علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۱۸، شماره ۱، ۱۳۹۹
صفحات ۸۲- ۶۶
مقاله عادی

مدل موقعیت‌یابی متون در داده‌های عظیم تصویری

امین‌اله مه‌آبادی[1*]

*نویسنده مسئول، دریافت: ۹۹/۰۹/۱۳، بازنگری: ۹۹/۰۹/۱۷، پذیرش: ۹۹/۰۹/۲۱

[1]استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی‌مهندسی، دانشگاه شاهد، تهران، ایران

چکیده

روش‌های متداول موقعیت‌یابی متن در تصاویر طبیعی دارای چالش فضای زیاد جستجو جهت تشخیص دقیق، صحیح، سریع و کارا است و برای پشتیبانی از داده‌های عظیم تصویری به یک مسئله سخت لاینحل تبدیل می‌شود. می‌توان کاهش محاسبات پیچیده را با هدف کنترل فضای جستجو، منابع مصرفی و هزینه‌ها، با برخورداری از تکنیک‌های نرم‌افزاری در فضای سخت‌افزاری چندهسته‌ای، ساختارهای گرید و رایانش ابری شکلداد. موقعیت‌یابی و خواندن نوری متون در سطوح مختلف پاراگراف، خط متن، کلمه و کاراکتر در داده‌های عظیم به‌طور وحشتناکی به پیچیدگی آن می‌افزاید. در این مقاله روش‌های عمومی غیریادگیر موقعیت‌یابی متون در تصاویر کم‌حجم و روش‌های متداول داده‌های حجیم طبیعی بررسی شده و مدلی مناسب دارای تحلیل‌گر استراتژیک تصویر با استفاده از عامل‌های هوشمند و بات‌های زیرک برای پردازش داده‌های عظیم تصویری جهت موقعیت‌یابی متون توسط روبات‌ها ارائه و در کنار مجموعه داده‌های متفاوت و معیارهای ارزیابی مختلف تشریح شده است.

کلمات کلیدی: موقعیت‌یابی متون، داده‌های عظیم تصویری، پردازش تصویر، تصاویر ویدئویی، تصاویر طبیعی.

۱- مقدمه

امروزه دسترسی همگان به دوربین‌های دیجیتال ارزان‌قیمت در قالب سیستم‌های نهفته بی‌درنگ[1] تا سیستم‌های حرفه‌ای پزشکی، سبب تولید حجم عظیمی از تصاویر و چالش‌های مهم پردازش در رایانش ابری[2] شده است. دستیابی به اطلاعات سودمند از این حجم عظیم داده[3]، نیازمند پردازش‌های کارا و سریع برای تصمیم‌سازی دقیق است. دنیای آینده با ساختارهای ربات‌های فیزیکی نیازمند توجه به بات‌های سایبری[4] با معماری نوین و پردازشگرهای مناسب در قالب‌های سخت‌افزاری و نرم‌افزاری دارد. داده‌های عظیم تصویری[5] دارای ابعاد مختلف حجم زیاد، تنوع داده، ارزش مقداری است که پیچیدگی بیشتر آن از نظر وجود نویز، ابعاد زیاد تصویر، رزولوشن نامتناسب، زمینه پیچیده پردازش، نیازمندی توجه خاص به طراحی سیستم‌های پردازش ابری خاص‌منظوره مانند پزشکی و عمومی تجاری است. تصمیم‌سازی سریع و دقیق در آینده باید از هوشمندی بالای عامل‌ها[6]، قدرت زیاد پردازشگرها، معماری‌های مناسب و کارا، سرعت دستیابی و جمع‌آوری داده‌ها، درک مفید و سریع اطلاعات محلی[7]، مبادله هوشمند و بوقت اطلاعات سراسری[8]، جستجوی مداوم در یادگیری سریع و عمیق، عامل‌گرایی در قالب شبکه بات‌های سایبری[9] وسوسه‌گر، و زیرکی و هوشمندی در یادگیری برخوردار باشد. طراحی سیستم‌های فیزیکی و سایبری نوین برای داده‌های عظیم تصویری یک استلزام منطقی است که گرفتار چالش مهم دنیای پسا ابری است. حتی ضرورت توجه به سیستم‌عامل‌های خاص منظوره پردازش تصاویر ماشین‌های سایبری، تحلیل‌گرهای هوشمند[10] نوین، بات‌های سایبری جستجوگر، ساختارهای خاص ذخیره‌سازی داده‌های ورودی و متاداده‌های تولیدی موردنیاز به‌عنوان چالش‌هایی برای تحقیقات آینده وجود دارد.

ردیابی تغییرات در تصاویر و ردیابی تصویری تغییرات [۱] از چالش‌های اصلی دنیای نوین در سیستم‌های پویا برای دریافت اطلاعات، موقعیت‌یابی تغییرات، تشخیص تغییرات، اخذ پروفایل تغییرات، تصمیم‌سازی محلی، بررسی اثرات سراسری تصمیم‌ها، برنامه‌ریزی استراتژیک[11]، اصلاح تصمیم، اصلاح اشتباهات و ترمیم خرابی مصنوعی ایجادی، ردیابی سلسله تصمیمات، یادگیری هوشمند تغییرات، ردیابی تغییرات، استخراج دانش، انتخاب استراتژی مناسب ادامه عملیات، زمان‌بندی و تخصیص مناسب منابع، ترمیم سیستم، ذخیره و بازیابی موقت و دائمی اطلاعات، بهبود محیط و ساختار و مانند آن است. طراحی عامل‌گرای سلسله مراتبی[12] از لایه پایین عملیات بی‌درنگ برخوردار از اطلاعات محلی برای تصمیم‌سازی عملیاتی (لایه عملیاتی)، تا سطح میانی عملیات برخط برخوردار از اطلاعات سراسری برای اصلاح

تصمیم‌های عملیاتی (لایه مشورت)، نیازمند داشتن لایه سطح بالای برنامه‌ریزی استراتژیک (لایه برنامه‌ریزی) جهت انتخاب استراتژی مناسب برای بازی موفق عامل-های هوشمند هر سطح (شکل۱) در کارهای حرفه‌ای مانند جراحی‌های خاص پزشکی جهت پردازش داده‌های عظیم تصویری از سطوح پایینِ به‌کارگیری سیستم‌ها هوشمند نهفته بی‌درنگ با تصویربرداری غیرواقعی و مخدوش، سطوح میانی اخذ و ردیابی اطلاعات حیاتی همراه با خطا، تا سطوح بالای تصمیم‌سازی استراتژیک جراحی بر اساس اطلاعات مخدوش، نیازمند سیستم‌های با قابلیت برنامه‌ریزی استراتژیک، دارای مدل‌های خاص تصمیم‌سازی با تولید اطلاعات هر سطح و درج ریسک تصمیم‌گیری و ریسک عملیات در فضای عدم اطمینان است. در بسیاری از سیستم‌های تصمیم‌سازی آینده برای عملیات حساس مستلزم بهره‌گیری از شبیه-سازی هوشمند در هر سطح از عملیات است. پیچیدگی دنیای پسا ابری ضرورت ارائه چالش‌های اساسی برای تحقیقات آینده توسط محققان برجسته از سطوح بالای انتخاب استراتژی فعالیت سیستمی تا سطوح پایین تصمیمات استراتژیک عملیات بی‌درنگ را می‌طلبد. در این مقاله با تمرکز بر ساختارهای گذشته، ایده دنیای نوین در موضوع موقعیت‌یابی متون[۱۳] تصویری و روش‌های سنتی ارائه شده بررسی، مشخصات سیستم‌های نوین پردازشگر تصویر همراه با تعاریف داده‌های عظیم تصویری بیان و معیارهای ارزیابی روش‌ها جهت توجه به چالش آینده پردازش داده‌های عظیم تصاویر شرح داده می‌شود.

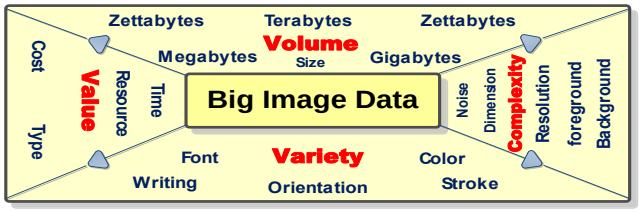
۱-۱ موقعیت‌یابی متون تصویری

موقعیت‌یابی یعنی تشخیص متون در تصاویر و یافتن موقعیت محل آن‌ها در ابعاد پاراگراف[۱۴]، جمله[۱۵]، کلمه[۱۶] یا حرف[۱۷] است (شکل۲). یافتن بی‌درنگ محل متن به‌عنوان هسته اصلی پردازش در تبدیل تصاویر به نوشتار در سه‌گام، موقعیت-یابی، تشخیص و خواندن متون است [۸۵] . تشخیص این مسئله در زبان‌هایی مانند فارسی و عربی با کلمات متصل که حروف دارای کاراکترهای منفک نیستند و از شکل‌های مختلف حروف متصل برخوردارند بسیار مشکلاست. بررسی‌ها بیانگر آن است که کارایی موقعیت‌یابی متن برای خواندن و تبدیل تصویر به نوشتار در دنیای کلان داده‌ها، وابستگی بسیاری به تنوع‌متون[۱۸]، حجم داده‌ها[۱۹] و میزان پیچیدگی[۲۰] متن و زمینه تصاویر دارد. همچنین عوامل دیگر جهت افزایش دقت، صحت و سرعت یافتن متن در مناظر طبیعی با ابعاد بزرگ و متون متنوع زیاد در داده‌های عظیم تصویری به پیچیده‌تر کردن این فرآیند کمک می‌کند. مشکلات مهم دیگر از قبیل محل دوربین[۲۱] ، درجه دانه‌درشتی متن، نوع متن، رزولوشن[۲۲] تصویر، روش نورپردازی، زاویه تصویربرداری، شرایط آب‌وهوایی، درجه کنترل محیط، نویزهای طبیعی[۲۳] و مصنوعی[۲۴] در لاینحل کردن مسئله بسیار مؤثر است [۱۵، ۴۱].

هر سیستم استخراج اطلاعات متن از تصویر، یک ورودی به شکلتصویر ساکن یا فریم ویدئو را دریافت می‌کند [۳]. تصاویر آن می‌تواند دارای سطوح خاکستری یا رنگی با فونت‌های متنوع و نگارش‌های گوناگون با پهنای قلم و ضخامت‌های متفاوت و فانتزی باشد (شکل۲). موقعیت‌یابی متون از سطح بالای یافتن توده‌های حجیم متنی و پاراگراف‌ها، سطح میانی ادراک جملات، تا سطح تشخیص حروف بر اساس ساختار هر زبان بر پیچیدگی آن می‌افزاید و تشخیص دقیق محل متن علاوه بر استفاده از روش‌های موازی، الگوریتم‌های توزیعی و معیارهای اندازه‌گیری آن‌ها نیازمند برخورداری از بستر پردازشگرهای سریع، کارا و توزیع‌شده هستند. افزایش حجم داده‌ها در کنار درجه پیچیدگی محاسبات جهت عملیات اصلاح اشکالات تصویربرداری، پیش‌پردازش و بهبود تصاویر، ارتقاء کیفیت و واضح‌سازی متون، تفکیک متون واقعی از متون تصادفی غیرواقعی، حذف نویزهای تصویر، تفکیک متون هم‌پوشان، تعیین ابعاد قلم نگارش، اصلاح زاویه چرخش متون و میزان تاری تصویر، بهبود تحدب و تقعر زمینه تصاویر و مانند آن بر لاینحل بودن آن می‌افزاید. این امر نیازمندی به سیستم‌های نوین موقعیت‌یابی تصاویر طبیعی را علاوه بر برخورداری از روش‌های نرم‌افزاری توزیعی و موازی کارا به داشتن ساختارهای قوی پردازشی، معماری‌های موازی سریع و الگوریتم‌های توزیعی مربوطه مخصوصاً برای پردازش تصاویر داده‌های عظیم و کلان داده‌ها سوق می‌دهد [۱].

شکل۱: سیمای عامل‌گرای سلسله مراتبی برای پردازش داده‌های عظیم تصویری.

شکل۲: تشخیص متون متنوع در تصاویر گوناگون طبیعی و مصنوعی.



