مهندس | هرآنچه یک دانشجوی مهندسی لازم دارد

دانلود رایگان : کتاب، جزوه، مقاله، پروژه، گزارشکار و ...

WWW.MOHANDES.ORG

دانشگاه آزاد اسلامی
واحد تهران جنوب
(دانشکده فنی)

پـــردازش تـــصـــویـــــر

بررسی روشها و الگوریتمها

الگوریتمهایی برای OCR وتشخیص چهره

گردآورنده: سیدرضا اشرفیدوست

چکیده

در این مقاله مبانی نظری، تکنیکی و جنبه‌های کاربردی مباحث پردازش تصویر با بررسی دقیق و تکنیکها و الگوریتمهای تشخیص چهرهٔ انسان و بازشناسی نوری حروف (OCR) بصورت جامع مورد بررسی قرار می‌گیرند و زیربخش‌ها و بلوک‌های پردازشی آن معرفی می‌گردند.

بطور کلی موراد مطرح شده عبارتند از :

- پردازش تصویر و کاربردهای عمومی آن .
- تشخیص الگو .
- پردازش متون و مفاهیم پایه‌ای .
- خصایص و پیچیدگی‌های مختص نگارش زبان فارسی در یک نرم‌افزار OCR .
- تحقیقات داخلی انجام‌شده در زمینه OCR .
- روشهای پردازش چهرهٔ انسان
- الگوریتهای پردازش تصویر
  - روش مینیمم هیستوگرام (mode)
  - روش درصد دنباله هیستوگرام (p-tile)
  - روش برگشتی سد احمد (sid-ahmed)
  - روش ماکزیمم واریانس
- ویژگیهای خود همبستگی

مقدمه

پیدایش علوم و فنون جدید جوامع بشری را با شکلهای مختلفی روبرو نموده است. سطح توسعه یک جامعه را می‌توان با مقدار اطلاعات ودانش تولید شده در آن ارزیابی کرد.

تولید فزاینده اطلاعات به شکلهای مختلف صورت می‌گیرد و با درجات متفاوتی از پیچیدگی همراه است. در نتیجه نیاز با سیسمتهای پردازس اطلاعات بصورت روزافزون افزایش می‌یابد، یکی از این مسائل مهم که شاخه‌ای از هوش مصنوعی است، پردازش تصویر است.

در این مقاله ابتدا به شرحی مختصری در مورد جنبه‌های مختلف و کاربردهای پردازش تصویر و تشخیص الگو می‌پردازیم و سپس به تشریح OCR ادامه می‌دهیم.

کاربردها و روشهای استفاده متعددی از آن پردازش تصویر در موارد مختلف وجود دارد، در ادامه به برخی از آنها اشاره شده‌است.

زمینه‌های کاربردی پردازش تصویر و جنبه‌های مختلف آن:

1- بینایی ماشین (machine vision) :

شاید بتوان گفت یکی از بزرگترین مشکلات بشر در تکمیل سیسمتمهای هوشمند مشکل بینایی آنان بود. با پیشرفت علم در عرصه پردازش دیجیتال سیستم ماشین بینایی جای خود را در صنعت باز کرد و روز به روز در حال تکمیل شدن است. یک سیستم ماشین بینایی مجموعه‌ای از عناصر لازم برای گرفتن عکس دیجیتال و اصلاح اطلاعات و عرضهٔ آنها به دنیای بیرون می‌باشد. سه بخش این سیستم عبارتند از:

1- Image Acquisition (بدست آوردن تصویر)
2- Proccessing
3- Output or Display

هر سیستم برای ماشین بینایی شامل این سه بخش کلی خواهد بود. یک سیستم مبتنی بر ماشین بینایی عمده تشخیص (sense) خود را با استفاده از حسگرهای بینایی و پردازشی که روی اطلاعات انجام می‌دهد بدست می‌آورد.

بعنوان مثال تصویر گرفته شده توسط دوربین در روبوتها و پردازش آنها امکان عکس‌العمل مناسب را در قبال اطلاعات دریافت شده می‌دهد. (روبوتهای فوتبالیست)

2- تشخیص چهره انسان:

شناسایی چهره توسط رایانه یکی از جذاب ترین زمینه‌های تحقیقات بیومتریک است که زمینه‌های علمی مختلفی از جمله بینایی ماشین، هوش محاسباتی، شناسایی الگو و روانشناسی را در بر می‌گیرد.
با توجه به این که از هر فرد مورد شناسایی، چندین تصویر با حالات مختلف چهره در حافظه سیستم ثبت می‌شود، تغییر حالات چهره نظیر اصلاح سر و صورت و تغییر وضعیت ابروها، بینی و سایر اجزای چهره نیز تاثیر چندانی در دقت سیستم ندارد؛ بنابراین استفاده از نرم‌افزارهای مجهز به سیستم تشخیص چهرهٔ انسان، می‌تواند ضریب اطمینان سیستم‌های کنترل تردد را به ویژه در مرزها و مبادی ورودی کشور به حد چشمگیری افزایش دهد

در دنیای ارتباطات امروز نیاز به امنیت دسترسی اطلاعات چندرسانه ای در شبکه‌های کامپیوتری کاملاً محسوس است. هر روز خبرهایی درباره دزدی ها و جرائم اینترنتی یا نقض امنیت در یک کمپانی یا سازمان دولتی می شنویم. در بیشتر این موارد بزهکاران از یک نقص اساسی در سیستمهای کنترل دسترسی سود می بردند؛

سیستمها بدون توجه به اینکه " ما چه کسی هستیم"با در نظر گرفتن اینکه " چه چیزی داریم" مثل کارتهای ID ،کلیدها و رمز عبور ویا شماره PIN مجوز دسترسی را اعطاء می کنند. در حالیکه هیچ یک از اینها نمی توانند حقیقت واقعی ما را نشان دهند .

اخیراً تکنولوژیی جهت بررسی هویت صحیح فرد مطرح شده که بر پایه فیلدی به نام **بیومتریک** استوار است .
کنترل دسترسی بیومتریک، روشهای خودکاری را برای بررسی و شناسایی هویت فرد ارائه می‌دهد که بر اساس مشخصات فیزیکی، مانند اثر انگشت، تصویر عنبیه، ویژگی های صورت، DNA و یا برخی جلوه های رفتاری فرد مانندنوع صحبت کردن، سبک دستخط یا الگوهای ضربه زدن به کلید ، به شناسایی فرد می پردازند .
در میان روشهای مختلف بیومتریک ویژگیهای فیزیکی پایدارتر از ویژگیهای رفتاری هستند. زیرا ویژگیهای فیزیکی، باستثنای آسیبهای جدی که منجر به از دست دادن ویژگی می شوند، غیر قابل تغییرند، اما ویژگیهای رفتاری ممکن است در اثر هیجان، خستگی و یا بیماری تغییر کنند.

شناسایی یا بررسی هویت بر اساس تصویر صورت یکی از معدود روشهایی است که در عین حال که مزاحمت زیادی برای فرد ایجاد نمی کند، از دقت بالایی برخوردار است. برای تهیه بیومتریکهای فیزیکی دیگر از جمله اثر انگشت و تصویر عنبیه که لازمه آن تماس نزدیک با فرد است باید از فرد خواست تا مدتی از وقت خود را در اختیار ما قرار داده و یا در وضعیت مشخصی برای تهیه تصویر مورد نظر قرار بگیرد. در حالیکه با نصب یک دوربین می توان بدون ایجاد مزاحمت برای فرد از فاصله دور تصویر چهره او را تهیه و جهت شناسایی یا بررسی هویت استفاده کرد .

شناسایی چهره انسان دارای کاربردهای زیادی در زمینه‌های تجاری، امنیتی و قانونی است و در سالهای اخیر بسیار موردتوجه بوده‌است.
برای مثال استفاده از آن در پایگاه تصاویر مجرمین در ادارهٔ پلیس و تطابق آنها از کاربردهای این روش است.
شناسایی چهره انسان توسط ماشین یکی از موضوعات جالب توجه در علم کامپیوتر و هوش ماشین است که تحقیقاتی را در زمینه پردازش تصویر، شناسایی الگو، بینایی ماشین و شبکه های عصبی برانگیخته‌است. شناسایی چهرهٔ انسان توسط ماشین بدین معنی است که با استفاده از پایگاه داده ذخیره شده از چهره‌هایی که به ماشین معرفی شده‌اند، ماشین بتواند یک یا چند نفر را در تصویر ثابت و یا ویدئویی از یک صحنه شناسایی نماید.

