL and ActiveX formats.

Sakhr OCR engine has been optimized using Intel tools to better run on the latest Pentium IV processors that support Hyper-Threading.

Sakhr's Automatic Reader transforms scanned images into a grid of millions of dots, optically recognizes the characters found in them and ultimately converts them into text. The complex nature of the Arabic language is evident in the cursiveness of the text, character overlapping, various character shapes, diacritics and the variety of calligraphic Arabic fonts that

exist. As a result, these specific Arabic language complexities present major technical challenges in the Arabic OCR industry. The Automatic Reader, backed by Sakhr's extensive experience in Arabic Natural Language Processing (NLP) technologies, addresses these challenges effectively, thus providing Arabic users with an award-winning and high quality OCR solution.

2.4. KEY FEATURES

2.4.1. Performance and Accuracy:

- 800 characters per second on PIII based computer.
- Up to 99% accuracy in recognizing Arabic books, newspapers, etc...
- Runs under Windows 98 (A or E), NT or 2000 or XP (Arabic-Enabled).

2.4.2. Recognition Engines:

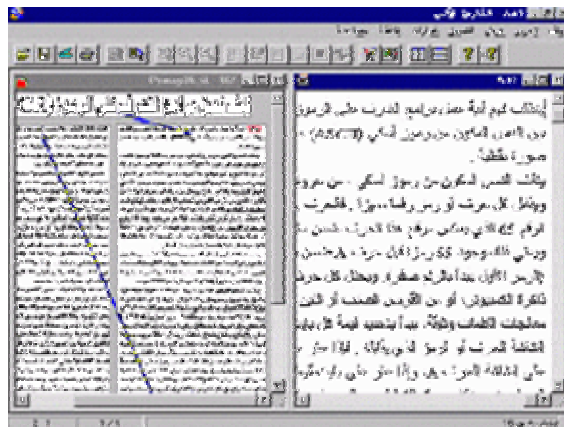
- Supports Arabic, Persian, English, French and 15 other languages.
- OCR bilingual documents (Arabic/English, Persian/English and Arabic/French).
- Supports both OMNI & Learning technologies to obtain higher accuracy in different fonts.

2.4.3. Supported Formats:

- Sakhr OCR can deal with different image format standards (.bmp, .tiff, .pcx, etc...).
- Sakhr OCR can save the output text in different formats such as .txt, .rtf, and HTML.

2.4.4. Supported Scanners:

- Twain, ISIS and KOFAX protocols.
- Works with any type of scanner.
- Supports simplex and duplex scanners



2.4.5. Unique Features:

- Recognizes the diacritics in Arabic images.
- Opens multiple documents at the same time.
- Recognizes tables in scanned images.
- Special support for ill-formed tables.
- Recognizes underlined words.
- Recognizes broken and stickled characters.

- Automatic style detection for fonts (Regular or Bold).
- Uses Arabic linguistic rules with recognition (Artificial Intelligence).
- Supports non-rectangular frames.
- Supports color documents.

2.4.6. Powerful Imaging Tools:

- Automatic and manual image rotation.
- Automatic and manual image fixing.

2.4.7. Integration with Other Applications:

- Can easily integrate with any third party application.
- Software Development Kit (SDK) (in Platinum version).

2.4.8. Other Important Features:

- Program interface is available in both Arabic and English.
- Includes bilingual spellchecker.
- Supports both automatic and manual framing modes.
- Ability to send OCR results by e-mail.

2.4.9. Bonus Package:

- Personal Archive: This program enables users to organize images in structured and hierarchical folders.
- The OCR can distinguish between 26 true types of Arabic fonts. The Pro version of the program is designed to enable the user scan large amounts
- of documents, save them as graphic files and classify them to be recognized later to save time. The program recognizes graphics and puts them in their proper place on the page. The program also saves page format without any modification to tables, columns or graphics.