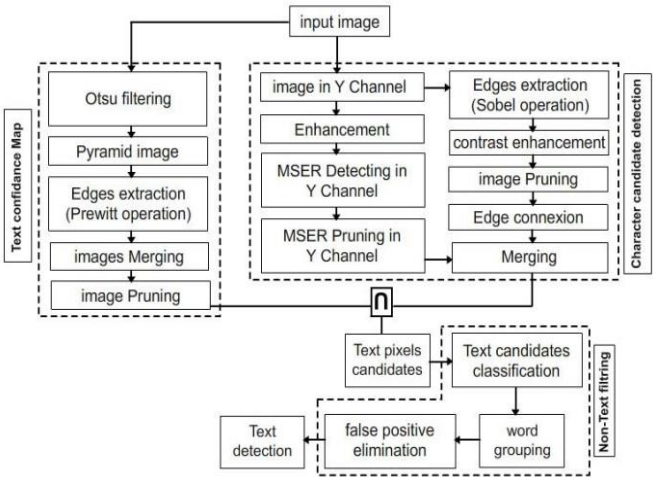
شکل۳: مشخصات و ساختار داده‌های عظیم تصویری.

متون موجود در تصاویر شامل دو نوع متون حقیقی و مصنوعی ساختاریافته منظم و نامنظم در مناظر طبیعی و متون فانتزی گرافیکی هستند. متون مناظر طبیعی از تصاویری هستند که توسط دوربین‌های عکس‌برداری متداول اخذ شده‌اند (شکل۲). موقعیت‌یابی متون برای کاربردهای خودکار نوین مانند شناسایی علائم، خواندن متون قدیمی، بایگانی اسناد، خبر یابی تا تشخیص تابلوها برای علائم جاده‌ها، متون آگهی‌ها، متون خودروهای اورژانس، متون تابلو مغازه‌ها، نوشته‌های پیراهن افراد و متون زیرنویس فیلم‌ها توسط ربات‌های ساکن و پرنده توسعه بسیاری یافته است. متون مناظر و چشم‌اندازهای طبیعی در مقایسه با متون گرافیکی ساختاریافته، متنوع و غیرقابل‌پیش‌بینی هستند [۱،۵۱]. استخراج متون مناظر طبیعی توسط ریزپرنده‌ها یا ربات‌های سیار برای تشخیص مسیر بر اساس تابلوهای راهنمایی، کشف تشخیص پلاک خودرو، شناسایی اشیاء و مانند آن از کاربردهای نوین تشخیص خودکار متون هستند [۲].

۱-۲ داده‌های عظیم تصویری

گرچه برای کلان داده‌ها از چالش حجم به‌عنوان بعد اصلی استفاده می‌شود و هنوز تعریف مناسبی برای آن وجود ندارد ولی اکنون در ابعاد تنوع، حجم و شتاب تولید[۲۵] از سرعت رشد بسیاری خطرناکی برخوردار است. تاکنون هیچ تعریف اساسی برای کلان‌داده از نوع تصویری یا *داده‌های عظیم تصویری* ارائه نشده است [۶۵].

حتی برای داده‌های عظیم تولیدی اینترنت اشیاء که از سال ۲۰۱۵ به فرم تصویر، آدیو و ویدئو مطرح‌شده فقط به مسئله حجم داده در مبادلات پرداخته شده و ابعاد دیگری برای آن متصور نشده‌اند.



شکل۴: پردازش موقعیت‌یابی در روش مبتنی بر مؤلفه‌های همبند [۸۷].

ما برای داده‌های عظیم تصویری خصوصاً در ساختار پیچیده موقعیت‌یابی متون، یک ساختار چهار بعدی با ابعاد تنوع، حجم، ارزش۲۶ و پیچیدگی متون در تصویر تعریف می‌کنیم (شکل۳). تنوع به ناهمگونی متون درون تصویر از نظر ابعاد فونت، پهنای قلم، تنوع رنگ و نوع نگارش توجه دارد. حجم به میزان داده‌های ساختاریافته یا نیافته متون درون تصویر اشاره دارد که تابع توسعه تکنولوژی تصویربرداری، خطوط انتقال، روش فشرده‌سازی، قدرت ذخیره، نوع بازیابی تا رسانه ذخیره‌سازی است. ارزش بر میزان ارزش‌گذاری هزینه پردازش از نظر نوع، میزان و اهمیت منابع تخصیصی مانند ارزش توان مصرفی و زمان استخراج تمرکز می‌کند [۶۴]. پیچیدگی به میزان شلوغی یا شفافیت تصویر اشاره دارد که به رزولوشن، ابعاد و درجه پیچیدگی تصویر وابسته است لذا در سرعت و کارایی پردازش آن تصویر مؤثر است. ابعاد بزرگ تصاویر و رزولوشن زیاد آن نیز باعث افزایش حجم پردازش می‌شود. برای فهم بیشتر در شکل۳، مدل جامع ارتباط بین ابعاد داده‌های تصویری عظیم، نشان داده شده که مبین ارتباطات درونی و رشد متوازن در ابعاد مختلف می‌باشد و نیازمند سیستم پردازشی بسیار پیچیده و گران‌قیمت است. چهار لبه مدل فوق مبین چهار بعد داده های عظیم تصویری یعنی تنوع، حجم، ارزش و پیچیدگی است. درون این مربع چهار بعد داده‌های بزرگ به‌صورت مثبت تحت تأثیر افزایش دقت، تنوع، حجم، ارزش، پیچیدگی، رنگ، رزولوشن، ارزش، زاویه و فضای تصویربرداری است.

مدل فوق نشان می‌دهد که داده‌های تصویری سنتی با بعد افزایش حجم فقط زیرمجموعه‌ای از تعریف جدید داده‌های عظیم تصویری در چهار بعد هستند. رشد تصویر در آن چهار بعد متناسب با افزایش داده‌های ساختاری و بدون ساختار گسترش می‌یابند. جهت فلش در شکل۳ نشان‌دهنده جهت رشد هر کدام از ابعاد است و پردازش آن بسیار مهم می‌باشد که مشخصات سیستم موردنیاز پیشنهادی آن نیز در بخش ۳ شرح داده شده است. سیستم پیشنهادی به یک هسته مرکزی "تحلیل‌گر استراتژیک تصویری هوشمند" نیاز دارد که مشخصات داده عظیم تصویری مورد بررسی را از نظر چهار بعد پیشنهادی شکل۳، به‌صورت متاداده ارائه دهد. این امر به استراتژی انتخاب منابع برای پردازش این تصاویر در کارهای تخصصی و حرفه‌ای مانند پزشکی کمک بسیار خواهد کرد. مثلاً در بعد پیچیدگی، در تصاویر داده‌های مختلف نویزی با زمینه پیچیده و در بعد تنوع، در تصاویر با داده‌های متنوع و متفاوت نیازمند پردازشگر قوی و ساختار نوین هوشمند عامل‌گرا برای پردازش‌های اطلاعات محلی و استفاده از اطلاعات سراسری موقعیت‌یابی است. همچنین در بعد حجم، در تصاویر با حجم داده زیاد نیازمند حافظه زیاد و سریع و در بعد ارزش، برای پردازش‌های سریع و کم‌هزینه باید از منابع قدرتمند استفاده شود؛ لذا این تحلیل‌گر استراتژیک تصویری با طراحی عامل‌گرا یا بات‌گرا و بهره‌برداری از سرورهای هوشمند می‌تواند در دسته‌بندی نوع و ابعاد داده‌های عظیم تصویری کمک بسیاری به سیستم مکان‌یابی کند.

۱-۳ سیستم موقعیت‌یابی متن

استخراج و فهم اطلاعات متون بسته به نوع کاربرد دارای مراحل (۱) آشکارسازی۲۷، (۲) موقعیت‌یابی۲۸، (۳) خواندن محتوا۲۹ و (۴) ردیابی ۳۰ است. ساختار سنتی سیستم خودکار ماشینی استخراج متن شامل تشخیص متن، موقعیت-یابی، شاخص کردن متن (در سطح بلاک، خط متن، کلمه و حرف)، خواندن آن‌ها و ردیابی در تصاویر متوالی مانند نمایش شکل۴ است. در این مقاله به بررسی موقعیت-یابی متون با توجه به افزایش دقت تفکیک‌پذیری توجه می‌شود. به دلیل رابطه تنگاتنگ آشکارسازی و موقعیت‌یابی متن، محققان آن را آشکارسازی متن و در برخی موارد موقعیت‌یابی متن معرفی کرده‌اند [۱۲]. در مرحله آشکارسازی و موقعیت‌یابی متن، تصویر برای یافتن احتمال وجود متن کاوش می‌گردد و موقعیت مکانی متون برای شناسایی و عمل قاب‌گذاری محدوده ۳۱ متن‌ها انجام می‌شود. خروجی حاصل از تعیین قاب بر محدوده متن، ابعاد آن را جهت حذف نواحی غیر متن مشخص می‌کند. درنهایت کاراکترهای متن از زمینه جدا و محتوای متن به روش تشخیص نوری کاراکتر قابل‌خواندن و فهم می‌شود. ازآنجاکه معمولاً نواحی متن در مناظر طبیعی در معرض نویز قرار می‌گیرند و برای پردازش از درجه تفکیک‌پذیری پایینی برخوردارند باید جهت کاهش خطا، تصویر قاب متن استخراجی در سطح تشخیص بلاک، خط متن، کلمه و کاراکتر، از ارتقاء کیفیت برخوردار شود [۵۲]. برای موقعیت-یابی متن باید متون به طور کامل از زمینه که ممکن است نویزی باشد جدا و متون اصلی شاخص گردد [۵۵-۵۷]. سپس قاب استخراجی به شیء تصویری تبدیل شود و به‌منظور افزایش دقت تشخیص، بلاک حاوی آن به سطح کیفیت تصویری بهتری ارتقاء یابد. نمونه سیستم سنتی تشخیص و گام‌های پردازشی تصویری مبتنی بر مؤلفه‌های همبند در شکل۴ ارائه شده است که مراحل مختلف و سنگین موقعیت‌یابی را نشان می‌دهد.

۱-۴ چالش‌های موقعیت‌یابی

گرچه حجم زیاد پردازش و درجه پیچیدگی محاسبات برای آماده‌سازی متون جهت تشخیص و موقعیت‌یابی بر تشخیص دقیق و صحیح آن مؤثر است ولی مقدمات لازم برای حل مشکلات ابتدایی آن تصویربرداری نادرست، درجه کاربرد آن را برای مصارف بی‌درنگ و برخط غیر بی‌درنگ مشکل می‌سازد. حتی روش‌هایی که در گذشته دارای کارایی بسیاری بودند در کاربردهای داده‌های تصویری عظیم ناکارا و بدون استفاده هستند. ازاین‌رو تغییر استراتژی در راستای داده‌های عظیم تصویری نیازمند رفع مشکلات سیستم‌های موجود و روش‌های نوین است. نیاز به روش‌های نوین برای رفع چالش‌های اساسی از قبیل تشخیص الگوها جهت سرعت تشخیص، مدل‌سازی سلسله‌مراتبی برای تفکیک پردازش‌ها، ساختار لوله‌ای برای موازی‌سازی مجازی عملیات، نگرش توزیعی برای توازی واقعی، مدل‌سازی عامل‌گرای[۳۲] هوشمند برای مقیاس‌پذیری[۳۳]، و توزیع هوشمندانه محاسبات جهت کاهش فضای جستجو[۳۴] در تصمیم‌سازی با اطلاعات محلی و سراسری وجود دارد. در این نگرش، الگوهای موجود موقعیت‌یابی متن دارای چالش‌های بسیاری از قبیل نوع مدل‌سازی[۳۵]، معماری روش تشخیص[۳۶]، شیوه تصویربرداری[۳۷]، درجه پیچیدگی محاسبات[۳۸]، دقت تشخیص[۳۹]، میزان وابستگی به زبان متن[۴۰]، تشخیص متون تصاویر طبیعی[۴۱]، پردازش داده‌های عظیم تصویری، و بهینه‌سازی منابع مصرفی[۴۲] مخصوصاً در دنیای نانوتکنولوژی[۴۳] (در غلبه جریان استاتیک[۴۴] بر جریان دینامیک[۴۵] مدارها) هستند که در زیر بیان می‌شود.