سیستمهای شناسایی چهره کاربردهای متعددی دارند که می‌توان به طور خلاصه آنها را به صورت زیر طبقه بندی کرد:
- تشخیص هویت با استفاده از کارتهای اعتباری ، پاسپورت و غیره
- استفاده از تصویر سابقه داران، برای تشخیص مجرمین
- حفاظت از بانک ها و فروشگاهها با آشکارسازی صورت در تصاویر
- یافتن صورت در یک صحنه شلوغ

یکی از مهمترین محورهای مورد توجه در شناسایی چهره شامل جداسازی تصویر چهره و محل قرارگیری آن از سایر اجزای تصویر است که باعث افزایش دقت ویژگی‌های استخراج شده می‌شود.
در این تحقیق استفاده از شکل صورت در تصاویر تمام رخ با تغییراتی به عنوان ابزاری برای یافتن محل چهره به کار رفته و همچنین معیار فاصله جدیدی برای حذف تصویر بدون چهره از یک تصویر دارای چهره ارائه شده است. محور بعدی، استخراج ویژگی‌هایی از چهره انسان است که از بیشترین کارایی در شناسایی تصاویر متنوع از یک چهره انسان برخوردار می‌باشند.

تست آنلاین تشخیص چهره توسط پروفسور جوزف بیگان (Josef Begun) در آدرس ذیل می‌تواند ذهنیت جالبی از اینگونه نرم‌افزارها در اختیار خواننده قرار دهد.
http://www2.hh.se/staff/josef/public/sa/face_recog_test

3- پردازش نقشه و تصاویر ماهواره‌ای:

از کاربردهای آن می‌توان در سیستمهای اطلاعات جغرافیایی- Geographical information system (GIS) نام برد.
در GIS واقعیت ها بصورت یکسری عارضه های جغرافیائی بر اساس دو نوع اطلاعات پایه ای (DATA ELEMENT) نمایش داده شده. اطلاعات پایه‌ای جغرافیایی (GEOGRAPHICAL DATA) برای تعیین موقعیت جغرافیائی عارضه و اطلاعات توصیفی (ATTRIBUTE DATA) برای تعریف مشخصات عارضه استفاده میشود. برای مثال : مرزهای جغرافیایی ، شبکه رودخانه ها و نقاط ارتفاعی همه عارضه های جغرافیائی اند که به ترتیب برای تعیین موقعیت املاک ، جریان آب رودخانه ها و پستی و بلندی ها مورد استفاده قرار می گیرند.
در GIS اطلاعات پایه جغرافیائی(GEOGRAPHICAL DATA ELEMET) مهم تر از اطلاعات پایه توصیفی (ATTRIBUTE ELEMENT)می باشد و این جنبه است که GIS را از سایر سیستمها متفاوت می سازد.

مختصات جغرافیائی صریح ، شامل طول و عرض جغرافیایی یا مختصات کارتزین است و مختصات مجازی نام یک واحد مدیریتی ، یک شهر، یک عارضه فیزیکی به مانند کوه و غیره می باشد .در بسیاری موارد مختصات صریح از مختصات مجازی مشتق میگردد و این مختصات جغرافیائی در اتصال مجموعه اطلاعات به یکدیگر بسیار مؤ ثر بوده و بیش از هر عامل دیگری ، دلیل بر موفقیت GIS گردیده است.

بسیاری از بسته های نرم افزاری دارای چند صد دستور و انواع کاربردهای عملی می باشند هم چنین برنامه ریزی و قابلیت نرم افزارهای GIS ی گوناگونند مانند:

- قابلیت بازیابی اطلاعات برای هر منطقه جغرافیائی از طریق نام منطقه ، طول و عرض جغرافیائی و یا هر ترکیبی از این دو عامل و یا تعیین محل هر منطقه نسبت به منطقه دیگر.
- داشتن امکانات تصمیم‌گیری بمنظور انتخاب مناسبترین عملیات برای اطلاعات موجود با توجه به هدف نهایی.
- امکانات لازم جهت تغییر سیستم مختصات از یک سیستم تصویر به سیستم دیگر.
- امکانات لازم جهت به روزرسانی اطلاعات جغرافیائی .
- استنتاج از جز به کل یا جمع بندی پارامترهای مربوط به یک یا چند صفت مربوط به یک منطق جعرافیایی و نشان دادن نتایج ان بصورت نقشه و یا جداول اطلاعاتی مورد نیاز.
- قابلیت ازکل به جز پارامترهای مربوط به یک یا چند صفت جغرافیایئ و یا غیر جغرافیائی و نشان دادن نتایج آن بصورت نقشه و یا جدول
- قابلیت انجام محاسبات اماری و تهیه اطلاعات در فرمتهای مختلف برای مدلهای اماری و شبیه سازیها مورد نیاز
- قابلیت ایجاد جدول ، گراف و هسیتوگرام ، نقشه و سایر روشهای جاری برای ارائه نتایج.
- قابلیت هدایت عملیات و محاسبات با استفاده از مقادیر مناسب برای پارامترهای مورد نیاز.
- قابلیت دریافت اطلاعات از سیستمهای مختلف و تهیه اطلاعات برای سیستمهای مختلف.
- قابلیت برقراری ارتباط بین اطلاعات جغرافیائی (نقشه) و اطلاعات غیر جغرافیایی (جداول اطلاعاتی) و ایجاد امکانات تجزیه و تحلیل اطلاعات جغرافیایی با استفاده از اطلاعات غیر جغرافیایی و بالعکس این قابلیت برقراری ارتباط بین یک جدول و نقشه ، دو نقشه از طریق یک جدول طبقه بندی اطلاعات یک نفشه بر اساس یک جدول و انجام محاسبات ریاضی بین هر تعداد نقشه را شامل می گردد.
- قابلیت محاسبه تعداد وقوع یک حادثه در فاصله مشخص از نقطه و یا نقاط معین:
- قابلیت محاسبه یک فرمول و یاعملگر در موقعیتهای مختلف.
- قابلیت روی هم قراردادن و انجام محاسبات نقشه های متعدد مختلف.
- قابلیت پیدا کردن اشیا مختلف با استفاده از خاصیت نزدیکی انها به یک شی خاص.
- قابلیت شبیه سازی و پیش بینی اثرات یک پروسه در طول زمان معین و با توجه به یک طرح و برنامه مشخص.

4- کاربرد در پزشکی :

1- جهت شبیه سازی اندازه گیری فواصل اپتیکی چشم
2- توسعه یک سیستم اندازه گیری سیستماتیک حرکات اندام به روش تصویر برداری ویدئویی
3- دستگاههای رادیولژی و تفسیر اشعه X
4- پردازش تصویر در سیستم های آندوسکوپی.
5- تفسیر الکتروکاردیوگرام

5- پردازش فضا:

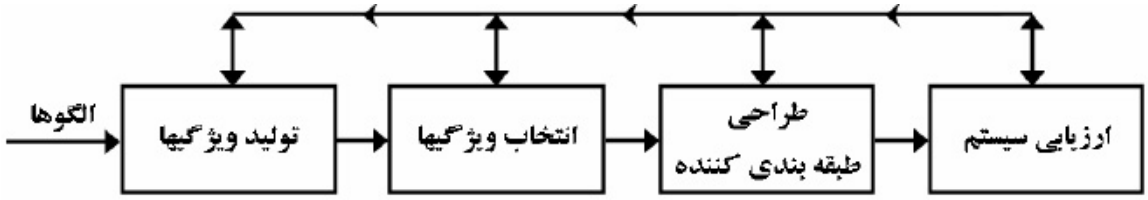
بهترین عکس‌های آماتوری از مریخ با استفاده از وِب‌کَم‌های متصل به‌تلسکوپ‌های ۸ اینچ یا بزرگتر تهیه می‌شود. با وب‌کم، فیلمی از سیاره گرفته می‌شود و بعد در نرم‌افزارهای ساده پردازش تصاویر نجومی مثل نرم‌افزار رایگانRegistax فریم‌های مناسب انتخاب و روی هم انداخته می‌شود و عکس خیره‌کننده‌ای حاصل می‌شود.