2.4.10. The PROFESSIONAL Version:

- Automatic Reader Pro supports the following languages: Arabic, Farsi, French, English, Czech, Danish, Dutch, Finnish, German, Greek, Hungarian, Italian, Norwegian, Polish, Portuguese, Russian, Spanish, Swedish, and Turkish.. This version also adds training technology to Omni technology to further raise accuracy levels. Users access 4 batch modes, can train the program in specific fonts, use spell checkers, Arabic linguistic rules, and OLE and DDE features.

2.4.11. Price:

- Sakhr's Automatic Reader Gold Edition Arabic/Multilingual OCR:
- *US \$2499.95; Sale Price: US \$1999.95.*
- Sakhr's Automatic Reader Platinum Edition Arabic/Multilingual OCR with SDK Kit:
- *US \$4999.95; Sale Price: US \$3999.95.*



2.5. Technologies and approaches used in Sakhr OCR

2.5.1. Contour creation:

In Sakhr OCR, after image is captured as bitmap from either scanner or image file, contours are created for all objects in the input bitmap and thereafter, bitmap is no longer used and contours are used instead to represent the input. A new contour creation algorithm is developed and used in Sakhr OCR. The algorithm is deterministic and succeeds in creating the contour in all cases regardless the contour complexity.

2.5.2. Creation of character Blocks:

In Arabic, dots and other what we call external objects are used heavily. This may make lines overlap. In such cases linking external objects to main objects to create what is called block of characters (BC) becomes a very difficult process. Sakhr OCR uses an algorithm which contains a big number of intelligent heuristics to create BCs. This algorithm gives almost 100% accuracy.

2.5.3. Segmentation:

We believe that the segmentation algorithm use in Sakhr OCR is almost the most powerful algorithm available now to solve this problem. It gives over 98% success and even in cases the algorithm fails; we use what we call User-defined ligatures to allow users to create his own ligatures.

2.5.4. Line separation:

As I mentioned in the above point and in section 2, Arabic uses a lot of external objects such as dots, 'hamza', 'madda', and diacritic marks. This makes the line separation much more difficult problem than in Latin languages. However, we have a strong line separation algorithm that even handles to good extent the problem of lines which may be stuck in part of it.

2.5.5. Framing:

There is a module in Sakhr OCR for what we call 'automatic framing'. This function is to divide the input page into columns if any, to exclude what are images or graphics from the recognition process in 'exclusion frames' and to include the paragraphs if any in several inclusion frames. Also the table and image detection are processed in this phase.

2.5.6. Recognition:

The recognition process in Sakhr OCR uses the contour and opological features in addition to the bitmap features using eural networks in OMNI recognition and matching in learning. The accuracy in the best quality can reach 99.8%. Also linguistic models are used to increase accuracy in bad quality.

2.5.7. Post-recognition:

One of the major critical processes in OCR systems in general is the post-recognition operations. For instance the correction process. Sakhr OCR relies on a very powerful linking between the input image and the output text which facilitate the correction process. Also Sakhr OCR uses a very powerful Spelling checker which is built on an accumulative research work in Sakhr for more than five years.

2.5.8. Output:

Sakhr OCR can export the recognized text to different word processors and in different formats. In addition to the text format, it can generate SWF format which can be read by all Sakhr Word processors, RTF to be read by most other word processors such as MS Arabic Word retaining the original document layout (columns, tables, images, indentation, etc...). Also Sakhr OCR can pass the recognized text directly to different word processors: Al-moharer and MS Arabic Word.

2.5.9. Integration:

Sakhr OCR can integrate with other applications though DDE communication protocol and OLE Automation.

2.6. Sakhr OCR - Frequently Asked Questions

That was an Older Version 2.0, Current Version is 7.0.

2.6.1. Is there a shortcut to delete the last frame created?

The user can delete the last frame created simply by clicking the right button of the mouse where a warning dialog will pop up ensuring the user for frame deletion.