(۱) مدل‌سازی و تصمیم‌سازی: از چالش‌های عمده برای کاهش حجم محاسبات در تشخیص می‌توان به نوع مدل‌سازی مسئله تشخیص از نظر مرکزیت-گرایی[۴۶]، غیر مرکزیت‌گرایی[۴۷]، و توزیعی[۴۸] یا عامل‌گرا توجه کرد. جایی که وجود مدل مرکزیت‌گرا، درجه دقت را افزایش می‌دهد ولی محاسبات را بسیار پیچیده می‌کند؛ مدل غیر مرکزیت گرا باعث کاهش درجه محاسبات و محلی سازی آن است ولی دقت را نیز کاهش می‌دهد؛ در مدل توزیعی عامل‌ها مبادلات داده‌ها را کمی افزایش می‌دهند زیرا نیازمند مشاوره[۴۹] با هم هستند ولی درجه تفکیک‌پذیری محاسبات را در کنار کاهش فضای جستجو افزایش داده و همچنین با مشاوره عامل‌ها میزان دقت تشخیص را افزایش می‌دهند؛ مدل‌های غیر مرکزیت‌گرا و عامل‌گرا به‌طور قابل ملاحظه‌ای درجه مقیاس‌پذیری مدل را بالا می‌برند و سرعت تصمیم‌سازی را به نحو مؤثری بهبود می‌بخشند [۶۰]. تمرکز آن‌ها بر درجه تفکیک‌پذیری نواحی متون و تفکیک هم‌پوشانی نواحی متفاوت متنی است. در بسیاری از موارد مانند تشخیص بیماری‌ها برای اخذ سریع دانش و تصمیم‌سازی جهت جستجوی محلی اطلاعات، تکنولوژی شبکه بات‌ها به‌جای استفاده از عامل‌ها لازم است و نیازمند ارائه روش‌های بات‌گرا جهت جستجوی بی‌درنگ هستیم.

(۲) معماری روش تشخیص: از چالش‌های عمده معماری روش تشخیص می‌توان به درجه کنترل ابعاد و پیچیدگی تصویر از نظر ساختار سلسله‌مراتبی، درجه توازی زیربخش‌های تشخیص، و درجه تفکیک الگوریتم‌ها برای پشتیبانی از تنوع داده‌ها در ساختارهای بی‌درنگ اشاره کرد [۶۳]. مثال آن متون خبری در حال گسترش صفحات وب است که توازی تشخیص و سرعت پردازش در پشتیبانی از تنوع، حجم و تغییرات زیاد داده‌های آن لازم است [۳۶، ۶۱ ، ۶۳].

(۳) شیوه تصویربرداری: از نظر مشخصات تصویر ورودی به شرایط کنترل نشده تصویربرداری مانند کیفیت پایین تصویر، نورپردازی ناهموار، سطوح غیرمسطح و عمق‌دار، فاصله دوربین، وجود اشیاء با بافت‌های مشابه متن، زمینه‌های پیچیده با ابعاد مختلف و رنگ متغیر فونت متون در تصاویر طبیعی می‌توان اشاره کرد [۱۷]. در اکثر روش‌های موجود با هدف چابک‌سازی[۵۰] روش فقط به حل چالش‌های این بعد اشاره می‌شود.

(۴) پیچیدگی محاسبات: گرچه به نظر نمی‌آید که انجام محاسباتِ یافتن متن بسیار پیچیده باشد ولی سرعت روش در کاربردهای بی‌درنگ برخط و توجه به آن برای کاهش محاسبات یا درجه پیچیدگی هر روش جهت پشتیبانی از داده‌های کلان لازم است. در این بعد، روش‌های چابک تشخیص، کاهش فضای جستجو، جستجوی موازی یا سلسله مراتبی در فضای تصویر جهت افزایش سرعت و کاهش پیچیدگی مدنظر است [۵۹] .

(۵) سرعت در دقت تشخیص: در بعضی کاربردها مانند تصاویر کتب چاپی از نظر دقت و صحت تشخیص، روش‌ها از معیارهای اندازه‌گیری خاص خود بهره می‌برند که با تغییر تنوع متون و چندمقیاسی[۵۱]، از دقت تشخیص آن‌ها کاسته می‌شود و بعضاً آن روش‌ها کاربرد عمومی ندارد [۵۳]. بیشتر روش‌های موجود به دلیل تکیه بر ویژگی‌های خاص تصویر، قادر به تشخیص متون خاص (مانند کتاب یا صفحات تایپی وب) هستند و ممکن است متون دیگر را تشخیص ندهند. لذا لازم است به افزایش دقت در کنار سرعت تشخیص تصاویر طبیعی متنوع توجه شود.

(۶) وابستگی‌های زبانی: در تشخیص متن بر اساس کاربردهای خاص (غیرعمومی) مسئله وابستگی به زبان(های) خاص مطرح می‌شود که از جامعیت آن روش می‌کاهد. در بعضی کشورها مانند کره، چین و ژاپن برای راهنمایی خارجی‌ها، استفاده از چندین زبان در حمل‌ونقل، تابلوهای ترافیکی و مانند آن مرسوم است. بیشتر روش‌های تشخیص چند زبانی برای کاربرد زبان(های) خاص طراحی می‌شوند و قادر به تشخیص همزمان زبان‌های دیگر نیستند [۳۶-۳۷]. تشخیص متون چند زبانی برای یک سیستم قابل‌اعتماد[۵۲] لازم است و با افزایش تعداد زبان، امکان کاربرد آن بسیار زیاد می‌شود. برای فرار از این تله، لازم است روش‌های آینده به‌سوی مفهوم تشخیص شیء[۵۳] سوق می‌یابند تا متون مختلف در قالب اشیاء متفاوت با ویژگی‌های خاص خود تعریف و همزمان بازیابی شوند [۶۲] .

(۷) تنوع تصاویر طبیعی: از چالش تشخیص متن در تصاویر داده‌های عظیم طبیعی به دلیل شرایط کنترل نشده، وجود زمینه پیچیده و تغییرات شدید الگوهای متن (مانند نوع فونت[۵۴]، زبان[۵۵]، رنگ[۵۶]، مقیاس[۵۷] و جهت[۵۸] آن) است که در تشخیص بی‌درنگ متون طبیعی و چند زبانی آن را بدون راه‌حل رها ساخته است [۸۵] .در بسیاری از کاربردها مانند ربات‌های خدمات‌رسان[۵۹] پردازش تصاویر طبیعی امری اجتناب‌ناپذیر و لازم است لذا شرایط کنترل‌شده که در تشخیص یک ماشین ربات وجود ندارد بر مشکلات قبلی می‌افزاید.

(۸) داده‌های عظیم تصویری: با افزایش سریع کاربردهای تصویر، چالش توسعه *داده‌های عظیم تصویری* از نظر پردازش در تشخیص سطوح مختلف توده متن، پاراگراف، خط متن، کلمه و کاراکترها به دلیل حجم زیاد و تنوع متون داده‌های عظیم تصویری از مشکلات دنیای آینده است [۶۵]. وجود روش‌های تشخیص بسیار سریع و مقیاس‌پذیر در آینده نه‌چندان دور از مشکلات بشریت خواهد بود لذا ارائه روش‌های دارای فضای جستجوی کم، قدرت پردازش زیاد و قابلیت مقیاس‌پذیری در چند مقیاس امری حیاتی خواهد بود [۶۳-۶۱] . این روش‌ها باید با برخورداری از قالب کار مناسب و ساختار توزیعی یا موازی واقعی و ساختار سلسله مراتبی لوله‌ای یا شبه موازی بتواند بر روی یک ماشین دارای چندین هسته‌ای برای سیستم‌های نهفته بی‌درنگ و بر روی ماشین‌های مختلف دارای پردازشگر جی‌پی‌یو[۶۰] برای سیستم‌های آنلاین با حجم داده بسیار عظیم اجرا شود.

(۹) هوشمندی و تطابق: در پردازش داده‌های عظیم سیستم نیازمند انتخاب هوشمند استراتژی‌های متفاوت در سطوح مختلف عملیاتی برای تطابق با فضای کاری، وضعیت منابع، قابلیت توزیع و ادراک شرایط موردنیاز است. با افزایش تنوع و حجم داده‌ها، سیستم‌های آینده نیازمند پردازش سریع با استفاده از اطلاعات محلی و تطابق با شرایط با استفاده از اطلاعات سراسری در هدایت سریع عامل‌ها یا بات‌ها هستند. لذا یادگیری سریع و هوشمندی فوق‌العاده جهت تطبیق‌پذیری با محیط جهت بهینه‌سازی فضای تصمیمات و تعامل با محیط در ادراکات سریع دانش[۶۱] و تولید و بهره‌گیری از متاداده‌ها[۶۲] امری ضروری است.

(۱۰) بهینه‌سازی تخصیص منابع: حجم زیاد محاسبات پیچیده در ساختارهای موجود با منابع محدود و لزوم پردازش داده‌های کلان آینده توجه به بهینه‌سازی مصرف منابع (مانند توان مصرفی)، درجه اهمیت منابع مصرفی و بی‌درنگی زمان پاسخ را اجتناب‌ناپذیر می‌سازد [۶۴]. به دلیل غلبه جریان استاتیک بر جریان دینامیک تراشه‌ها در کاربردهای سریع نانوتکنولوژی آینده و قیود جدید مانند مصرف توان و دمای تولیدی، استفاده بهینه از منابع بسیار مهم است. به دلیل پدیده سیلیکون خاموش[۶۳] در ساختارهای چندهسته‌ای و نیازمندی به خاموش کردن بخش‌های عظیم پردازشگرها، توجه به مصرف توان و دمای تولیدی تراشه‌ها و بهینه‌سازی استفاده از منابع امری اجتناب‌ناپذیر است و قیود طراحی سیستم‌های خودکار ماشینی را بیشتر می‌کند [۶۱،۶۶]. روش‌های موجود به دلیل چالش‌های مذکور و محدودیت‌های مقیاس‌بندی در دنیای داده‌های عظیم از کارایی و کاربرد کمتری برخوردار خواهند بود [۶۵] . در فضای ابری استفاده اشتراکی از قدرت پردازش موبایل‌ها اجتناب‌ناپذیر و توزیع مصرف منابع بسیار گسترده خواهد بود.

(۱۱) تصمیم‌سازی عدم اطمینان: سیستم در عملیات بی‌درنگ در حالت عدم اطمینان[۶۴] قرار دارد و تصمیم‌سازی[۶۵] در این وضعیت بر اساس اطلاعات محلی قرار دارد و همراه با ریسک[۶۶] است. باید از وضعیت ریسک در عملیات بی‌درنگ مطلع بود تا بتوان از تصمیم‌گیری نامناسب پیش‌گیری کرد و در صورت نیاز به تصمیم‌سازی ریسک‌گریز[۶۷] و ریسک‌گرا[۶۸] روی آورد. در صورت اتصال و ارتباط عامل‌ها و اطلاع آن‌ها از اطلاعات سراسری باید تصمیمات اصلاح یا تنظیمات دوباره برقرار شود تا ریسک عملیات سراسری رفع شود. این چالش باید به دقت مطالعه و مدلسازی شود

تا از مدل آن برای تشخیص و اصلاح عملیات بهره برد. باید به تناسب اهداف پردازش در هر کاربرد تصویری با تعریف تهدیدات[۶۹] نسبت به ارزیابی عملیات و عدم ورود به منطقه ریسک مطمئن بود.

در راستای پشتیبانی از داده‌های عظیم تصویری، روش‌شناسایی برای تشخیص متون تصاویرطبیعی که بتواند همزمان به داده‌های متنوع زیاد و زبان‌های مختلف در سطوح زیاد بلاک، خط متن، کلمه و حرف راهنمایی خاص در ابعاد ذکر شده ارائه دهد مشاهده نشد. توجه به چالش‌های تصاویر پیچیده حجیم و مقاومت در برابر مشکلات گوناگون آن در کنار بستر مناسب برای کاهش زمان‌محاسبات و زمان-جستجو لازمه دنیای آینده است. برای حرکت به‌سمت این سیستم‌ها در این مقاله با بررسی روش موجود و توجه به‌پشتیبانی از تنوع داده‌ها در ابعاد مختلف ضمن ارائه راه‌کارها، به‌مسیر تحقیقات آینده در پردازش‌تصویر اشاره خواهد شد. هدف این مقاله بررسی ابعاد مسئله در موارد زیر است.

- بررسی ساختار روش‌های عمومی دقیق و سریع تشخیص در تصاویر طبیعی در سطوح بلاک، خط متن، کلمه و حرف دارای قدرت پیش‌بینی و سرعت با فراخوانی بالا،
- کلاسه‌بندی مناسب روش‌ها جهت درک سریع و حذف نواحی غیرمتن برای افزایش دقت و تشخیص هم‌زمان متون چند مقیاسی در کنار چند زبانی،
- توجه به ارائه معیارهای و مجموعه داده‌های جدید جهت حذف بلاک‌های غیرمتن در جهت افزایش درجهٔ تفکیک‌پذیری کلمات، و
- قابلیت تشخیص هم‌زمان داده‌های متنوع، زبان‌های مختلف و متون با مقیاس متفاوت در تصاویر دوربین و ویدئو در راستای پشتیبانی از داده-های عظیم تصویری است.