تشخیص الگو(pattern recognition):

شناسایی الگو شاخه‌ای از هوش مصنوعی است که با طبقه بندی (classification) و توصیف مشاهدات سروکار دارد. شناسایی الگو به ما کمک می‌کند تا داده‌های (الگوها) را با تکیه بر دانش قبلی یا اطلاعات آماری استخراج شده از الگوها طبقه بندی نماییم. الگوهایی که می‌بایست کلاسه بندی شوند معمولا گروههایی از سنجشها یا مشاهدات هستند که مجموعه نقاطی را در یک فضای چند بعدی مناسب تعریف می‌نماییم.

یک سیستم شناسایی الگوی کامل متشکل از یک حسگر (sensor) که مشاهداتی را که باید کلاسه‌بندی شوند را جمع‌اوری میکند، یک مکانیزم استخراج ویژگیها (feature extraction) که اطلاعات عددی یا سمبولیک را از مشاهدات محاسبه می‌کند و یک نظام کلاسه بندی که وظیفه اصلی طبقه بندی یا توصیف الگوها را با تکیه بر ویژگیهای استخراج شده عهده‌دار است.

شکل زیر بلوک دیاگرام یک سیستم شناسایی الگو را نشان می‌دهد.

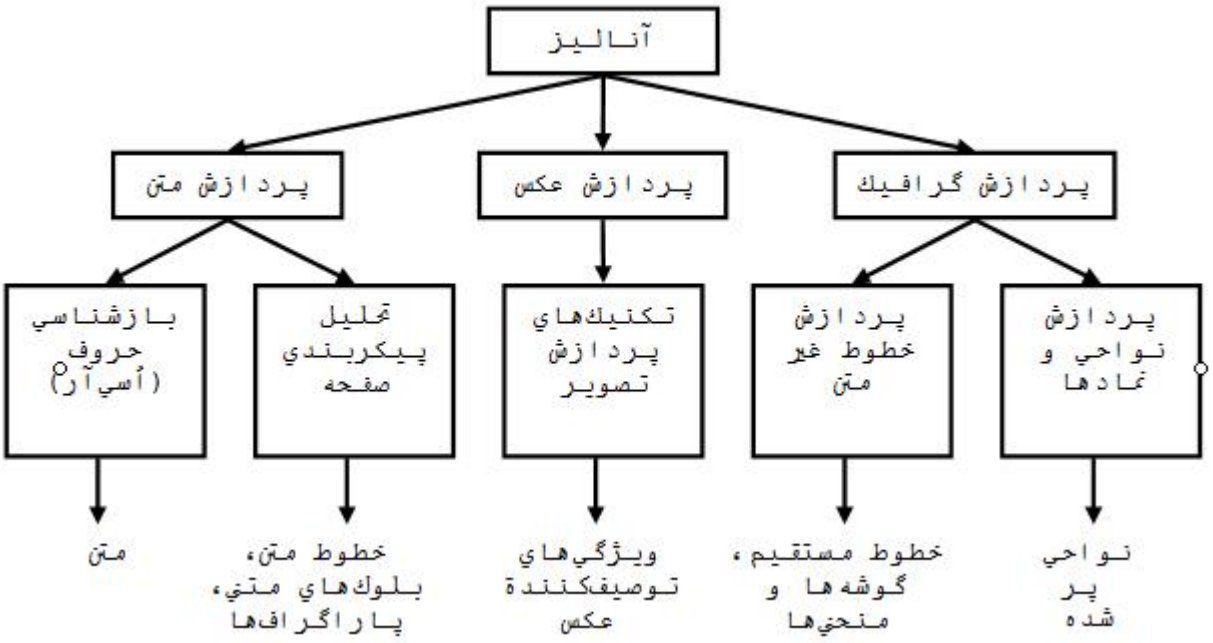


همانطور که از فلشهای برگشتی مشخص است این بلوکها لزوما مستقل نیستند و بسته به نتایج حاصله ممکن است لازم باشد که بلوکهای اولیه مجددا طراحی گردند تا راندمان کلی سیستم بهبود یابد.

نظام کلاسه بندی یا توصیف معمولا مبتنی بر وجود یک مجموعه از الگوهایی است که قبلا کلاسه بندی شده‌اند . این مجموعه را مجموعهٔ آموزشی (training set) گویند و قانون یادگیری منتج شده را قانون یادگیری با نظارت گویند.

مبحث «آنالیز تصویر اسناد» (DID) از جمله شاخه‌های فعال در شناسایی الگو و پردازش تصاویر می‌باشد و مشتمل بر کلیه مراحل پردازشی است که محتویات یک سند اسکن یا فکس شده را به یک فرم الکترونیکی مناسب، تبدیل

می‌نماید. تکنیک‌های DID اجزای مختلف ساختاری سند، یعنی قسمت‌های متنی (پاراگراف‌ها، کلمات، حروف، ...)، قسمت‌های گرافیکی (خطوط، نمادها، نمودارها، ...) و قسمت‌های تصویری (تصاویر موجود در متن) را از یکدیگر تفکیک می‌کنند و پردازش مناسب را بر روی هر دسته از اجزا، اعمال می‌نمایند و نیز با توجه به **ارتباط منطقی** بین اجزای مختلف، نقش هر یک از این اجزا را در سند مشخص می‌سازند. شکل زیر ساختار سلسله‌مراتبی DID را به نمایش می‌گذارد.



همانگونه که شکل فوق نشان می‌دهد، DID دربردارنده دسته بزرگی از تکنیک‌ها بنام تکنیک‌های بازشناسی نوری حروف OCR است. این تکنیک‌ها در مورد اجزایی از تصویر سند که توسط تکنیک‌های تحلیل ساختاری در DID به عنوان متن تشخیص داده شده‌اند، اعمال می‌گردند و تصویر سند را به یک متن قابل ویرایش توسط رایانه تبدیل می‌نمایند. سیستم‌های OCR با حذف نقش تایپیست‌ها در فرایند تبدیل اسناد کاغذی به قالب الکترونیکی، سرعت ورود اطلاعات به رایانه را دهها برابر افزایش می‌دهند و روند انجام این فرایند را به میزان قابل توجهی تسهیل می‌کنند. امروزه بازار مصرف سیستم‌های OCR، طیف بسیار وسیعی از مؤسسات (شامل مراکز نشر، دانشگاه‌ها، کتابخانه‌ها، بانک‌ها، ادارات پستی، شرکت‌های بیمه، و ...) را دربرمی‌گیرد. در نتیجه آشنایی اولیه با مبانی این سیستم‌ها برای کلیه افرادی که به نحوی با اسناد و مدارک سر و کار دارند، ضروری به نظر می‌رسد.

الگوریتمهای تشخیص الگو:

به منظور بررسی و انتخاب االگوریتمهای مناسب در مراحل مختلف شناسایی الگو (چهرهٔ انسان) لازم است آنها را بر روی تصاویری آزمایش نماییم . این تصاویر باید به نحوی باشند که بتوانند نقاط ضعف و قوت الگوریتمها را در شرایط مختلف نشان دهند و بهمین دلیل از حساسیت برخوردار هستند.

- **روش مینیمم هیستوگرام (mode):** در این روش ابتدا آستانه برابر مینیمم بین دو قله در هیستوگرام عکس موردنظر در نظر گرفته می‌شود. هیستوگرام تصاویر معمولا دارای سه قله هستند. یک قله مربوط به مو، قلهٔ دیگر که بین دوقله قرار دارد مربوط به صورت است و بسته به رنگ پوست افراد دارای سطوح خاکستری متفاوت می‌باشد بنابراین باید مینیمم بین قلهٔ دوم و سوم را بیابیم و بعنوان آستانه در نظر گیریم.

- **روش درصد دنباله هیستوگرام (p-tile):** در این روش فرض می‌کنیم می‌دانیم چند درصد از کل تصویر را شیء موردنظر اشغال کرده‌است. بنابراین آستانه را طوری درنظر می‌گیریم که درصد موردنظر از سطوح خاکستری در یک طرف آستانه قرار گیرد. بهمین دلیل این روش را Percentage tile (p-tile) می‌نامند.

- **روش برگشتی سد احمد (sid-ahmed):** این روش را می‌توان برای پیدا کردن اتوماتیک آستانه‌هایی بین تعدای سطح خاکستری در تصویر بکار برد. چنین روشی با توجه به آنکه برای هر تصویر چهره مقدار آسنانه‌ای مناسب آنرا پیدا می‌کند از دو روش قبل مناسبتر به نظر می‌رسد. در این روش تصاویر را سه‌سطحی درنظر می‌گیریم و دو آستانه توسط این روش محاسبه می‌نماییم . سطوح بین دو آستانه مربوط به صورت می‌باشند. از آنجاییکه می‌خواهیم زمینه را سفید و سر را سیاه کنیم، از آستانه بزرگتر برای Binary کردن تصویر استفاده می‌کنیم .