2.6.2. How many image formats OCR version 2.0 can support?

The supported image formats are: *.TIF, *.PCX, *.BMP, *.MAG, *.TGA, *.GIF, *.DCX, *.JPG.

2.6.3. To what applications can the user export an (*.ART) file?

An (*.ART) file can be exported to either Al Ustaz with Ustaz format (*.SWW) or MS-Word with (*.RTF) format.

2.6.4. What does file subformat means in "Save image as" dialog?

This version can save an image (*.Tiff) format with a subformat and this feature is supported with the (*.TIF) format only. The available subformats for a (*.Tiff) images are:

- Tiff Uncompressed
- Tiff Huffman
- Tiff CCITT G3 Fax
- Tiff LZW
- Tiff CCITT G4 Fax
- Tiff Packbits

2.6.5. How many user defined ligatures OCR version 2.0 can support?

OCR version 2.0 can support over 256 user defined ligatures per font which is already a very sufficient number for a user to define his own ligatures per font.

2.6.6. How can the user define his own ligatures?

During the learning process, the user can create his own ligatures (which may be composed of 2 characters up to 6 characters) by typing the ligature in the character field editor then press OK button. A sample dialog will pop up containing the typed ligature.

Note: The user defined ligatures will all be found in the Ligatures combo box, to be selected when needed.

2.6.7. How to learn a ligature using the mouse?

Press CTRL key then double click the ligature characters in the learning dialog then press OK.

2.6.8. What are the different file formats the user will deal with in OCR version 2.0?

(*.AFN) Font File, (*.AFL) Font Library, (*.ART) Auto Reader Text File, (*.AFR) Frame file

2.6.9. How many fonts can be included in one Font library?

One font library can include up to 250 fonts.

2.6.10. How does OCR version 2.0 handle colored images?

OCR can simply learn and recognize a colored image after converting it to 2 colors (black and white), and this is available when the image is already saved. OCR can NOT scan a colored image.

..... And what is the case if the image is gray?

OCR does NOT support gray scale images.

Gray images can neither be opened, nor learned or recognized.

2.6.11. How can the user create a default frame file of his own to be applied on every opened (or scanned) image?

- The user can create his own frame file and make it the default frame by following these steps:
- Create your own inclusion frames on a certain image
- Frame menu, Frames file menu item, Save As option then save the frame file (*.AFR)
- Choose Default Frames then in the select default frames file dialog select your (*.AFR) file, hence your frame file will be applied on any further opened image automatically.

..... And how could the user cancel a default frame?

To cancel the default frame file, click None button in the Select default frames file dialog in default frames option in frames file menu item in Frame menu...

2.6.12. What is the difference between learning a failure and learning a word after the recognition process?

To learn a failure is to learn an unrecognized character, but to learn a word is to learn wrongly recognized characters, hence the font file or the font library will be updated accordingly.

2.6.13. What is the difference between DOS text format (*.TXT) and DTPtext format (*.TXT)?

The main difference is that the DOS text format takes the carriage return into consideration, while DTP text format does not.

i.e. when exporting an (*.ART) file with DOS text format, sentences are exported as it is exactly with the carriage return positions, while sentences exported with DTP text format will wrap.

2.6.14. The minimum hard disk space required for installing OCR for 16 bit is 10 MB while that required for 32 bit is 20 MB. What is that big difference for?

Actually the real free hard disk space required to install OCR 16 bit is about 9 MB and that required for 32 bit is about 10 MB, but OCR creates temporary files on the hard disk during processing and deletes them right after that. So the actual free disk space required to run OCR is greater than that required to just install it.

2.6.15. What are the supported features regarding scanning multiple images process?

The process of scanning multiple images is supported by the following features: Ability to choose the image files format of the images to be scanned (*.TIFF) (*.PCX) (*.BMP) (*.WPG) (*.DCX) and choosing the images file subformat in case of TIFF format. Sequential images

scanning: scans the images sequentially (i.e. face only) as in TWAIN protocol and in this case the images will be saved sequentially (i.e. image01, image02, image03, ...). Duplex images scanning: scans the paper both sides as in Kofax protocol and the scanned images will be saved as image01a, image01b, image02a, image02b,...