در ادامه مقاله و در بخش۲ مهم‌ترین کارهای موجود علمی سنتی و کارهای نوین مرتبط با موقعیت‌یابی متون بررسی می‌شود. در بخش۳ ضمن تعریف مسئله موقعیت‌یابی در تصاویر رنگی طبیعی سنتی، مدل‌سازی و روش‌های نوین ارائه می-شود. در بخش ۴ به ایده موقعیت‌یابی بات‌گرا و جزئیات آن پرداخته می‌شود. در بخش ۵ به بررسی مجموعه داده‌ها و معیارهای ارزیابی توجه شده و نهایتاً در بخش۶ نتیجه‌گیری می‌شود.

۲- کارهای مرتبط

به دلیل فقدان روش‌های نوین در راستای پردازش داده‌های عظیم تصویری در ابتدا به روش‌های سنتی اشاره می‌شود که ضعف عمده آن‌ها عدم مقیاس‌پذیری و برخورداری از محاسبات پیچیده است. با وجود اهمیت داده‌های عظیم تصویری تاکنون مقاله‌های مهم و ویژه‌ای در این راستا ارائه نشده است. البته چندین روش نوین پردازش تصویر در راستای داده‌های عظیم تصویری معرفی شده‌اند که فقط به غلبه بر بعد حجم زیاد (در موضوع خاص تصاویر پزشکی) پرداخته‌اند. در تحقیقات گذشته اثری از مسئله پیچیده موقعیت‌یابی متن در چند بعد بیان‌شده نیست. روش-های عمده موجود عبارت از روش‌های بر اساس پنجره لغزنده، مولفه‌های هم‌بند، ترکیبی، بافت‌گرا، گوشه‌گرا، تولید شبه‌خودکار حقیقت‌پایه[۷۰]، پرسش تصاویر مورد تقاضا، و نواحی قوی کاراکتری هستند [۷۲]. روش‌های پنجره لغزنده دارای عیب عمده کندی به دلیل پردازش چندمقیاسی هستند که از محدودیت فضای جستجو برخوردارند ولی روش‌های مبتنی بر مؤلفه‌های همبند در برابر کنتراست کم متن و شدت زیاد روشنایی ضعف دارند. مؤلفه‌های هم‌بند بر اساس لبه‌یابی (ضعف تشخیص نواحی کوچک تصویر و پهنای کم قلم)، بر اساس رنگ (ضعف در درجه خوشه‌بندی و تعداد خوشه‌ها) و یا ترکیب این دو با نتایج بهتر کار می‌کنند. روش‌های ترکیبی با آشکارسازی مناطق کاندیدای متن و حذف غیرمتن (ضعف سگمنت‌بندی) صورت می‌گیرد. روش‌های بافت‌گرا (ضعف حدس احتمال‌بافت و مقیاس متن) بر اساس ساختار بافت، متن را تشخیص می‌دهند. روش‌های سریع گوشه‌گرا بر اساس نقاط زیاد گوشه‌ها (ضعف کارایی) عمل می‌کنند.

به دلیل توجه زیاد محققان به روش‌های مبتنی بر مؤلفه‌های همبند و اهمیت آن برای مکان‌یابی متن در سال‌های اخیر، اکثر کارهای موجود در این حوزه قرار دارند [۹-۲۰]. روش ارائه‌شده جین و همکاران [۱۰] تصاویر را توسط خوشه‌بندی رنگی، مؤلفه‌های گروه‌بندی‌شده در خط متن، تجزیه‌وتحلیل مؤلفه‌ها و حذف مؤلفه-های غیرمتنی بر اساس قوانین هندسی به چندین بخش بدون تداخل افراز می‌کند. این روش به دلیل تنظیم دستی قوانین و پارامترها برای تصاویر پیچیدهٔ طبیعی کند است و خوب عمل نمی‌کند. استفاده از ویژگی پهنای قلم[۷۱] نزدیک به هم کاراکترها توسط اپشتاین با ارائه عملگر جدیدی به نام تغییر پهنای قلم[۷۲] انجام شد [۱۱]. روشی آسان برای بازیابی پهنای قلم کاراکترها در نقشهٔلبه بیان کردند که قابلیت استخراج مؤلفه‌های متنی با مقیاس‌های متفاوت و جهت‌های مختلف از تصاویر پیچیدهٔ طبیعی را دارد. ولی این روش نیازمند تعریف قوانین و پارامترها توسط انسان است و تنها به متون افقی‌متن توجه دارد. در این راستا تحقیق دیگری برای تصاویر طبیعی با زمینه پیچیده و چند زبانی ارائه شد که با بهبود تشخیص لبه بتواند تشخیص بهتر بدهد نواحی غیرمتن را حذف کند که نتایج خوبی دارد ولی برای تصاویر بزرگ مناسب نیست [۱۰۱] .

نویمان و همکاران [۱۲]، الگوریتم تشخیص متن بر اساس نواحی حدی بیشینه-پایدار[۷۳] پیشنهاد کردند که با استفاده از یک طبقه‌بند یادگیر این نواحی را از تصویر اصلی استخراج و مؤلفه‌های نامعتبر آن حذف می‌کند. سپس مؤلفه‌های باقی‌مانده خط متن از طریق مجموعه قوانین، گروه‌بندی می‌شوند. به دلیل استفاده از قوانین خاص مرتبط با متون افقی یا نزدیک به افق، روش قادر به کشف متون با زوایای مختلف نیست. روش تغییر پهنای قلم و روش نواحی حدی بیشینه پایدار دو روش اصلی در زمینهٔ تشخیص متون مناظر مختلف هستند [۱۲-۱۱] که در بسیاری از آثار از جمله [۱۳-۱۷] استفاده‌شده و عامل موفقیت‌های بزرگی در تشخیص چهره [۱۸] و حذف نویز تصویر [۱۹] به شمار می‌آیند. ژائو و همکاران [۲۰]، یک فرهنگ لغت از نمونه‌های آموزشی ساختند و از آن به‌عنوان تصمیم‌گیرنده در مورد نواحی متنی استفاده کردند. تعمیم فرهنگ لغت آن محدودیت دارد و این روش قادر به حل مشکلات چرخش و تغییر مقیاس متن نیست.

شیواکومارا و همکاران [۲۱] ابتدا تصویر ورودی را به کمک لاپلاسین فیلتر و سپس کلاسه‌بند K-means را در تشخیص محل کاندیداهای متن بر مبنای بیشینهٔ اختلاف بکار بردند. برای جداسازی رشته‌های متنی از یکدیگر از اسکلت هر مؤلفهٔ همبند استفاده کردند. در نهایت میزان صافی‌طول[۷۴] و چگالی لبه رشتهٔ متنی را برای حذف نواحی اشتباه[۷۵] تشخیص‌داده‌شده اعمال کردند. این روش در سطح خط متن خوب عمل می‌کند ولی در سطح کلمه نتایج مطلوبی ندارد. همچنین شیواکومارا و همکاران [۸]، بردار شارگرادیان[۷۶] تصویر را استخراج کردند تا بتوانند پیکسل‌های شاخص متن در تصویر لبه را شناسایی کنند. سپس مؤلفه‌های لبه که در تصویر لبه متناظر با آن پیکسل‌های شاخص است را استخراج و نهایتاً با گروه‌بندی و حذف مؤلفه‌های غیرمتن، مؤلفه‌های نهایی را استخراج کردند. این روش نیز در سطح خط متن به‌خوبی عمل می‌کند ولی در سطح کلمه نتیجه خوبی ندارد. این روند با استفاده از یادگیری عمیق و داده‌های زیاد رشد کرد و ایده‌های جدیدی ارائه شد که باعث بهبود تشخیص آن گردید [۳۱-۳۸,۳۵-۴۰].

افزایش عملکرد این روش‌ها تا حد زیادی به دلیل آموزش و به‌کارگیری داده‌های بزرگ آموزشی است که در دسترس عموم نیست. ازآنجاکه کلید مؤثر آشکارسازی و تشخیص متن در ساختار تقسیم‌بندی کاراکترها و روابط بین آن‌ها نهفته است این روش را از یک روش عمومی خارج و به حوزه خاص محصور می‌کند. در آخرین روش‌های علمی تعداد کمی از تحقیقات به مباحث قبلی توجه کرده‌اند [۳۸-۴۰]. روش یانگ و همکاران [۳۹] مبتنی بر یادگیری در تصاویر طبیعی و مناسب تشخیص

متون افقی است که فقط در سطح خط متن عمل می‌کند. این روش دارای کارایی خوبی است و فقط مشکل کندی و یادگیری شبکه دارد. روش واسی‌پولوس [۴۰] از روش‌های مبتنی بر ناحیه است که تغییرات رنگ زیاد در متن را قبول نمی‌کند و کنتراست متن و زمینه باید به‌اندازه کافی زیاد باشد. همچنین نیاز به دانش از شکل کاراکترها دارد ولی نسبت به نوع فونت و ابعاد آن مقاوم و بیشتر مناسب متون کتاب و مجلات است.

روش‌های شبه خودکار حقیقت‌پایه با اجازه به اصلاح دستی ناحیه مشخص‌شده و ضعف استفاده از ویژگی‌های زیاد دارای کارایی کمی هستند. روش پرسش تصاویر مورد تقاضا با قدرت تشخیص از لایه پایین (کاراکتر) به بالا (کلمه، خط متن، پاراگراف و توده متنی) با ضعف یادگیری صورت می‌گیرد. روش‌های مبتنی بر نواحی قوی کاراکتری از چند منظر بر اساس احتمال بالای محلی در تشخیص کاراکتر بودن عمل می‌کنند. آن‌ها در برابر تصاویر با کیفیت پایین و کنتراست پایین پس‌زمینه و پیش‌زمینه ضعیف هستند. برای مقایسه سریع روش‌های مهم سنتی، خلاصه نتایج بررسی نقاط قوت و ضعف مهم‌ترین تحقیقات روش‌های عمومی در جدول۱ آمده است.

اصل روش‌های داده‌های عظیم تصویری با پیروی از ساختار سلسله مراتبی [۶۸] در کنار قابلیت‌های توازی عملیات [۵۹] همراه تلفیق شیءگرایی/عامل‌گرایی [۶۲] با ساختار توزیعی [۶۳] به موقعیت‌یابی متون می‌پردازد. آن‌ها باید بتوانند تنوع و درون‌زایی [۶۲] را در ساختار عامل‌گرایی هوشمند با توزیع پردازش‌ها [۷۱] در کنار قدرت پردازش موازی [۶۰] در اختیار بگیرند و بر چالش‌های داده‌های عظیم طبیعی در راستای نگرش ابر رایانش [۶۵] غلبه کنند. لذا ارائه روش‌های عمومی دقیق تشخیص تصاویر متنوع طبیعی با زبان‌های مختلف دارای سطوح تشخیص توده متن یا بلاک داده، پاراگراف، خط متن، کلمه و حرف بسیار پیچیده است. این روش‌ها باید ضمن حل چالش‌های قبلی، از کاهش بسیار زمان جستجو و افزایش قدرت زیاد پردازش نیز برخوردار گردند. سیستم‌های تولیدی نوین برای پشتیبانی از داده‌های تصویری در مقیاس بزرگ باید از قابلیت تنوع زیاد متون مورد جستجو و حجم عظیم داده‌های مورد پردازش پشتیبانی کنند. آن‌ها باید از ساختار اساسی پشتیبانی از چالش داده‌های متنوع حجیم و ارائه روش مقیاس‌پذیر برخوردار گردند.

لذا ارائه روش‌های عمومیِ تشخیص تصاویر متنوع طبیعی با نتایج دقیق در کلان داده‌ها مخصوصاً از نوع تصویری که باید ضمن حل چالش‌های قبلی و توجه به افزایش زیاد قیود حل مسئله و رشد بسیار داده‌ها، از کاهش شدید زمان جستجو و افزایش قدرت پردازش برخوردار گردند. باید بتوانند ضمن پشتیبانی از چالش تنوع داده‌ها، حجم آن‌ها و برآوردن قیود جدید برای تخصیص منابع مانند هزینه توان مصرفی از روش مقیاس‌پذیر برخوردار باشند [۴۲، ۴۱]. همچنین باید از تنوع داده‌های مختلف در ابعاد متفاوت پشتیبانی کنند و از سرعت لازم برای تشخیص سریع بی‌درنگ بهره‌مند شوند تا مناسب کاربرد دنیای آینده گردند [۶۵].