- روش ماکزیمم واریانس: در این روش آستانه‌ای پیدا می‌شود واریانس بین سطوح خاکستری شیء و زمینه با هم تداخل دارند و یا بعبارتی قله‌های هیستوگرام دو ناحیه مربوط به آنها روی هم می‌افتد، مفید می‌باشد.

همانطور که برای روش سداحمد ذکر شد، فرض می‌شود که تصویر سه سطحی می‌باشد. بهمبن دلیل در مرحله اول با اعمال الگوریتم به تصویر آستانه‌ای پیدا می‌کنیم که موها را متمایز کند در مرحله دوم الگوریتم را به بقیه هیستوگرام، یا بعبارتی سطوح خاکستری بزرگتر از آستانه بدست آمده در مرحله قبل، اعمال می‌نماییم و آستانه‌ای می‌یابیم که مرز بین سطوح خاکستری صورت و سطوح خاکستری زمینه است.

از این الگوریتم برای باینری کردن تصاویر استفاده می‌کنیم . به منظور سریع شدن الگوریتم به جای فرض اولیه زمینه بودن چهار گوشهٔ تصویر ، الگوریتم را با فرض یک مقدار اولیه برای آستانه شروع می‌کنیم و سپس الگوریتم را ادامه می‌دهیم.

ویژگیهای الگوریتم با روش ضریب خود همبستگی:

این ویژگیها با محاسبهٔ همبستگی تصویر با خودش محاسبه می‌شود و نسبت به انتقال ، دوران، و اندازهٔ شیء در تصویر مستقل هستند.
برای تصاویر باینری داریم :

$$A= \Sigma(y)\Sigma(x)\ f(x,y)$$

$$G(a\sqrt{A}\ \mathrm{Cos}\ t,\ a\sqrt{A}\ \mathrm{Sin}\ t) = 1/A\ \Sigma(y)\Sigma(x)\ f(x,y).\ F(x+a\sqrt{A}\ \mathrm{Cos}\ t,\ y+a\sqrt{A}\ \mathrm{Sin}t)$$

با استفاده از رابطهٔ 2 می‌توان برای زوایای t و مقادیر جابجایی مختلف a بردار ویژگی G را محاسبه نمود

آزمایشها نشان داده‌اند که در صورت حداقل بودن تعداد تصاویر یک فرد نرخ شناسایی بسیار بالاست. اما تغییرات نور و همچنین زمینه در مورد چهره و هم در مورد متون تاثیر بسزایی در تشخیص نرم‌افزار هوشمند خواهد داشت.

پردازش متن (Optical Character Recognition)

رایانه تنها یک ابزار است، و ما انسان‌ها می‌باید شیوهٔ کار کردن را به او بیاموزیم، و امکانات کار را هم برایش فراهم آوریم، نرم‌افزارهای مختلف به همین منظور طراحی می‌شوند. یکی از آنها نرم‌افزار OCR است که برای تشخیص و بازیابی الفبا (نوشته‌های دست‌نویس یا تایپ‌شده) توسط کامپیوتر طراحی می‌شوند.

OCR سرنام اصطلاحی است که صورت کامل آن در واژه‌نامه‌های انگلیسی به دو صورت آمده است:

- Optical Character Recognition
- Optical Character Reader

در چند دهه گذشته بازشناسی الگوهای نوشتاری شامل حروف، ارقام و دیگر نمادهای متداول در اسناد نوشته‌شده 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 به منظور وارد نمودن اطلاعات موجود در اسناد، مدارک، کتاب‌ها و سایر مکتوبات تایپی و حتی دست‌نوشت به داخل رایانه شده است. اصطلاح OCR به تکنیک‌هایی اطلاق می‌شود که در تصاویر اسکن یا فکس شده، نواحی متنی را تشخیص می‌دهند و سپس این نواحی (تصویری) را به متن قابل ویرایش تبدیل می‌نمایند.

یک نرم‌افزار OCR، تصویر اسکن‌شده را می‌خواند، محتویات آن (شامل متن، خطوط، تصاویر، جداول، ...) را شناسایی می‌نماید، و سپس آن را به یک قالب قابل ویرایش (در واژه‌پردازها) تبدیل می‌کند. امروزه بیشتر دستگاه‌های اسکنر به نرم‌افزارهای OCR مجهز گردیده‌اند و قادرند متن موجود در یک سند اسکن‌شده را تشخیص دهند و آن را با همان نحوه قالب‌بندی، ستون‌بندی، جدول‌بندی و نوع فونت مطابق با سند کاغذی اصلی، در قالب یک فایل متنی با قالب‌بندی مناسب ذخیره نمایند.

استفاده از سیستم‌های OCR دو مزیت عمده دارد:

الف. افزایش چشمگیر سرعت دسترسی به اطلاعات؛ زیرا در متن بر خلاف تصویر، امکان جستجو و ویرایش وجود دارد.
ب. کاهش فضای ذخیره‌سازی؛ زیرا حجم فایل متنی استخراج‌شده از یک تصویر، معمولاً بسیار کمتر از حجم خود فایل تصویری است.

چنین قابلیتی امکان استفاده گسترده از رایانه را در پردازش سریع حجم وسیعی از داده‌های مکتوب شرکت‌ها و مؤسسات مختلف (نظیر بانک‌ها، شرکت‌های بیمه، مؤسسات خدمات عمومی، اداره پست، و دیگر نهادهایی که سالانه با میلیون‌ها مورد پرداخت، دریافت و حسابرسی امور مشتریان خود مواجه‌اند فراهم می‌آورد.

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

الف. مرحله تکوین (از 1900 تا 1980): رد پای اولیه اقدامات صورت‌گرفته در زمینه بازشناسی حروف را در سال‌های اول دهه 1900 می‌توان یافت و آن زمانی است که «تیورینگ» دانشمند روسی بر آن بود که به افراد مبتلا به نارسایی‌های بینایی کمک نماید. اولین اختراع‌های ثبت‌شده در این زمینه مربوط به سال‌های 1929 و 1933 میلادی هستند. این سیستم‌ها حروف چاپی را با روش تطابق قالبی شناسایی می‌کردند؛ به این صورت که ماسک‌های مکانیکی مختلفی از مقابل تصویر حرف عبور می‌کردند (مکانیکی) و نور از یک سو به آن تابانده می‌شد و از سوی دیگر توسط یک آشکارساز نوری دریافت می‌گردید (اپتیکی). وقتی یک انطباق کامل صورت می‌گرفت، نور به آشکارساز می‌رسید و حرف ورودی بازشناسی می‌شد. این اختراع به دلیل فناوری اپتومکانیکی مورد استفاده در آن، کاربردی نبود. تصور دسترسی به دستگاهی برای بازشناسی حروف تا دهه 1940 میلادی و ظهور رایانه‌های دیجیتال، به صورت یک رؤیا باقی ماند.

اقدامات اولیه در زمینه بازشناسی حروف، بر متون چاپی یا مجموعه کوچکی از حروف و نمادهای دستنوشت که براحتی قابل تشخیص بودند، متمرکز گردیده بود. سیستم‌های بازشناسی حروف چاپی که در این مقطع زمانی عرضه شدند، عمدتاً از روش تطابق قالبی استفاده می‌نمودند که در آن، تصویر ورودی با مجموعه بزرگی از تصاویر حروف، مورد مقایسه قرار می‌گرفت. در مورد متون دستنوشت نیز الگوریتم‌های پردازش تصویر که ویژگی‌های سطح پایین (ویژگی‌هایی که مستقیماً و بدون اعمال هیچ تبدیلی، از تصاویر استخراج می‌شوند) را از تصاویر استخراج می‌کنند، در مورد تصاویر دوسطحی اعمال می‌شدند تا بردارهای ویژگی استخراج گردند. سپس این بردارهای ویژگی به طبقه‌بندی‌کننده‌های آماری سپرده می‌شدند.