2.6.16. How can a user having a scanner using Twain protocol scan double sided papers through OCR?

A user can scan double sided papers through OCR given a scanner that supports Twain protocol by the following method:

- File menu, Scan multiple menu item, Scan and save dialog will pop up, choose "Odd" combo box from the sequence conventions, then scan one face for all papers. The resulting scanned images will be saved as image01, image03, image05,
- Reverse the papers so as to scan the other face, and this time choose "Even" combo box from sequence conventions from the above mentioned dialog, then scan again. The resulting scanned images will be saved as image02, image 04, image 06,...

2.6.17. The "Fix Image" check box is pale in Scan & Save dialog (File menu, Scan multiple menu item) and is active when the Load and recognize check box is selected in the same dialog?

The "Fix image" feature is used while recognition only and not while scanning the image because fixing the image means adjusting the skewed lines and converting the black backgrounds into white ones and all these matters are requested during the recognition process.

2.6.18. In the Scan & Save dialog (File menu, Scan multiple menu item), sequence conventions, one of the two combo boxes "Sequential" and "Duplex" will be pale. Why?

Because these two combo boxes "Sequential" and "Duplex" depends on the currently installed scanner type. If the scanner supports TWAIN Protocol, "Sequential" will be active. If the scanner supports Kofax protocol, "Duplex" will be active.

2.6.19. Is it available to print an opened image?

No, printing an opened image is not supported, but instead, printing the (*.ART) file is supported.

2.6.20. While using DDE with Access version 2.0 as a client application and OCR as a server application sometimes with very large frames this message appears "Timeout while waiting for DDE response", how to overcome this?

We can overcome this problem from Access version 2.0 settings. Select View menu, Options menu item, then select "General" as the category. Change OLE/DDE Timeout from 30 to 300 seconds, (for example).

2.6.21. What is the meaning of "Use Morphology in recognition" (Preferences dialog)?

This check box is used to apply the Morphological rules on the recognition process. These morphological rules are simply based on some characters can never be connected to another so as to compose one correct word, such cases will result in recognition failures.

2.6.22. What is the difference between Logical layout and Visual layout in the output page setup?

The difference between the logical and the visual layout in exporting an ART file to MS Word (i.e. the RTF format) is that "Retain the logical layout" means that if the image includes columns, or if it is divided into sections, it will be exported

to Word with the same layout (i.e. same columns and same sections) as if it is written using Word from the beginning, while "Retain Visual layout" means that the (*.ART) exported to Word will be exported as frames, frame by frame as they are in the image itself.

2.6.23. What is the difference between "Restore Settings" and "Restore defaults" (options menu)?

Restore settings is used to return the settings to the last saved settings, but Restore defaults is used to return the OCR settings to the default settings.

2.6.24. What is new in OCR version 2.0 about the feature of associating the caret position in the text file with its location in the image file?

Associating the caret position in the text file (*.ART) with its location in the image file is enhanced in OCR version 2.0 in a way such that if the image file was closed and the user selected Edit menu, Show Image menu item, the related image will be opened and the operation of associating the caret position in the ART file with the image will be completed.