چند کار جدید وجود دارد که در راستای روش‌های پردازش‌های سریالی قدیمی فقط به مسئله پردازش داده‌ها در تصاویر حجیم حوزه پزشکی یا تعداد زیاد تصاویر همزمان در قالب ارائه قالب کار توجه کرده‌اند [۶۸، ۶۱]. آن‌ها فقط به افزایش سرعت محاسبات در قالب ارائه معماری موازی لوله‌ای پردازنده‌های گرافیکی و سیستم‌های توزیعی در پردازش‌های پزشکی پرداخته‌اند زیرا از حجم زیاد داده در فایل موضوعی خود برخوردار بوده‌اند. آن‌ها مسئله حجم داده در مقیاس بزرگ تصویر رنگی با رزولوشن 100000×100000 پیکسل را در حد ۳۰ گیگابایت تعریف کرده‌اند [۶۱]. در تحقیقی یک قالب کاری با مپ‌ردیوس[۷۷] ارائه شد که داده‌های آن دارای ساختار بزرگ تصویر نیست ولی تعداد تصاویر ورودی آن برای پردازش زیاد است [۶۸]. نتایج آن‌ها نشان از بهبود در سرعت پردازش داشت. گرچه در موضوع موقعیت‌یابی سعی بر ارائه روش‌های مؤثر مقاوم کار چندانی نشده، روش پیشنهادی [۸۸] مبتنی بر MSER با کاهش تکرارها و پیچیدگی مسئله توانسته سرعت خوبی ارائه کند. اما همچنان در مقابل چالش‌های اولیه متن از جمله انعکاس و کنتراست پایین و نیز چندزبانه بودن دارای اشکال است. برخی روش‌ها همچون [۸۹-۹۲] علی‌رغم غلبه بر برخی چالش‌ها، همچنان دارای ایراداتی از جمله سرعت پایین، دقت ناکافی، عدم تشخیص متون ریز و نقطه‌ها و یا تشخیص صرفاً تا مرحله خط متن و ضعف در تشخیص دقیق کلمات می‌باشند.

جدول۱: مقایسه الگوریتم‌های مکان‌یابی متون و روش‌های مهم.

الگوریتم	نقاط قوت	نقاط ضعف
[۵۰]	– مناسب تصویر با زمینه ساده. – عمل خوب در تصاویر طبیعی	– ضعیف در تصاویر با زمینه پیچیده. – فقدان مقیاس‌پذیری.
[۱۰]	– مناسب تصویر با زمینه ساده. – عمل خوب در تصاویر طبیعی.	– ضعیف در تصاویر با زمینه پیچیده. – متکی به قوانین تعریف‌شده دستی. – فقدان مقیاس‌پذیری.
[۵۱]	– مناسب آشکارسازی متون متحرک ویدئویی.	– تنها مناسب متون افقی.
[۵۳]	– مناسب تصویر با زمینه ساده. – عمل خوب در تصاویر طبیعی و ویدئویی.	– ضعیف در تصاویر با زمینه پیچیده. – مناسب متون افقی. – فقدان مقیاس‌پذیری.
[۵۷]	– آشکارسازی تصاویر با زمینه پیچیده. – عملکرد سریع.	– مناسب متون افقی. – فقدان مقیاس‌پذیری
[۵۴]	– مناسب آشکارسازی فریم‌های ویدئویی – آشکارسازی چند زبان مختلف.	– تنها مناسب متون افقی. – فقدان مقیاس‌پذیری.
[۵۵]	– عمل خوب در تصاویر طبیعی.	– مناسب متون افقی.
[۵۶]	– آشکارسازی در تصاویر با زمینه پیچیده.	– مناسب متون افقی. – نیازمند لغت‌نامه برای تصویر. – فقدان مقیاس‌پذیری.
[۱۱]	– آشکارسازی در تصاویر با زمینه پیچیده. – آشکارسازی در چند زبان مختلف. – عملکرد سریع.	– تنها مناسب متون افقی. – متکی به قوانین تعریف‌شده دستی. – فقدان مقیاس‌پذیری.
[۱۲]	– آشکارسازی در تصاویر با زمینه پیچیده. – عملکرد سریع.	– مناسب متون افقی. – فقدان مقیاس‌پذیری.
[۵۲]	– آشکارسازی در جهات مختلف. – آشکارسازی چند زبان مختلف.	– متکی به قوانین تعریف‌شده. – ضعیف در تصاویر با زمینه پیچیده.
[۲۱]	– آشکارسازی در جهات مختلف. – مناسب خط متن.	– متکی به قوانین تعریف‌شده. – فقدان مقیاس‌پذیری.
[۱۳]	– آشکارسازی سریع در جهات مختلف. – آشکارسازی چند زبان مختلف.	– متکی به قوانین تعریف‌شده. – فقدان مقیاس‌پذیری.
[۴۸]	– آشکارسازی در تصاویر با زمینه پیچیده.	– مناسب متون افقی. – متکی به قوانین تعریف‌شده. – فقدان مقیاس‌پذیری.
[۴۹]	– آشکارسازی در تصاویر طبیعی. – عملکرد خوب.	– تنها مناسب متون افقی. – فقدان مقیاس‌پذیری
[۵۸]	– آشکارسازی در تصاویر طبیعی. – تشخیص فقط خط متن. – روش کند با عملکرد خوب.	– یادگیری شبکه عصبی. – مناسب متون افقی. – فقدان مقیاس‌پذیری
[۴۰]	– روش مبتنی بر ناحیه و تصویر دوربین – متون هم رنگ غیر متنوع. – تفاوت کنتراست زیاد متن و زمینه.	– نیازمند دانش شکل کاراکترها. – عدم تنوع رنگ متون. – فقدان مقیاس‌پذیری.
[۸۸]	– سرعت بالا – پایین بودن میزان تکرار نواحی MSER – پیچیدگی محاسباتی پایین	– عدم پوشش چند زبانی – دقت کم برای متون با کنتراست پایین – اشکال در مقابله با چالش انعکاس نور – فقدان مقیاس‌پذیری
[۸۹]	– مستقل از نوع زبان – نرخ فراخوانی بالا – مقاوم در برابر چالش انحنای متون	– دقت پایین در تصاویر با پس‌زمینه پیچیده – فقط استخراج خط متن – مشکل متون دارای نقطه گاما – فقدان مقیاس‌پذیری
[۹۰]	– ساختار موازی بودن – غلبه بر چالش پس‌زمینه پیچیده	– عدم تشخیص متون ریز گاها – نیازمند کنتراست بالای متن و پس‌زمینه – فقط استخراج خط متن – فقدان مقیاس‌پذیری
[۹۱]	– مقاوم در برابر چالش انعکاس – مقاوم در برابر چالش وضوح پایین تصویر	– سرعت پایین به‌خاطر طولانی بودن مراحل * – فقدان مقیاس‌پذیری
[۹۲]	– سرعت بالا – عدم وابستگی به زبان و جهت متن	– تشخیص اشتباه مناطق غیرمتن – فقط استخراج خط متن – فقدان مقیاس‌پذیری
[۱۰۱]	– دقت و سرعت متوسط – عدم وابستگی به زبان، پهنای قلم و جهت متن	– سرعت بسیار کم – فقط استخراج متون ساده – فقدان مقیاس‌پذیری

در راستای حرکت به ساختارهای تحقیقات نوین، توجه به لزوم افزایش قیود مسئله به پیچیدگی بیشتر حل این مسئله کمک شده است. با رشد موبایل‌های ارزان بحث بهینه‌سازی منابع مصرفی مانند مصرف انرژی طرح شد. قیود جدید همراه با نیازمندی معیارهای نوین اندازه‌گیری آن، دامنه وسیع‌تری به تحقیقات قبلی در کاربرد داده‌های عظیم تصویری افزوده است [۶۹، ۶۶] که برای توزیع‌پذیری مسئله، تقسیم پردازش‌ها، هوشمندی اشیاء موجود در تصاویر، و جامعیت کاربرد روش در تصاویر طبیعی ساختار دیگری به نام تشخیص شیء طرح و در حال گسترش است تا اشیاء مختلف با متون متفاوت مشخص و تعیین موقعیت شوند [۶۷] . نهایتاً کلاس آن‌ها تفکیک گردد تا بتوانند به‌صورت مجزا درون‌یابی و ردیابی شوند و به دلیل پیچیدگی محاسبات آن برای داده‌های عظیم تصویری هنوز روش مناسبی ارائه نداده‌اند [۷۶] . گرچه در پژوهشی ادعا یافتن سریع اشیاء تصویر با روش نمونه‌برداری هرمی و با داشتن قید کاهش مصرف توان ارائه شده [۶۶] ولی مقایسه معتبری ارائه نداده است.

جدول۲: مقایسه الگوریتم‌های مکان‌یابی متن در داده‌های عظیم تصویری.

الگوریتم	نقاط قوت	نقاط ضعف
[۶۱]	– بررسی تصاویر کلان‌داده پزشکی. – افزایش قدرت پردازش.	– عدم تناسب جهت حوزه مکان‌یابی تصویر. – فقدان مقیاس‌پذیری.
[۶۸]	– قالب کار عمومی توزیعی موازی پردازش تصویر. – پردازش تصاویر مقیاس بزرگ توزیعی. – جداسازی داده‌های خالص و مشخصات تصاویر.	– عدم تناسب جهت حوزه مکان‌یابی تصویر. – کارایی نامناسب پردازش.
[۴۱]	– قالب کار عمومی توزیعی موازی پردازش تصویر. – روش مناسب تشخیص متن در جهات مختلف. – زبان‌های متفاوت، کاراکتر متفاوت و متون رنگی.	– خاص تصاویر ویدئویی. – نیازمند یادگیری و بر اساس بافت. – فقدان گزارش کارایی.
[۷۳]	– بهبود مکانیسم خوشه‌بندی **kmeans** – مصرف پایین حافظه	– عدم تناسب جهت حوزه مکان‌یابی تصویر – دقت و سرعت پایین
[۹۸]	– افزایش نرخ فراخوانی و دقت نواحی انتخاب شده	– نیازمند یادگیری (شبکه‌های عصبی)

روش‌های پردازش تصویر تشخیص اشیاء به چند دسته شکل‌گراء، حرکت‌گرا، رنگ‌گرا و بافت‌گرا تقسیم می‌شوند. روش‌های شکل‌گرا دارای زمان پردازش کم و دقت متوسط بر اساس الگوریتم‌های تطبیق‌هستند و مناسب حالت‌های پویا نیستند. روش‌های حرکت‌گرا دارای زمان زیاد پردازش و دقت متوسط بر اساس تشخیص حرکت هستند. روش‌های رنگ‌گرا به دلیل افزایش محاسبات بهبود کیفیت تصویر دارای زمان پردازش بالا ولی دقت زیاد هستند. روش‌های بافت‌گرا به دلیل بخش‌بندی تصویر و اشیاء تصویر دارای زمان پردازش بالا و دقت زیاد هستند [۷۶]. هویت‌بخشی به موضوعات در راستای مجرد سازی و قابلیت تشخیص جهت تسریع امور پردازش و تشخیص هم‌زمان اشیاء متفاوت در تصاویر با قیود جدید صورت می‌گیرد [۶۹] .

الگوریتم‌های تشخیص اشیاء دارای ماهیت هم‌زمانی و توزیعی هستند تا بتوانند مقیاس‌پذیری را مجسم کنند ولی به دلیل محاسبات پیچیده هنوز از قابلیت کاربرد در داده عظیم تصویری برخوردار نیستند [۷۰] . دامنه این مسئله به موقعیت‌یابی سنتی متن نیز کشیده شده است ولی هنوز با وجود کاهش محاسبات روش مناسبی در راستای کلان‌داده نیست [۷۱] . گرچه به سمت تشخیص سریع محل متن کارهایی ارائه شده ولی نیازمند یادگیری سیستم هستند و بر روی داده‌های کلان آزمون نشده‌اند [۸۰-۷۶] . بیشتر کارهای موجود سعی کرده‌اند که از ابزارهای موجود مانند مپ‌ردیوس با تقسیم مجموعه داده‌ها در توزیع دسترسی و پردازش جهت کاهش زمان استفاده کنند [۴۱، ۶۳] .

همچنین سعی کرده‌اند تا بر اساس ابزار قالب کاری متن‌باز آن یعنی هدوپ[۷۸] که برای پردازش انبوه متن ارائه شد با تبدیل داده‌های تصویر به جریان داده‌های باینری [۸۲] و سپس با پردازش ساختارهای پیش‌ساخته هدوپ به سمت پردازش بی‌درنگ مقیاس‌پذیر تصاویر بزرگ حرکت کنند [۸۱] . البته بعضی با ارائه فرمت I/O مرتبط با تصاویر در هدوپ به بهبود پردازش موازی و افزایش کارایی الگوریتم‌های پردازش‌تصویر این ابزار کمک کردند [۸۴-۸۳] . نیازهای جدید داده‌های عظیم تصویری در حوزه‌هایی مانند ساختمان [۷۴] بسیار سریع گسترش‌یافته ولی هنوز برای حوزه موقعیت‌یابی متن، مفهوم و ابعاد آن تعریف نشده است [۷۵]. برای مقایسه سریع روش‌های نوین در داده‌های عظیم تصویری و انتخاب استراتژی بر اساس شرایط تصویر، خلاصه نتایج بررسی نقاط قوت و ضعف مهم‌ترین تحقیقات در جدول۲ آمده است.