در این دوره، تحقیقات موفق اما مقید (منظور از مقید، مفروض دانستن شرایط و پیش‌فرض‌های خاص برای کاراکترهای ورودی است)، بیشتر بر روی حروف و اعداد لاتین انجام گرفت. با این حال مطالعات چندی نیز بر روی حروف ژاپنی، چینی، عبری، هندی، سیریلیکی، یونانی و عربی در هر دو زمینه حروف چاپی و دستنوشت آغاز گردید. با ظهور صفحات رقومی‌کننده در دهه 1950 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 بازشناسی برخط حروف دستنوشت، فعالیت خود را آغاز نمایند. «Suen، 1992» یک منبع مناسب درباره اقدامات صورت‌گرفته بر روی بازشناسی برخط حروف تا سال 1980 می‌باشد.
ب. مرحله توسعه (از 1980 تا 1990): مطالعات صورت گرفته تا قبل از سال 1980 بدلیل فقدان سخت‌افزارهای رایانه‌ای قدرتمند و دستگاه‌های اخذ داده‌ها با مشکل همراه بودند. در این دهه بواسطه رشد انفجارگونه فناوری اطلاعات، وضعیت بسیار مناسبی برای تحقیقات مختلف از جمله بازشناسی حروف فراهم گردید. روش‌های ساختاری به همراه روش‌های آماری در بسیاری از سیستم‌ها استفاده شدند. تحقیقات در زمینه OCR اساساً به توسعه روش‌های

بازشناسی معطوف گردید، بی آنکه مسئله استفاده از اطلاعات معناشناختی به منظور افزایش دقت بازشناسی مورد توجه قرار گیرد. این امر سبب گردید که دقت بازشناسی (نرخ بازشناسی) از یک حد خاص فراتر نرود، که در بسیاری از کاربردهای OCR، قابل قبول نبود. مروری بر تحقیقات و پیشرفت‌های حاصل‌شده در مورد OCR در این برهه زمانی را می‌توان در «Mori، 1992» و «Suen، 1992»، بترتیب برای بازشناسی برخط و برون‌خط ، جستجو نمود.
ج. مرحله بهبود (از 1990 به بعد): در این مقطع زمانی بود که با تکوین ابزارها و تکنیک‌های پردازشی جدید، پیشرفت واقعی در سیستم‌های OCR محقق گردید. در اوایل دهه 90، روش‌های پردازش تصویر و بازشناسی الگو با تکنیک‌های کارآمد هوش مصنوعی ادغام گشتند. محققان، الگوریتم‌های پیچیده‌ای را در بازشناسی حروف ابداع نمودند که قادر بودند داده‌های ورودی با تفکیک‌پذیری بالا را دریافت کنند و در مرحله پیاده‌سازی، محاسبات بسیار زیادی را بر روی داده‌ها انجام دهند. امروزه علاوه بر وجود رایانه‌های قدرتمندتر و تجهیزات الکترونیکی دقیق‌تر مانند اسکنرها، دوربین‌ها و صفحات رقومی‌کننده، استفاده از تکنیک‌های پردازشی مدرن و توانمند همچون شبکه‌های عصبی ، مدل‌های مارکوف پنهان ، منطق‌های مجموعه فازی و مدل‌های پردازش زبان طبیعی امکان‌پذیر گشته است.
جدول زیر نشاندهندهٔ جایگاه کنونی تحقیقات در زمینه سیستم های OCR لاتین است که نیازمند تحقیقات بیشتر در این زمینه است.

		متون چاپی			متون دستنوشت		
		یک نوع فونت	چند نوع فونت	همه نوع فونت	گسسته	پیوسته	مخلوط
برخط	مقید						
	نامقید						
برون‌خط	بدون نویز						
	نویزی						

سیستم‌های جدید OCR برون‌خط متون چاپی و برخط متون دستنوشت با واژگان محدود و وابسته به نویسنده، در کاربردهای محدود به نحو کاملاً رضایت‌بخشی عمل می‌کنند. اما به منظور دستیابی به هدف نهایی در شبیه‌سازی ماشینی نگارش انسانی و متون چاپی، هنوز راه درازی در پیش است. «آریکا» برخی از تکنیک‌های عرضه‌شده در این دوره را مرور کرده‌اند. جدول فوق جایگاه کنونی پیشرفت‌های حاصل‌شده در زمینه سیستم‌های OCR برای متون لاتین را به نمایش می‌گذارد. توجه شود که برای متون چاپی، پردازش برخط تعریف نمی‌شود.

4. تحقیقات انجام‌شده در داخل کشور در زمینه تولید OCR فارسی با توجه به اهمیت طیف وسیع کاربردها و نیاز شدید بازار تجاری، در سال‌های اخیر تحقیقات قابل ملاحظه‌ای در کشور در زمینه OCR توسط دانشگاه‌ها، برخی نهادهای دولتی، و شرکت‌های خصوصی صورت گرفته است که متاسفانه از آمار دقیق آن‌ها اطلاعی در دست نیست. اما قدر مسلم این که برای OCR متون چاپی تاکنون هیچ نرم‌افزار کارآمد OCR تجاری که محصول تحقیقات داخل کشور باشد، عرضه نگردیده است. در ادامه به برخی از تلاش‌های صورت گرفته در این زمینه اشاره می‌شود:
- در حوزه تحقیقات دانشگاهی، تعداد نسبتاً زیادی پایان‌نامه (بخصوص در مقاطع کارشناسی ارشد و دکتری) و مقاله در این زمینه منتشر شده‌اند که نقطه تمرکز بیشتر آن‌ها، ارائه روش‌هایی به منظور قطعه‌بندی درونی ، بازنمایی و بازشناسی حروف بوده است و سایر بخش‌ها شامل پیش‌پردازش ، قطعه‌بندی بیرونی و پس‌پردازش کمتر مورد توجه قرار گرفته‌اند . بخش‌های مختلف پردازشی یک سیستم OCR شامل پیش‌پردازش، قطعه‌بندی، بازنمایی، بازشناسی و پس‌پردازش مورد بررسی قرار خواهند گرفت.

انواع OCR:

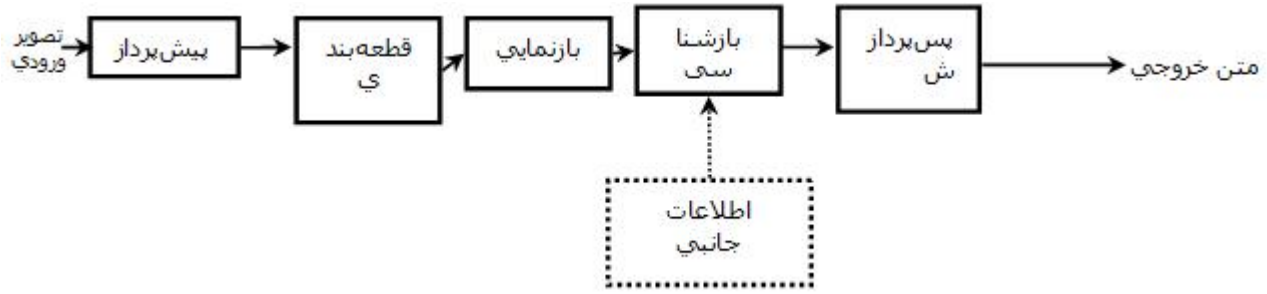
انواع سیستم‌های OCR از لحاظ نوع الگوی ورودی

سیستم‌های OCR را می‌توان از لحاظ نوع الگوی ورودی به دو گروه اصلی تقسیم کرد:
الف. سیستم‌های برخط،
ب. سیستم‌های برون‌خط.

در بازشناسی برخط، حروف در همان زمان نگارش توسط سیستم تشخیص داده می‌شوند و 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 رقومی‌کننده 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 دوبعدی متن ورودی استفاده می‌شود. در این روش به هیچ نوع وسیله نگارش خاصی نیاز نیست و تفسیر داده‌ها مستقل از فرآیند تولید آن‌ها و تنها براساس تصویر متن صورت می‌گیرد. این روش به نحوه بازشناسی توسط انسان شباهت بیشتری دارد.

معرفی بخش‌های مختلف یک سیستم OCR

شکل زیر نمودار بلوکی یک سیستم OCR را نمایش می‌دهد. لازم به ذکر است که بسته به الگوریتم کلی به کار رفته و سطح انتظارات از عملکرد نرم‌فزار، ممکن است برخی سیستم‌ها فاقد یک یا چند مرحله از مراحل فوق باشند.



نمودار بلوکی دیاگرام یک سیستم OCR

در ادامه هر یک از این بلوک‌ها مورد بررسی قرار می‌گیرند.