فهرست اصطلاحات

آ-۱

Meta-structure	ابرساختار
Blobs.....	اجزاء لکه‌مانند
Orthogonal.....	ارتوگونال (متعامد)
Skew	اریب
Office Document Architecture (ODA) Standard.....	استاندارد ساختار مستندات اداری
Feature extraction	استخراج ویژگیها
Thresholding.....	اعمال سطح آستانه
Curvilinear distortion	اعوجاج منحنی-خطی
Smearing.....	آغشته‌سازی
Alphanumeric	الفبایی-عددی
Dilation.....	انبساط
Reject Option.....	انتخاب وازدگی
Connected component analysis	آنالیز اجزاء پیوسته
Document Analysis (DA) or Document Image Analysis (DIA)	آنالیز اسناد (مستندات)
Vertical projection analysis	آنالیز تصویرنمایی عمودی
Multi-Resolution Analysis (MRA).....	آنالیز چند تفکیکه
Structural analysis	آنالیز ساختاری
Functional analysis.....	آنالیز عملکردی
Linear Discriminant Analysis (LDA).....	آنالیز مبین خطی

ب

Optical Character Recognition (OCR)	بازشناسی نوری حروف
Textured	بافت‌گونه
Top-Down	بالا به پایین

Representation	بازنمایی
Binary	باینری (دو سطحی)
Line Fitting	برآزش خطی
Labeling	برچسب گذاری
Online	برخط
Offline	برون خط
Series Expansion	بسط سری

پ

Footer	پاصفحه
Bottom-Up	پایین به بالا
Image processing	پردازش تصویر
Natural Language Processing (NLP)	پردازش زبان طبیعی
Picture Processing or Photographic Processing	پردازش عکس
Graphical Processing or Graphics Processing	پردازش گرافیک
Query	پرس وجو
Postprocessing	پس پردازش
Projection profile	پروفایل تصویرنمایی
Preprocessing	پیش پردازش
Layout	پیکربندی

ت

Probability Density Function (PDF)	تابع چگالی احتمال
Point-spread function	تابع گسترش نقاط
Discrete Fourier Transform (DFT)	تبدیل فوریه گسسته
Karhunen-Loeve (KL) transform	تبدیل کارهونن لُوئو
Wavelet transform	تبدیل موجک
Hough transform	تبدیل هاف
Alignment	ترازبندی
Dissection	تشریح

Template matching..... تطابق قالبی

Vertical Crossing Counts..... تعداد گذرهای عمودی

Resolution..... تفکیک پذیری (رزولوشن)

Splitting and Merging..... تقسیم و ادغام

Interactive..... تعاملی

Feedforward تغذیه رو به جلو

Dilation..... تغییر شکل (برای حروف)

Scaling..... تغییر مقیاس

Crossing..... تلاقی

Full image..... تمام تصویر

ج

Time–frequency localization..... جانمایی زمانی – فرکانسی

Field Separator جداکننده ناحیه‌ای

چ

Machine printed..... چاپی

Multiscale چند مقیاسه

ح

Sensor حسگر

خ

Baselines..... خط زمینه

Erosion خوردگی (برای حروف)

Clustering خوشه‌بندی (کلاستر بندی)

د

Domain Dependant..... دامنه-محدود

Kurtosis درجه تیزی نقطه اوج منحنی

Decision Tree درخت تصمیم‌گیری

Script دست خط (خط)

Handwritten دستنویس

Facsimile دورنگاری

Compactly supported دوره-محدود

ر

Orientation راستا

Selective attention رسیدگی انتخابی

Flowchart روندنما (فلوچارت)

Segmentation-free approach رویکرد مبتنی بر بازشناسی کلمه به عنوان یک الگوی واحد

Segmentation-based approach رویکرد مبتنی بر قطعه‌بندی کلمات

ز

Stroke زیرحرف

س

Tabular Structure ساختار جدولی

Cursive سرهم

Global سراسری

Header سرصفحه

Gray levels سطوح خاکستری

Sonar سونار

ش

Indexing شاخص‌گذاری

Sharpening شارپ کردن

Neural Network (NN) شبکه عصبی

Pattern Recognition شناسایی الگو

ص

Digitizer صفحه رقومی‌کننده (دیجیتایزر)

ع

Uncertainty	عدم قطعیت
Morphological operations.....	عملگرهای مورفولوژی
Fractal signature	علائم فرکتالی
Landmarks	علائم مشخصه