۳- مدل‌های موقعیت‌یابی

از نظر الگوی متن‌یابی روش‌های مهم موجود در سه گروه مبتنی بر بافت[۷۹]، مبتنی بر مؤلفه‌های همبند[۸۰] و روش‌های ترکیبی[۸۱] طبقه‌بندی می‌شوند [۱]. اساس روش موقعیت‌یابی مبتنی بر بافت، توجه به ساختار بافت تصویر است. با تحلیل بافت خاص هر تصویر و استفاده از بردارهای ویژگی استخراجی، متن موجود را (مثلاً با ماشین‌های یادگیری یا رده‌بند) می‌یابند [۳۵]. به‌طورکلی در زمینه‌های پیچیده، مدل‌های مبتنی بر بافت از مدل‌های مبتنی بر مؤلفه‌های همبند پایدارترند و بیشتر به‌سرعت پردازش توجه دارند. باتوجه به آنکه تصاویر دارای ویژگی متفاوت بافت با متن هستند روش‌های مبتنی بر بافت می‌توانند به‌راحتی متن را از زمینه تفکیک کنند. عیب اصلی مدل‌های مبتنی بر بافت، بخش‌بندی پیچیدهٔ بافت است که نسبت به چرخش و تغییر مقیاس متن حساس و لذا نیازمند پردازش‌های موازی متناسب با ابعاد تصویر است. روش‌های مبتنی بر بافت عمدتاً بر پایهٔ روش‌های آماری[۸۲]، طیفی و ساختاری[۸۳] بنا شده‌اند. با افزایش ابعاد تصویر و پیچیدگی متون کارایی آن در داده‌های عظیم کاهش می‌یابد و باید با استفاده از تکنیک‌های موازی، پردازش‌های محلی بیشتر و توزیع پردازش‌ها آن را ارتقاء داد.

۳-۱موقعیت‌یابی سنتی

مدل‌های مبتنی بر مؤلفه‌های همبند بر اساس تحلیل نظم هندسی لبه‌ها یا رنگ‌های همسان کاراکترها عمل می‌کنند. بیشتر روش‌های مبتنی بر ویژگی گرادیان و رنگ [۲-۷] به‌عنوان روش‌های مبتنی بر مؤلفه‌های همبند نیز دسته‌بندی می‌شوند. برای متون زیرنویس و با رنگ یکنواخت مناسب هستند ولی برای متون با کاراکترهای چند رنگی و پس‌زمینه‌های پیچیده در ابعاد بزرگ کارایی ندارند. بیشتر توجه آن‌ها به‌دقت و فراخوانی تشخیص است و فاقد سرعت لازم در داده‌های عظیم بر اساس انتخاب استراتژی‌های مناسب تصمیم‌سازی هستند. دسته‌های ترکیبی معمولاً ترکیبی از هر دو روش فوق هستند که از مزایای هر دو بهره می‌برند.

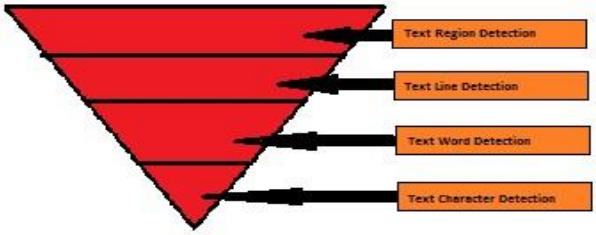
در سال‌های اخیر، روش ترکیبی به دلیل سختی کار، پیچیدگی‌های پیاده‌سازی، دقت نامناسب و سرعت کم در دو بعد کاراکتر و سطح کلمه به‌ندرت مورداستفاده قرار گرفته است [۴۲]. این روش‌ها باید برای کاربرد در داده‌های عظیم تصویر از مدل‌های عامل‌گرا و تکنیک‌های تشخیص اشیاء برخوردار شوند تا در جهت افزایش سرعت حرکت کنند. در بیشتر کاربردهای آینده مسئله حجم داده‌ها و ارائه پاسخ بی‌درنگ از چالش‌های عمده است. این امر نیازمندی به مدل، معماری، تکنیک و معیارهای نوین را طلب می‌کند تا بتوان روش‌های چابک برای تشخیص و موقعیت‌یابی متون ارائه داد. نگاه سلسله‌مراتبی کارهای تحقیقاتی موجود فاقد معماری و مدل مناسب برای توسعه و مقیاس‌پذیری هستند.

پس از مطالعه آخرین روش‌های علمی و مشاهدهٔ مشکلات موجود آن‌ها برای تصاویر ثابت [۳۸-۴۴] و ویدئویی [۲۶-۳۰]، فقدان روش عمومی دقیق که ضمن تشخیص در سطوح پاراگراف، خط متن، کلمه و حرف از سرعت کافی در تصاویر با متون و ابعاد زیاد مناسب کلان‌داده باشد احساس می‌شود. روش‌های موجود نتوانسته‌اند در دو جهت عمومیت روش و دقت کافی تشخیص خط متن و کلمه برخوردار باشند و از سرعت لازم تشخیص بی‌درنگ بهره ببرند. مثلاً بعضی الگوریتم‌ها

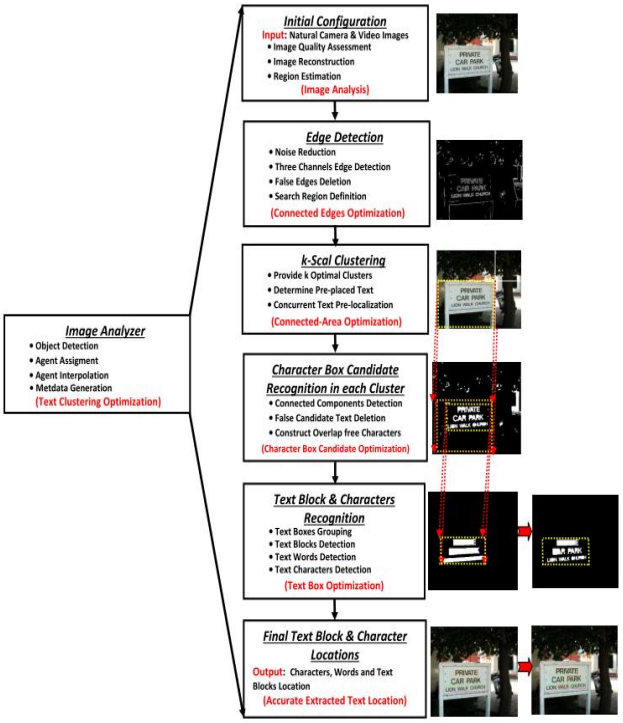
در سطح خط متن نتایج نسبتاً خوبی دارند ولی در سطح کلمه مطلوب نیستند [۵۵]. در کار دیگری در آغاز برای تولید کاراکتر کاندیدا از یک ساختمان داده موازی و سپس از یک فیلتر دو لایه برای حذف نواحی کاندیدای غیرمتن استفاده شد و با به‌کارگیری SVM نواحی متصل کاراکتری استخراج شد [۱۰۰]. برای از مشکلات عمده الگوریتم‌های سنتی فقدان مقیاس‌پذیری روش، سرعت محدود و برخورداری از پردازش‌های پیچیده را می‌توان نام برد.

۲-۳ موقعیت‌یابی داده‌های عظیم تصویری

دو مشخصه اصلی داده‌های عظیم تصویری عبارت از تنوع داده و حجم بسیار زیاد آن است. لذا یک سیستم کاربردی برای کلان داده‌ها باید بتواند تنوع و حجم زیاد داده‌ها را پشتیبانی کند. مقیاس‌پذیری روش برای تصاویر بسیار بزرگ با متون زیاد لازمه سیستم‌های آینده است و کاهش فضای جستجو[۸۴]در تصویر در کنار افزایش سرعت پردازش آن لازم است.



شکل۵: نگاه سلسله‌مراتبی به موقعیت‌یابی متن در داده‌های عظیم تصویری.



شکل۶: نمونه مراحل تشخیص متون داده‌های عظیم تصویری در پردازش بی‌درنگ.

روش‌های جدید تشخیص متن در حال تمرکز بر کاربرد در داده‌های عظیم تصویری هستند و از حجم زیاد داده‌ها در کنار تنوع زیاد آن‌ها برخوردارند [۴۱-۴۲]. برای این مسئله می‌توان تشخیص بی‌درنگ را اضافه کرد که سیستم‌های موجود عموماً فاقد آن هستند. باید با توجه به چالش‌های مطرح‌شده به برخورداری از قابلیت معماری موازی در ساختارهای مدل‌سازی مانند مدل‌های توزیعی و بستر اجرا مانند ماشین‌های موازی در کنار بی‌درنگی توجه شود.

داده‌های عظیم تصویری درسه بخش داده‌ها (ماهیت وحجم داده‌های واقعی)، پردازش (پیچیدگی پردازش‌ها و منابع موردنیاز) و مدیریت (داده‌های واقعی و متاداده‌های کنترلی) دارای چالش‌های اساسی هستند. در این مقاله بیشتر به‌تعریف از جنبه داده‌ها و پردازش توجه‌می‌شود وبه جنبه مدیریت فقط از نظر لزوم وجود ساختار متاداده در کلان‌داده فقط اشاره خواهد شد که در پردازش‌های با کارایی بالا لازم است.این امر منجر به‌جداسازی پردازش شناسنامه داده‌ها، محتوای آن‌ها و جستجوی مداوم اخذ دانش در تولیدات حجیم داده‌های جدید می‌شود [۵۹، ۷۱، ۷۴]. از یک نگاه در حوزه تصاویر عظیم جهت تشخیص کلان داده‌ها، ملاک‌های ارزشمند تشخیص فرآیندهای مهم پردازشی شامل نوع‌فرآیند (پردازشی یا استنباطی)، نوع‌عملیات آن (پردازشی یا حافظه‌ای)، ناحیه تحت‌پوشش (اختصاصی یا اشتراکی)، تعداد عملیات (سری، بازگشتی یا موازی)، مراحل عملیات (سبک یا سنگین)، درجه اهمیت هر پردازش (اساسی یا بهبودسازی) و معیارهای سنجش آن است که سرعت تشخیص را در روش توزیعی بی‌درنگ مبتنی بر مولفه‌های‌همبند تحت‌الشعاع قرار می‌دهد. در روش مورد بررسی برای نگاشت در داده‌های عظیم تصویری، تصاویر برای خواندن از دو بخش غیرفعال ِمرده زمینه و بخش فعال ِزنده متن تشکیل می‌شوند. معمولاً زمینه تصاویر به دوبخش زمینه‌های مضر فاقد اطلاعات و زمینه‌های‌مفید دارای اطلاعات جهت کمک غیرمستقیم به تشخیص متن تقسیم می‌گردد. بخش‌های مضر دارای ارزش پردازش نیستند و باید حذف شوند. اما نواحی-زمینه مفید دارای اطلاعاتی هستند که به یافتن سریع نواحی‌متنی کمک می‌کنند و باید تقویت شوند. در سیستم‌های یادگیر این نواحی بسیار مهم از نواحی مهم غیرمتن تلقی می‌شوند که معمولاً زمینه‌های‌پیچیده یا شلوغ در تفکیک و تشخیص این نواحی، ایجاد خطا می‌کنند.

معمولاً تصاویر طبیعی از وضوح مناسب زمینه برخوردار نیستند و شلوغی و تنوع متون آن، عملیات موقعیت‌یابی متن را بسیار دشوار می‌سازد. در نتیجه از اعمال مهم پردازشی در داده‌های عظیم تصویری، جداسازی و تفکیک نواحی غیرفعال جهت حذف و نواحی فعال جهت خواندن است. متونی مانند کتاب یا صفحات وب در بعضی موضوعات تصویری از شفافیت و درجه وضوح مناسب زمینه برخوردارند که از این نظر به پردازش کمتری نیاز دارند. معمولاً پردازش‌ها در این بخش باید در جهت تضعیف نواحی مرده و تقویت نواحی زنده تصویر صورت گیرد. در تصاویر پیچیده نسبت نواحی زنده به نواحی مرده بسیار بالا است که معیار جدیدی برای سنجش آن باید تعریف گردد. می‌توان این معیار را با عنوان درجه تفکیک‌پذیری نواحی زمینه مرده مضر و نواحی زمینه زنده مفید ذکر کرد که باید نواحی دارای متن را کاملاً مشخص کنند. ضمناً ارزیابی قدرت پردازش هر روش در این حوزه علاوه بر درجه تفکیک‌پذیری به میزان و درجه تضعیف بلاک‌های مرده، میزان تقویت بلاک‌های زنده و درجه توازی عملیات نیز مرتبط می‌شود.