1. پیش‌پردازش:

کلیه اعمالی که روی تصویر صورت می‌گیرند تا موجب تسهیل در روند اجرای فازهای بعدی گردد؛ مانند دوگانی‌کردن تصویر، حذف نویز ، هموارسازی ، نازکسازی ، تشخیص زبان و فونت کلمات، و نظایر این‌ها. از مجموعه این پردازش‌ها، هدف‌های زیر دنبال می‌شود:

1. کاهش نویز،
2. نرمالیزه کردن داده‌ها،
3. فشرده‌سازی میزان اطلاعاتی که می‌بایست محفوظ بماند.

کاهش نویز: نویز ایجادشده بواسطه دستگاه‌های اسکنر نوری منجر به ایجاد نقطه نقطه‌های لک‌مانند ، قطعه خط‌های گسسته ، اتصال بین خطوط، فضاهای خالی در خطوط متن، پرشدن حفره‌های موجود در تصویر برخی حروف، و ... می‌گردد. همچنین اعوجاج‌های مختلف شامل تغییرات محلی، منحنی‌شدن گوشه‌های حروف، تغییر شکل یا خوردگی حروف را نیز باید در نظر داشت. قبل از مرحله بازشناسی حروف، لازم است که این نقایص برطرف شوند. مهم‌ترین دلیل برای کاهش نویز، کم‌کردن خطا در مراحل قطعه‌بندی و بخصوص بازشناسی می‌باشد. کاهش

نویز همچنین سبب کم‌شدن اندازه فایل تصویر می‌شود که به نوبه خود، کاهش زمان مورد نیاز برای پردازش‌ها و ذخیره‌سازی‌های آینده را در پی خواهد داشت.

نرمالیزه‌کردن: روش‌های نرمالیزه‌کردن داده‌ها به حذف تغییرات نگارشی کمک می‌کند و داده‌های استانداردشده‌ای را نتیجه می‌دهد. روش‌های پایه نرمالیزه‌کردن عبارت‌اند از:

الف. نرمالیزه‌کردن کجی متن و استخراج خطوط زمینه : به دلیل بی‌دقتی در مرحله اسکن یا بی‌دقتی نویسنده در هنگام نگارش متن دست‌نوشت، ممکن است خطوط متن نسبت به تصویر، اندکی انحراف یا چرخش داشته باشند (شکل 5). این وضع ممکن است کارآیی الگوریتم‌های به کار رفته در طبقات بعدی سیستم OCR را تحت تأثیر قرار دهد؛ چرا که یکی از مفروضات در بیشتر روش‌های قطعه‌بندی، کج‌نبودن 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» در نگارش لاتین و «.» (نقطه) و «0» (صفر) در نگارش فارسی را بواسطه موقعیت نسبی‌شان نسبت به خط زمینه می‌توان آشکار ساخت.

کلیه الگوریتم‌های توسعه داده شده برای آشکارسازی کجی صفحه، بر روی صفحات متنی با ترازبندی یکنواخت، دقیق عمل می‌کنند. الگوریتمی کارآتر است که به واسطه حضور مواردی نظیر گرافیک، پاراگراف‌های دارای کجی متفاوت، اعوجاج‌های منحنی- خطی ظاهرشونده در کتاب‌های فتوکپی‌شده، نواحی وسیع پیکسل‌های سیاه نزدیک حاشیه صفحه و خطوط متنی مختصر و کوتاه، دقت آن کمتر دستخوش تغییر شود.

روش‌ها و الگوریتمهای به کار رفته برای تصحیح کجی خطوط زمینه در متون لاتین را می‌توان به چهار گروه اصلی دسته‌بندی کرد که عبارت‌اند از:

1. **به کارگیری هیستوگرام (پروفایل تصویرنمایی) تصویر**
2. **استفاده از روش خوشه‌بندی نزدیک‌ترین همسایه‌ها**
3. **روش همبستگی متقابل بین حروف** (اشاره شده در صفحات قبل)
4. **تبدیل کد هاف .**

اغلب پس از آشکارسازی کجی، تصویر صفحه در جهت اصلی چرخانده می‌شود تا عملیات تحلیل قالب‌بندی متن و OCR با سهولت و دقت بیشتری انجام پذیرد. نمونه‌برداری مجدد مورد نیاز برای این منظور که باید بر روی صفحات دوگانی‌شده اعمال گردد، ممکن است الگوی کاراکترها را تغییر دهد. در این حالت به جای چرخاندن تصویر می‌توان

الگوریتم‌های پردازشی را به نحوی اصلاح نمود که اثر چرخش در آن‌ها لحاظ گردد. همچنین می‌توان تصویر سند را قبل از دوگانی‌کردن، چرخش داد یا این که مقدار چرخش را از روی انتقال‌های کوچک و بدون اعوجاج کل بلوک‌های متنی، تقریب زد.

ب. نرمالیزه‌کردن اریب‌شدگی : در متون چاپی فارسی و لاتین، کاراکترهای دارای قالب ایتالیک از راستای عمود، انحراف دارند. در متون دست‌نوشت نیز برخی از نویسنده‌ها حروف را به صورت زاویه‌دار می‌نویسند. این پدیده با عنوان «اریب‌شدگی» شناخته می‌شود و ممکن است دقت برخی از الگوریتم‌های قطعه‌بندی یا بازشناسی را تحت تأثیر قرار دهد و و از این رو در این سیستم‌ها لازم است که در مرحله پیش‌پردازش، میزان اریب‌بودن کاراکترها شناسایی و تصحیح گردد.

ج. نرمالیزه‌کردن (تغییر مقیاس دادن) اندازه: در سیستم‌هایOCR، اغلب تصاویر کلمات یا حروف خیلی کوچک یا خیلی بزرگ، به یک اندازه استاندارد نرمالیزه می‌شوند تا بدین ترتیب عملیات بازشناسی، مستقل از اندازه فونت متن گردد.

د. هموارسازی کانتور : خط تشکیل‌دهنده مرز یک کاراکتر را کانتور آن کاراکتر گویند. در متون دست‌نوشت، به واسطه لرزش یا حرکات ناخواسته دست نویسنده در هنگام نگارش، ممکن است که کانتور حروف ناصاف شود. این وضع در سیستم‌های بازشناسی متون چاپی و دست‌نوشت نیز، به دلیل تغییر مقیاس حروف یا وجود نویز در مرحله اسکن تصاویر ممکن است ظاهر گردد. روش‌های هموارسازی کانتور، به منظور جبران این نقیصه مورد استفاده قرار می‌گیرند. به طورکلی هموارسازی کانتور، تعداد نقاط نمونه مورد نیاز برای نمایش کاراکتر را کاهش می‌دهد و در نتیجه کارآیی مراحل پردازشی باقیمانده را بهبود می‌بخشد.
فشرده‌سازی: این نکته پذیرفته شده است که تکنیک‌های کلاسیک فشرده‌سازی تصاویر که تصویر را از حوزه مکانی به حوزه‌های دیگر منتقل می‌کنند، برای بازشناسی حروف مناسب نیستند. در بازشناسی حروف، عمل فشرده‌سازی نیازمند آن دسته از تکنیک‌های حوزه مکانی است که اطلاعات شکلی را حفظ می‌نمایند. دو تکنیک متعارف فشرده‌سازی، یکی تکنیک اِعمال سطح آستانه (به منظور دوگانی یا دوسطحی کردن تصاویر سطح خاکستری) و دیگری نازک‌سازی می‌باشند.

الف. دوگانی (دوسطحی) کردن تصویر متن: تصاویر دیجیتالی به یکی از سه صورت- تصاویر رنگی، تصاویر سطح خاکستری (مشابه تصویر یک تلویزیون سیاه و سفید که رنگ تصویر به صورت سیاه، سفید و طیفی از رنگ‌های خاکستری ظاهر می‌شود)، و تصاویر دوگانی یا دوسطحی (مشابه تصویر یک سند فکس‌شده که رنگ پیکسلهای تصویر، تنها سیاه یا سفید است)- می‌باشند. به منظور کاهش حجم ذخیره‌سازی مورد نیاز و افزایش سرعت و سهولت

پردازش، اغلب مطلوب است که با انتخاب یک سطح آستانه، تصاویر سطح خاکستری یا رنگی را به تصاویر دوگانی تبدیل نمود.
ب. نازک‌سازی: با این عمل، تصویر کاراکترها به تصویری با عرض یک پیکسل تبدیل می‌شود؛ درست مثل این که کاراکترها با یک قلم نوک باریک نوشته شده باشند. نازک‌سازی در حالی که کاهش قابل‌ملاحظه‌ای در حجم داده‌ها ایجاد می‌کند، اطلاعات شکلی کاراکتر را نیز حفظ می‌نماید.