غ

Translation invariant.....	غیرحساس به انتقال
----------------------------	-------------------------

ف

Between-line Spacing.....	فاصله بین خطوط متن
Within-line Spacing.....	فاصله درون خطی (فاصله بین کلمات و بین کاراکترهای یک خط متن)
Freeman.....	فریمن (کد)
Compression.....	فشرده سازی
Table Of Contents (TOC).....	فهرست مطالب

ق

Page Border	قاب طرحدار (قاب دور صفحه)
Reconstructability.....	قابلیت بازسازی مجدد
Segmentation	قطعه بندی
Internal Segmentation.....	قطعه بندی بیرونی
External Segmentation	قطعه بندی درونی
Explicit Segmentation	قطعه بندی صریح
Implicit Segmentation	قطعه بندی ضمنی

ک

Bounding Box	کادر محصورکننده
Contour.....	کانتور
Convolute	کانولو کردن
Convolution.....	کانولوشن

Slant.....	کجی (در مورد تصویر)
Chain code.....	کد زنجیره‌ای
Classification.....	کلاسه‌بندی (طبقه‌بندی)
Contrast	کنتراست (پادنمایی)
Quantization	کوانتیزاسیون

گ

Gabor.....	گابور
Relational graph	گراف پیوندی
Attributed graph	گراف منتصب شده
Discretization	گسسته‌سازی

م

Gabor Dominant Orientation Matrix.....	ماتریس جهت غالب گابور
Local.....	محلی
Training set.....	مجموعه آموزشی
Hidden Markoff Model (HMM).....	مدل مارکوف پنهان
Context model	مدل متنی
Noise modeling	مدلسازی نویز
Evaluation criterion	معیار ارزیابی
Moment	ممان
Pseudo-Zernike moment.....	ممان شبه زرنیکه
Central moment.....	ممان مرکزی
Regular moment	ممان منظم
Wavelet.....	موجک

ن

Zoning	ناحیه‌بندی
Thinning	نازک‌سازی
Pixel wise thinning.....	نازک‌سازی از طریق پیکسل

Normalization.....	نرمالیزاسیون (نرمالیزه نمودن)
Nearest neighbor	نزدیکترین همسایه
Indication.....	نشانه گذاری
Curse of dimensionality.....	نفرین بعد
Fork point	نقطه انشعاب
Extreme point	نقطه حدی (کرانه)
Bending point	نقطه خمش
Resampling.....	نمونه برداری مجدد
Salt-and-Pepper Noise	نویز فلفل-نمکی

و

Graphic User Interface (GUI).....	واسط گرافیکی کاربر
Viterbi.....	ویتربی (الگوریتم)
Invariants	ویژگیهای پایا
Low level features	ویژگیهای سطح پایین

ه

Cross correlation	همبستگی متقابل
Artificial intelligence.....	هوش مصنوعی
Smoothing	هموارسازی
Histogram.....	هیستوگرام

ی

Supervised learning	یادگیری با سرپرستی (با نظارت)
Unsupervised learning	یادگیری بدون سرپرستی (بدون نظارت)
Decimate.....	یک چندی

فهرست علائم اختصاری

Database	DB
Discrete Fourier Transform	DFT
Document Analysis	DA
Document Image Analysis	DIA
Graphic User Interface	GUI
Hidden Markoff Model	HMM
Karhunen-Loeve Transform	KLT
Linear Discriminant Analysis	LDA
Main Page	MP
Multi-Resolution Analysis	MRA
Natural Language Processing	NLP
Neural Network	NN
Office Document Architecture	ODA
Optical Character Recognition	OCR
Page Number	PN
Portable Document Format	PDF
Probability Density Function	PDF
Run-Length Smearing Algorithm	RLSA
Software Developer's Kit (or Software Development Kit)	SDK
Standard generalized Markup Language	SGML
Table of Contents	TOC