نگاه دیگر به ماهیت سلسله‌مراتبی تشخیص موقعیت متون در تصویر و ادراک آن جهت خواندن در یک سیستم کلان موقعیت‌یابی است. پایه تمامی سیستم‌های موقعیت‌یابی حرف، کلمه، خط متن و پاراگراف در انواع متون مختلف دست‌نویس و تایپی هستند که بلاک‌های مهم متنی در سطوح مختلف تشخیص و خواندن متن را تشکیل می‌دهند. معمولاً واژه پاراگراف و یا خط متن فقط برای تشخیص نواحی کلان متنی و واژه کلمه برای نواحی خرد متنی لازم است. شکل۵ مراحل نگاه منطقی سلسله مراتبی به متون و سطوح مختلف ادراکی آن را برای کلان‌داده نشان می‌دهد. در پایین‌ترین سطح یا سطح اتمیک حرف، شکل حروف، نوع آن یا علایم هر زبان بسیار مهم هستند. در سطحی بالاتر یا سطح کلمه مجموعه‌ای از حروف یا علائم اساس کار هستند. در سطح میانی یا سطح خط متن اساس کار مجموعه کلمات هستند. در بالاترین سطح مجموعه خطوط متنی به‌عنوان پاراگراف، مبین بلاک‌های مهم متنی می‌باشند. در شکل۶ مراحل پردازشی و سطوح مختلف آن نشان داده شده است. این نگاه در داده‌های عظیم تصویری جهت کاهش پیچیدگی پردازش‌ها و تقسیم‌بندی عملیات برای تعیین میزان توزیع و موازی‌سازی آن‌ها کمک می‌کند. معمولاً پردازش‌های این بخش باید در جهت تفکیک نواحی حرفی برای تشخیص

کلمات و تقویت آن‌ها در خواندن نوری تصویر صورت گیرد. در تصاویر پیچیده نسبت وضوح حروف می‌تواند معیاری جدیدی برای سنجش عملکرد هر روش باشد. می‌توان این معیار را با عنوان درجه تفکیک‌پذیری حروف از زمینه ذکر کرد که باید نواحی دارای حرف را کاملاً تفکیک کند. ضمناً قدرت پردازش هر روش در این حوزه علاوه به درجه تفکیک‌پذیری حروف، به میزان و درجه تقویت آن‌ها نیز مرتبط می‌شود.

انواع مختلف متون مانند متون دست‌نویس، متون دارای حرف غیر هم‌راستا، و متون دارای حروف تبلیغاتی از پیچیدگی زیادی برخوردارند. هرچه درجه تفکیک-پذیری روش بالاتر باشد در سطح کلان‌داده دارای کمک زیادی به کارایی نیست مگر آن که درجه توازی عملیات بیشتر باشد. از این منظر در داده عظیم تصویری از سطوح بالا یعنی سطح پاراگراف عملیات سلسله‌مراتبی تشخیص در هر سطح برای موازی‌سازی مجازی خط لوله[۸۵] شروع می‌شود. پس از تشخیص بلاک‌های متنی در سطوح بالاتر جزئیات در سطوح پایین‌تر نمایان می‌گردند که در هر سطح معیارهای تشخیص تابع سطحِ خود هستند. معیارهای ارزیابی هر روش بر اساس سطح پردازش و درجه تفکیک‌پذیری در همان سطح تعیین می‌گردد. مثلاً در سطح پاراگراف درجه تفکیک بالاترین سطح بلاک‌های متن مهم است. در ساختار ویدئویی این نگاه به سرعت پردازش کمک بسیار می‌کند و در تصاویر ثابت، سرعت تزریق و توالی تصاویر بر سرعت تولید نتایج می‌افزاید. در این مقاله ما با مطالعه و پیاده‌سازی مقالات مهم، به بررسی موقعیت‌یابی مبتنی برای داده‌های عظیم تصویری متمرکز شده‌ایم [۶۱-۶۳]. ارائه روش‌های عمومی قابل‌اعتماد و سریع تشخیص متون متنوع در تصاویر داده‌های عظیم تصاویر طبیعی که بتواند از مزایای تمامی روش‌ها در سطوح تشخیص بلاک، خط متن، کلمه و حرف برخوردار باشد. با غلبه بر مشکل پیچیدگی متون و افزایش ابعاد تصویر ورودی، نتایج آن با روش‌های سنتی قابل رقابت باشد. روش‌های نوین با برخورداری از ساختارهای سلسله‌مراتبی یا موازی باید دارای قابلیت پردازش موازی یا توزیعی عملیات در کنار کاهش پیچیدگی‌های پیاده‌سازی باشند تا بتوانند از دقت و سرعت در سطوح مختلف تشخیص برخوردار گردند.

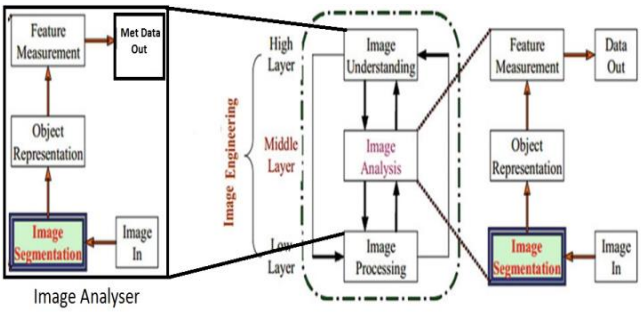
روش‌های عمومی باید در برابر چالش‌های نورپردازی، ناهمواری زمینه و غیر یکنواختی آن، سطوح ناهموار، تغییرات شدید رنگ، مقیاس متفاوت فونت، چند زبانی و پیچیدگی پس‌زمینهٔ تصویر مقاوم باشند. همچنین با هدف حرکت به سمت پشتیبانی از داده‌های عظیم تصویری، ارائه روش‌های نوین سریع چابک برای تشخیص‌های سریع اولیه در جهت کمک به رفع چالش‌های فوق در تصاویر مناظر طبیعی ضروری است. برای کاربرد وسیع آن باید از قابلیت شناسایی تصاویر دوربین و ویدئو و همچنین قابلیت پیاده‌سازی بر ساختارهای موازی برخوردار گردند. لذا تمرکز اصلی دنیای آینده برکاهش فضای جستجو برای افزایش سرعت‌تشخیص در کنار توزیع پردازش‌های زیاد و پیچیده برای برخورداری از پاسخ بی‌درنگی لازم است [۶۳].

۳-۳ مدل‌سازی مخاطرات عملیات

باید به تناسب اهداف پردازش در هر کاربرد تصویری و سطوح زیرین یا بالاترین آن و همترازی‌های در هر سطح با تعریف تهدیدات نسبت با ارزیابی عملیات و عدم ورود به منطقه ریسک مطمئن بود. تهدیدات موجود از تصویربرداری تا خطاهای انسانی که در عملیات مخاطره ایجاد می‌کند تعریف شود. ردیابی و کنترل از تهدیدات ممکن تا توقف عملیات بی‌درنگ باید در هر سیستم صورت گیرد و پایش شود تا اصلاحات یا استراتژی‌های مرحله‌ای ممکن انتخاب گردد. با ارزیابی مداوم می‌توان از ورود به وضعیت ریسک در عملیات بی‌درنگ پیشگیری کرد و در صورت عدم امکان محاسبات ریسک را وارد عملیات کرد. نگاه سلسله مراتبی شکل۵ می‌تواند از عملیات بی‌درنگ تا تصمیمات استراتژیکی را تحت‌الشعاع قرار دهد. ارائه مدل مناسب این امر لازم، محاسبات مداوم آن حیاتی، و توجه به تهدیدات و جلوگیری از تبدیل به مخاطرات و خطرات بالاتر در کنار تعریف معیارهای اندازه‌گیری در آینده با توجه به علایق عامل‌ها در تصمیم‌سازی، استفاده از تکنولوژی‌های نانو در تجهیزات نانو و کاهش شدید زمان بین خرابی‌ها اجتناب‌ناپذیر است. این مدل برای تحلیل‌گر استراتژیک سیستم پردازش لازم و مفید است.

۴- موقعیت‌یابی بات‌گرا

تشخیص دقیق و سریع محل متون و تعیین سطح تشخیص متن در محیط‌های واقعی مانند لوگوها، تابلوها و تصاویر طبیعی مختلف، وابستگی بسیاری به ابعاد تصویر و میزان تغییرات متن، موقعیت دوربین، پیچیدگی تصویر، تغییرات شدت روشنایی و نور محیط، میزان ضخامت متون یا پهنای قلم، زاویه دید دوربین و بستر مسطح یا غیرمسطح تصویر دارد.



شکل۷: دیاگرام بلاکی روش عامل‌گرای موقعیت‌یابی متن در تصاویر داده‌های عظیم.

روشی جامع برای دقت بالای تشخیص در سطوح بلاک و کلمه مناظر طبیعی در پس‌زمینهٔ پیچیده دارای قابلیت شناسایی تصاویر ویدئویی وجود ندارد ولی می-توان ساختاری مطابق دیاگرام شکل۶ برای تمامی روش‌ها ارائه داد. ممکن است بعضی از آن روش‌ها با انتخاب استراتژی مناسب برای بهبود سرعت، بخشی از پردازش‌ها مانند پیش‌پردازش یا بهبود متون را حذف کنند. دو ویژگی بارز در متون تصاویر وجود یک‌پارچگی و کنتراست خوب متن با زمینه، و شکل نواحی بسته لبه-های متون است. با توجه به این ویژگی‌ها، از استخراج لبه و ویژگی بسته بودن کاراکترهای متن جهت شناخت کاندیداهای اولیه متون استفاده و برای افزایش سرعت، آن را مقاوم‌سازی می‌کنند. ابتدا لبه‌های تصویر تعیین و سپس با تقسیم تصویر، کاراکترهای کاندیدا را تشخیص دهند. نهایتاً محل دقیق کاراکترها و کلمات و بلاک‌های متنی را تعیین و قاب‌گذاری می‌کنند که در ادامه به شرح جزئیات تکنیک‌های آن می‌پردازیم. البته ما ساختار هوشمندی برای داده‌های عظیم تصویری بر اساس انتخاب استراتژی مناسب تصمیم‌سازی ارائه داده‌ایم (شکل۷) که در زیر به مفاد مهم آن نیز می‌پردازیم.

۴-۱ پیش‌پردازش

معمولاً تحلیل زمینه تصویر روشی سلسله‌مراتبی، اساسی و لازم برای بهبود شانس تشخیص متون برای تصاویر خام است که کارایی آن باید برای کلان‌داده با نگاه به میزان عملیات کلان سراسری و عملیات محلی بررسی شود. این کار اساساً به دلیل اموری مانند روش نوشتن، سرعت نوشتن، اندازه حروف، شرایط فیزیکی و ذهنی نگارنده، هم‌پوشانی حروف و مانند آن صورت می‌گیرد تا برای تطبیق تصویر با موضوع مورد بررسی، آن را برای پردازش آماده سازد [۴۰] . کارهایی از قبیلِ حذف نویز، باینری‌سازی، اسکلت‌یابی، نازک‌سازی، تشخیص لبه، نرمال کردن، و تغییر اندازه از موارد پیش‌پردازش هستند. از آنجا که برای تعیین دقیق محل متن تصاویر باید کل تصویر مورد بررسی قرار گیرد این عمل زمان‌بر و از گره‌گاه‌های اصلی کلان داده‌ها است. اگر در انجام عملیات پیش‌پردازش بتوان اطلاعات اصلی موقعیت نسبی متون در تصویر را کسب کرد این اطلاعات اضافی کمک بسیاری به افزایش سرعت یافتن

محل متن خواهد کرد. این عمل گراف‌گرا علت تولید اطلاعات اضافی زیاد به‌عنوان متاداده است و زمینه مصرف زیاد حافظه سیستم پردازشگر و در نتیجه مصرف توان بیشتر را فراهم می‌سازد.

هوشمندی حدس نواحی غیرمتن و حذف تصویر آن‌ها از ساختار این گراف با استفاده از ساختار عامل‌ها یا سرورهای بات می‌تواند افزونگی آن را کاهش دهد. در پردازش کلان‌داده تولید این اطلاعات ظاهراً اضافی (متاداده) در کنار تغییر اندازه ناحیه یافت‌شده، ارزش‌گذاری نواحی، و نگهداری آن‌ها در ماشین‌های جدا کمک‌شایانی به‌تقلیل نواحی غیرمتن (فضای‌مرده) و نگهداری اطلاعات واقعی تصویر (فضای‌سفید) می‌کند. لذا در بالاترین سطح مدل‌سازی یعنی سطح ماکرو یا کلان (یافتن بلاک‌های کاندیداهای متون)، مسئله متاداده و ساختمان داده‌های لازم آن نیاز به پژوهش جدید دارد. این ساختارها می‌توانند در پایین‌ترین سطح مدل‌سازی یعنی سطح میکرو یا محلی تقلیل‌یابند یا حذف‌شوند. در ردیابی تصاویر ویدئویی متون، تشخیص فضای‌مرده می‌تواند به‌سرعت انجام گیرد.