شکل 6. اعمال عملیات نازک‌سازی بر روی یک تصویر متنی نمونه

بازشناسی خط ، زبان و فونت: بازشناسی خط، تعداد کلاس‌های مختلف نمادهایی را که باید مورد ملاحظه قرار گیرند کاهش می‌دهد. شناسایی زبان متن، به منظور به کارگیری مدل‌های متنی خاص ضرورت دارد. طبقه‌بندی فونت‌ها، کاهش تعدد شکل‌های مختلف حروف در هر کلاس که لازم است در فرایند بازشناسی لحاظ گردند را به دنبال دارد و سبب می‌شود که امر شناسایی، تنها به یک کلاس فونت محدود گردد. بازشناسی خط و زبان و فونت در کاربردهایی مانند نمایه‌سازی و دستکاری اسناد نیز مطلوب می‌باشد.

2. قطعه‌بندی (جداسازی)

قطعه‌بندی مرحله‌ای بسیار مهم برای سیستم‌های OCR مخصوصاً OCR فارسی و عربی (که حروف کلمات به صورت سرهم نوشته می‌شوند) می‌باشد. قطعه‌بندی به دو گونه تقسیم می‌شود:
1. قطعه‌بندی بیرونی که عبارت است از تفکیک قسمت‌های مختلف تصویر نظیر متن، گرافیک و خطوط و نیز جدا کردن بخش‌های مختلف متن مانند پاراگراف‌ها، جملات یا کلمات؛

2. قطعه‌بندی درونی که منظور از آن، جداسازی حروف کلمات مخصوصاً در مورد کلمات سر هم نوشته‌شده در متون لاتین، یا در رسم‌الخط‌های پیوسته نظیر فارسی و عربی است. همچنین حروفی که در متن اصلی جدا بوده‌اند، امَا به خاطر کیفیت پایین دستگاه اسکنر به هم چسبیده‌اند، توسط این دسته از تکنیک‌ها از یکدیگر جدا می‌شوند.

شکل 7. قطعه‌بندی یک کلمه به حروف

مرحله قطعه‌بندی بیرونی، بحرانی‌ترین و حساس‌ترین قسمت در حوزه تحلیل تصویر اسناد می‌باشد و یک مرحله ضروری برای سیستم‌های OCR برون‌خط محسوب می‌شود. گرچه مبحث تحلیل اسناد با روش‌ها و تکنیک‌های خاص خود یک حوزه تحقیقاتی تا حدی متفاوت نسبت به OCR است، امَا تقسیم‌بندی تصویر سند به نواحی متنی و غیرمتنی، یک بخش لاینفک در نرم‌افزارهای OCR به حساب می‌آید. بنابراین برای افرادی که در زمینه OCR تحقیق می‌نمایند، داشتن دانش عمومی از تکنیک‌های آنالیز اسناد ضرورت دارد.
تحلیل پیکربندی صفحات در سه مرحله انجام می‌گیرد: مرحله اول تفکیک نواحی متنی در تصویر از نواحی غیرمتنی (شامل گرافیک و خطوط) است. مرحله دوم، تحلیل ساختاری است که با قطعه‌بندی نواحی متنی به بلوک‌هایی از تصویر سند (نظیر پاراگراف، ردیف، کلمه، و...) مرتبط می‌باشد. مرحله سوم که تحلیل عملکردی نام دارد، با استفاده از اطلاعات مکانی، اندازه و قوانین مختلف صفحه‌بندی، عملکرد هر یک از اجزای سند (نظیر عنوان، چکیده، و ...) را تعیین می‌نماید.

نقطه تمایز اصلی میان سیستم‌های OCR لاتین و فارسی برای متون چاپی، در مرحله قطعه‌بندی درونی نهفته است؛ چرا که حروف کلمات در نگارش فارسی برخلاف نگارش رسمی لاتین به صورت سرهم نوشته می‌شوند و در نتیجه ضرورت انجام صحیح این مرحله برای متون فارسی و عربی نسبت به متون لاتین، اهمیت فوق‌العاده بیشتری دارد. با وجود فعالیت‌های نسبتاً چشمگیر دهه گذشته و تنوع تکنیک‌های عرضه‌شده، قطعه‌بندی متون پیوسته (بخصوص متون دست‌نوشت پیوسته) به حروف، هنوز هم یک مسئله قابل بررسی مانده است.

الگوریتمهای قطعه‌بندی حروف به سه دسته تقسیم می‌شوند:

1. **الگوریتمهای قطعه‌بندی صریح**
2. **الگوریتمهای قطعه‌بندی ضمنی**
3. **الگوریتمهای ادغام‌شده.**

در مواردی همچون متون فارسی که حروف به صورت سرهم نوشته می‌شوند، سه رویکرد مختلف در بازشناسی برون‌خط متون کلمات یا زیرکلمات وجود دارد:

1. رویکرد مبتنی بر قطعه‌بندی کلمات
2. رویکرد مبتنی بر بازشناسی کلمه به عنوان یک الگوی واحد
3. رویکرد ترکیبی

در رویکرد بازشناسی مبتنی بر قطعه‌بندی، ابتدا کلمه در مرحله جداسازی به حروف یا زیرحروف، شکسته می‌شود؛ آنگاه قطعات جداشده بازشناسی می‌شوند و از کنار هم قرارگرفتن آن‌ها، کلمه شناسایی خواهد شد. روش‌های به کار گرفته‌شده در این رویکرد به دو گروه مختلف تقسیم می‌شوند:

- تقطیع کلمه به حروف
- تقطیع کلمه به زیرحروف

در گروه اول، کلمه به حروف جداسازی می‌شود و با شناسایی حروف جداشده، کلمه بازشناسی می‌گردد.
در گروه دوم، کلمه به زیرحروف مثل پاره‌منحنی‌ها و ساختارهای پایه دیگر جداسازی می‌شود و با شناسایی زیرحروف‌ها و ترکیب آن‌ها، کلمه بازشناسی می‌گردد. در این رویکرد نمی‌توان در ابتدا مرز حروف را به طور کامل مشخص کرد، بلکه حروف به ترتیب از ابتدا به انتهای کلمه، بازشناسی و جداسازی می‌شوند. در هیچک از دو رویکرد نخست که مبتنی بر جداسازی هستند، 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.

قطعه‌بندی غلط کاراکترها، عامل بسیاری از خطاهای OCR است (مانند: **rn→m** یا **m→rn**). میزان دقت یک الگوریتم قطعه‌بندی به سبک نگارش حروف، کیفیت دستگاه چاپ (کاراکترهای ایتالیک لکه‌دار برای قطعه‌بندی دارای اشکال می‌باشند)، و نیز نسبت اندازه فونت به قدرت تفکیک دستگاه اسکنر (تابع گسترش نقاط و نرخ نمونه‌برداری مکانی) بستگی دارد.

نتیجه مطلوب مرحله قطعه‌بندی، تصویری است که تنها حاوی یک کاراکتر باشد و بجز پیکسل‌های پس‌زمینه، هیچ شیء دیگری را شامل نشود. اما هنگامی که اشیای چاپی، در تصویر ورودی خیلی نزدیک به هم ظاهر شوند (مانند نقشه‌های هیدروگرافی)، این منظور همواره قابل حصول نخواهد بود. غالباً در چنین حالتی دیگر کاراکترها یا اشیای چاپی، به طور تصادفی در داخل تصویر کاراکتر قرار می‌گیرند و احتمالاً ویژگی‌های استخراج‌شده را تحریف می‌نمایند. این مورد یکی از دلایلی است که بیان می‌دارد چرا هر سیستم بازشناسی حروف، یک گزینه وازدگی دارد.

3. بازنمایی (استخراج ویژگی‌ها) :

این مرحله یکی از مراحل بسیار با اهمیت در سیستم‌های OCR است؛ چرا که نتایج حاصل از این مرحله، مستقیماً بر روی کیفیت مرحله بازشناسی اثر می‌گذارد. در مرحله بازنمایی، به هر الگوی ورودی (کاراکتر یا کلمه- بر حسب آن که رویکرد سیستم، مبتنی بر کدامیک از دو گروه «قطعه‌بندی کلمات» یا «شناسایی کلمه به عنوان یک الگوی واحد» باشد)، یک کد یا بردار ویژگی نسبت داده می‌شود که معرف آن الگو در فضای ویژگی‌ها است و آن را از دیگر الگوها متمایز می‌سازد.

در انتخاب بردارهای ویژگی لازم است موارد زیر مورد توجه قرار گیرند:

1. بردار ویژگی هر الگو باید تا حد امکان از بردارهای ویژگی دیگر الگوها متمایز باشد (فاصله بین بردارهای ویژگی در فضای ویژگی‌ها، حداکثر باشد).
2. بردار ویژگی الگوها باید تا بیشترین حد ممکن، خصوصیات شکل و ساختار الگوها را از تصویر آن‌ها استخراج نماید.
3. تا حد امکان نسبت به نویز، تغییر اندازه و نوع فونت، چرخش، و دیگر تغییرات احتمالی الگوها دارای ثبات باشد.
4. شرایط، نوع و خصوصیات الگوهای ورودی در انتخاب بردارهای ویژگی اثر می‌گذارند. به عنوان مثال، باید تعیین نمود که آیا حروف یا کلماتی که می‌بایست تشخیص داده شوند جهت و اندازه مشخصی دارند یا خیر، دست‌نوشت هستند یا چاپی، یا این که تا چه حد بوسیله نویز، مغشوش شده‌اند. همچنین گاهی کفایت می‌کند که سیستم، تنها جوابگوی گروه محدودی از الگوها (مثلاً الگوهایی با اندازه یا نوع فونت از پیش مشخص شده) باشد.
5. در مورد حروفی که به چندین شکل نوشته می‌شوند (مانند ‘ ’ و ‘’a، 4»» و «4») لازم است که بیش از یک کلاس الگو به یک کاراکتر خاص تعلق یابد.

همانطور که عنوان شد، بازنمایی یک مرحله بسیار مهم در حصول راندمان مناسب برای سیستم‌های بازشناسی حروف است؛ ولی برای دستیابی به عملکرد بهینه، لازم است که دیگر مراحل نیز بهینه گردند و باید توجه نمود که این مراحل، مستقل از هم نیستند. یک روش خاص استخراج ویژگی‌ها، طبیعت خروجی مرحله پیش‌پردازش را به ما دیکته می‌کند یا حداقل ما را در انتخابمان محدود می‌سازد.

مراحل قطعه‌بندی و بازشناسی، دو وجه تمایز عمده میان سیستم‌های OCR فارسی و لاتین می‌باشند. بواسطه وجود تفاوت‌های اساسی بین نحوه نگارش فارسی و لاتین، امکان اعمال مستقیم تکنیک‌های قطعه‌بندی و بازشناسی مربوط به سیستم‌های OCR لاتین، برای متون فارسی وجود ندارد. پیچیدگی‌های مختص نگارش فارسی، بر پیچیدگی الگوریتم‌های این دو مرحله می‌افزاید. درست به همین دلیل است که بیشتر نرم‌افزارهای OCR تجاری لاتین، قادر به پشتیبانی زبان فارسی و عربی نمی‌باشند.

4. طبقه‌بندی و بازشناسی (با یک یا چند طبقه‌بندی‌کننده):

این مرحله شامل روش‌هایی برای متناظرساختن هر یک از الگوهای به دست‌آمده از مرحله استخراج ویژگی‌ها، با یکی از کلاس‌های فضای الگوهای مورد بحث است که از طریق کمینه ساختن فاصله بردار ویژگی‌های هر الگوی ورودی نسبت به یکی از بردارهای مرجع، انجام می‌گیرد. بردارهای مرجع، بردارهایی هستند که قبلاً از نمونه‌های آموزشی اخذ شده‌اند. تکنیک‌های عرضه‌شده برای این مرحله را می‌توان در روش‌های مربوط به چهار گروه عمومی مبحث شناسایی الگو، جستجو نمود:

- تطابق قالبی
- تکنیک‌های آماری
- تکنیک‌های ساختاری
- شبکه‌های عصبی

چهار گروه فوق لزوماً مستقل یا مجزا از یکدیگر نمی‌باشند و گاهی می‌توان تکنیک‌های یک گروه را در میان تکنیک‌های مربوط به دیگر گروه‌ها یافت.

5. به کارگیری اطلاعات جانبی (پس‌پردازش):

در این مرحله با استفاده از اطلاعات جانبی (نظیر مجموعه لغات معتبر، اطلاعات آماری مربوط به رخداد حروف، اطلاعات دستوری و معنایی) سعی در بهبود نتایج حاصل از مرحله بازشناسی می‌گردد. تا قبل از این مرحله، هیچگونه اطلاعات معناشناختی در طول مراحل پردازشی مورد استفاده قرار نگرفته بود. مرحله پیش‌پردازش سعی در تمیز کردن تصویر سند به نحو مقتضی دارد و به علت عدم دسترسی به اطلاعات مفهومی، ممکن است باعث حذف برخی از اطلاعات مهم تصویر گردد. فقدان اطلاعات معنایی در مرحله قطعه‌بندی می‌تواند به نتایج حادتر و غیرقابل برگشتی بینجامد؛ چرا که خروجی این مرحله، تعیین‌کنندهٔ مرز الگوهای ورودی می‌باشد. بنابراین واضح است که در صورت فراهم‌شدن اطلاعات معناشناختی، دقت نتایج بازشناسی به نحو چشمگیری افزایش می‌یابد. مروری بر تحقیقات اخیراً انجام‌شده در زمینه

بازشناسی حروف نشان می‌دهد که در صورت استفاده از اطلاعات شکلی بدون به‌کارگیری اطلاعات معناشناختی، افزایش دقت قابل توجهی نخواهیم داشت. در نتیجه، یکپارچه‌کردن اطلاعات معنایی و شکلی در کلیه بلوک‌های سیستم‌های OCR به منظور بهبود نرخ بازشناسی صحیح، ضرورت دارد.

ساده‌ترین راه برای این منظور، استفاده از یک فرهنگ لغات برای اصلاح خطاهای جزئی است. این کار توسط یک برنامه واژه‌پرداز (با قابلیت خطایاب املایی) که املای کلمات را کنترل و چندین پیشنهاد برای اصلاح املای کلمات نامأنوس ارائه می‌کند، انجام می‌گیرد.

نتیجه‌گیری

با توجه به افزایش نیاز سیستمهای اطلاعاتی و اهمیت E-Document در عرصهٔ تعاملات دیجیتال استفاده از الگوریتمهای پردازش تصاویر بخصوص OCR اهمیت فوق‌العاده‌ای پیدا کرده‌است.
البته با توجه به موانع و پیچیدگیهای موجود در پیاده‌سازی این سیستمها ، همانطور که در مقالهٔ حاضر مورد بحث و بررسی قرار گرفت یافتن الگوریتمهایی که با استفاده از هوش مصنوعی ماشین بتواند این هدف را تا حد قابل قبولی برآورده نماید بسیار ضروری می‌نماید.

سیدرضا اشرفی‌دوست سلماسی
1384

مراجع:

- **SLANT ESTIMATION ALGORITHM FOR OCR SYSTEMS**
  E.Kavallieratou, N.Fakotakis, and G.Kokkinakis
  Wire Communications Laboratory,
  University of Patras, 26500 Patras, Greece

- **An Interactive Spelling Correction System for OCR Errors in Text**
  Kazem Taghva, Eric Stofsky
  Information Science Research Institute
  University of Nevada, Las Vegas
  taghva@isri.unlv.edu

- پردازش تصویر به روش آنالیز مؤلفه‌های اصلی - غلامرضا رمضان - مجموعه مقالات پنجمین کنفرانس بین‌المللی سالانه کامپیوتر ایران (1378)
- شناسایی چهرهٔ انسان با استفاده از ضرایب خود همبستگی - رضا صفا بخش - مجموعه مقالات پنجمین کنفرانس بین‌المللی سالانه کامپیوتر ایران (1378)
- استخراج الگوهای غالب در سناسایی تصویر - محمد انشانیه - پایان نامهٔ کارشناسی ارشد دانشگاه علم‌وصنعت
- پردازش تصویر / پایان نامه امیر شاه‌حسینی دانشگاه صنعتی مالک اشتر 1382
- www.mathnetworks.com