در داده‌های عظیم تصویری تشکیل متاداده‌ها در سطح خوشه‌بندی، لازم و یافتن ساختمان‌داده مناسب هر سطح آن بسیار حیاتی است. در این راستا جداسازی اطلاعات پس‌زمینه تصویر (فضاهای‌سرد یا مرده)، اطلاعات متون (فضاهای‌داغ یا سفید) و اطلاعات‌مشکوک به‌متن (فضاهای‌گرم یا بایر) همچنین طراحی ساختمان داده‌های مناسب و جدا برای هر کدام ضرورت دارد. لذا در سطح ماکرو نیازمند روش‌های چابک به‌عنوان تحلیل‌گران استراتژیک تصویری برای تعیین‌سریع فضاهای مرده و در سطح میکرو نیازمند روش‌های یادگیر برای تعیین دقیق فضاهای داغ هستیم. برای تعیین نواحی از روش‌هایِ‌سریعِ ساده و برای تشخیص متون از روش‌های دقیقِ پیچیده می‌توان استفاده کرد. از پیش‌پردازش‌های مهم استفاده از موجک[۸۶] است که امکان لبه‌یابی در جهات مختلف و امکان فشرده‌سازی را نیز فراهم می‌کند. برای داده‌های عظیم تصویری (شکل۵) می‌توان گفت که در بعضی موارد مانند تصاویر پزشکی و کاربردهای آن نیازمند پیش‌پردازش هستیم و نمی‌توان این موارد را برای افزایش سرعت حذف کرد. پیش‌پردازش را برای کاهش فضای جستجو، نویزگیری، پیش‌بینی سیستم پردازش و غیره می‌توان به‌کار برد. از آنجا که در کاهش فضای جستجو نقش موثری ایفا میکند صرفنظر کردن از آن جهت افزایش سرعت به پردازش کمکی نمی‌کند مگر تصویر اولیه دارای بافت سبک و زمینه‌های ساده باشد.

۴-۲ خوشه‌بندی و شیءیابی

اشیاء نواحی با ارزشی هستند که مجموعه خط متن‌ها یا کلمات را تشکیل می‌دهند. برای یافتن اشیاء نیازمند خوشه‌بندی کارآ و حل مسئله عدم اطمینان در داده‌ها هستیم. فرآیند تجمیع عناصر بدون برچسب که به اشیاء مشابه گروه‌بندی می‌شوند و مطابق شباهت‌شان برچسب می‌خورند را خوشه‌بندی گویند. ابعاد خوشه‌بندی برای متون بسیار بزرگ و هنگامی که خوشه‌بندی یا توئیت می‌شوند چالش جدی محسوب می‌شود. برای خوشه‌بندی از روش‌های مختلف استفاده می شود [۹۶] که همزمانی آن‌ها مهم است [۹۷] و اکنون تئوری مجموعه‌های خشن[۸۷] برای کاهش ابعاد آن کاربرد یافته است [۹۳] . این روش‌ریاضی بالاترین و پایین‌ترین تقریب مجموعه برای حل عدم‌اطمینان در داده‌های خوشه‌بندی مستند بدون اطلاعات اضافی از پیش‌تعیین‌شده درباره داده‌ها را می‌دهد. در این تئوری، داده‌ها به‌عنوان یک سیستم اطلاعاتی به‌صورت جدول تصمیم‌گیری که ردیف‌های آن اشیاء و ستون‌های آن ویژگی‌های آن‌ها هستند به‌روش بدون‌ناظر برای متن‌کاوی ارائه می‌شوند. این ویژگی‌ها به‌سه تقریب یعنی (۱) تقریب‌پایین برای عناصر از طریق جمع‌آوری داده‌ها، (۲) تقریب بالا برای اشیاء با احتمال زیاد و (۳) ناحیه مرزی که تفاوت بین بالا و پایین است تقسیم می‌کند [۹۴] .

برای پردازش داده‌های عظیم تصویری ابتدا باید ابعاد تصویر را کاهش‌داد و با تقسیم نواحی دارای متن با خوشه‌بندی و تشخیص اشیاء با تخصیص عامل هوشمند به هر شیء بتوان عملیات همزمان پردازش درون‌یابی را برای کاهش پیچیدگی‌تشخیص تشکیل‌داد. لذا روش چاپک بسیار سریع برای تعیین نواحی بالفعل و بالقوه دارای متون برای خوشه‌بندی همزمان[۸۸] جهت یافتن الگوهای مخفی، تقلیل ابعاد، ارائه آسان و تقلیل زمان اجرا لازم است [۹۵] . حتما لازم نیست که دارای دقت بسیار زیاد باشد بلکه باید به‌سرعت بتواند نوحی دارای متن (نواحی‌سفید)، یا ناحیه فاقد متن (نواحی مرده) و نواحی بالقوه‌متن (نواحی بایر) را تعیین کند. این نگاه می‌تواند به‌صورت توزیعی و با پخش‌کردن این پردزاش در نواحی یک تصویر عظیم و همزمان صورت گیرد. البته در این بخش‌بندی تصویر باید با نظارت عامل‌های هربخش، عملیات مشاوره عامل‌ها برای تشخیص نواحی پیوسته‌متنی جهت ادغام نواحی مجزا، صورت گیرد. لذا پردازش سریع برای تعیین و بخش‌بندی همراه برچسب‌زدن نواحی کلان تصویر لازم است. این روش می‌تواند با تعیین خوشه‌های لازم و بهینه‌سازی تعداد آن‌ها پایان یابد. در نهایت این بذرها یا اشیاء در اختیار پردازشگرهای موازی درون‌یاب بعدی قرار می‌گیرد.

۴-۳ درون‌یابی‌اشیاء

بسیاری از روش‌ها برای یافتن متون بالقوه و مرزهای اشیاء از لبه‌یابی استفاده می‌کنند تا لبه‌های متن را به‌صورت محیط‌های کاملاً بسته آشکار سازند که دارای مزایای و معایب بسیاری هستند. با توجه به نقاط ضعف و قوت بعضی لبه‌یاب‌ها، برای بهره‌برداری از قدرت همه آن‌ها و برای یافتن دقیق‌تر لبه‌های‌متن به‌شکلمحیط‌های‌بسته از ترکیب هم‌زمان آن‌ها استفاده می‌شود. مثلاً پس از حذف نویز در هر سه کانال رنگی، با لبه‌یابی در هر کانال و ادغام آن‌ها تصویرلبهٔ مطلوب با هدف آماده‌سازی تصویر برای یافتن و تقلیل نواحی متن حاصل می‌شود. مزایای لبه‌یابی عبارتند از کاهش فضای جستجو در اثر حذف نواحی غیرلبه و خروجی باینری است. معایب آن عبارت از افزایش زمان پردازش، از دست رفتن اطلاعات رنگ، امکان ایجاد گسستگی در لبه و مانند آن است.

در الگوریتم‌هایی که زیربنای پردازش مبتنی بر مولفه‌های متصل‌به‌هم تعریفی برای دست‌یابی به مولفه‌ها، نیاز به در اختیار داشتن تصویر لبه باشد لبه‌یاب‌های متعددی وجود دارند که بسته به نوع نیاز از یک یا ترکیب چند نوع آن‌ها میتوان استفاده کرد. از معروفترین لبه‌یاب‌های درجه اول لاگ[۸۹]، رابرتز[۹۰]، زیروکراس[۹۱]، کنی[۹۲] و سوبل[۹۳] و برای لبه‌یاب‌های درجه دوم لاپلاس[۹۴] و مارهیلدرث[۹۵] را نام برد. بحث عدم تغییر لبه در نواحی مرده مبین نواحی نامناسب برای پردازش است. اگر در این مرحله با دستیابی به روش مناسب بتوان متون نواحی‌سفید را تقویت کرد می‌توان با وجود داشتن پردازش اضافی، مراحل بعدی پردازش و یا بخش‌هایی از آن را کاهش داد. اکنون در روش‌ها نوین جهت افزایش سرعت و دقت باید از تئوری مجموعه‌های خشن برای لبه‌یابی استفاده شود.

۴-۴ حذف شیء غیرمتن

پس از حدس نواحی متون بالقوه (سفید و بایر) باید به‌بررسی و پالایش دقیق جهت حذف بلاک‌های پیشنهادی غیرمتنی اقدام شود. حذف کاندیداهای غیرمتن یا فضای‌مرده تصویر برای برخورداری از دقت بالای تشخیص با سه تکنیک صورت می‌گیرد. این تکنیک‌های حذف لبه‌های‌غیرمتن بر اساس مشخصات همسایگی، مشخصات هندسی و مشخصات پهنای‌قلم است که در ادامه به شرح آن‌ها می‌پردازیم.

حذف همسایگی: آشکارسازی لبه با برچسب‌گذاری در همسایگی مؤلفه‌های همبند و اطلاعات بلاک‌های مجاور صورت می‌گیرد. هر عنصر بدست‌آمده را بلاک لبه[۹۶] می‌نامیم. فرضیات تجربی و علمی برای حذف محل‌های غیرمتن اعمال می‌-

کنیم. نسبت ابعاد بلاک‌های بین دو حد را برای بررسی نواحی باریک در نظر می‌گیرند که ابعاد EB باید بزرگتر از تعدادی پیکسل و کوچکتر از درصدی از ابعاد تصویر باشد. از آنجا که با لبهٔ آشکارشده، کناره‌های داخلی و خارجی کاراکترها را بدست می‌آورند لذا ممکن است که هر EB دارای یک یا چند EB دیگر باشد. مثلاً حرف "O"، به‌علت مرز داخلی EB_{int} و به‌علت مرز خارجی EB_{out} به دو عنصر تقسیم شود. لذا احتمال دارد که هر EB کامل، یک یا چند EB را احاطه کند. مثلاً حرف B سه مؤلفه، دو عدد برای مرزهای داخلی EB_{int} و سومی برای مرز خارجی EB_{out} دارد. اگر هر EB خاص بیش از دو EB داشته باشد که کاملاً داخلش باشند تنها EB داخلی می‌ماند و EB خارجی حذف می‌شود. بنابراین لبه‌های مؤلفه‌های غیرمتن با درنظرگرفتن محدودیت هر لبهٔ مؤلفه با شرط زیر حذف می‌شوند.

If (N_{int} > n) { Accept : EB_{int}, Reject: EB_{out}}

که در آن EB_{int} به EBهایی که کاملاً در EB فعلی قرار دارد و مساحت آن‌ها بزرگتر از حد خاصی است و N_{int} به‌تعداد EB_{int} اشاره می‌کند. محدودیت‌های روی لبهٔ مؤلفه‌های کاندیدا، مؤلفه‌های غیرمتن را حذف می‌کند. گودال‌های جامانده پس از حذف لبه‌های غیرمتن با استفاده از عملیات مورفولوژی باید پرُ شوند تا کاندیداهای اولیه متن به‌دست می‌آیند.

حذف هندسی: برای حذف نواحی غیرمتن، مجموعه‌ای از ویژگی‌های هندسی انعطاف‌پذیر مساحت‌ناحیه، گریز از مرکز[۹۷] (نسبت طول محور اصلی به‌محور فرعی) رابطه(۲) و استحکام[۹۸] رابطه(۳) بر روی هر مؤلفهٔ‌همبند تعریف می‌شوند. در ابتدا همهٔ مؤلفه‌های بسیار کوچک پراکنده و سپس مؤلفه‌های بزرگتر و با طول‌زیاد حذف می‌شوند. با انتخاب آستانهٔ محافظه‌کارانه می‌توان مطمئن شد که بعضی حروف‌کشیده مانند i و l حذف نخواهند شد. در انتها مؤلفه‌های دارای تعداد زیاد حفره، حذف می‌شوند زیرا این مؤلفه‌ها شبیه به کاندیدهای متن نیستند. قوانین تنظیم ویژگی‌های هندسی‌تجربی (مساحت، گریز از مرکز و استحکام) جهت حذف مولفه‌های غیرمشابه متن عبارت از مؤلفه‌هایی با مساحت خاص، گریز از مرکز خاص و استحکام کمتر حد خاصی است.

$$Eccentricity_i = \frac{a_1}{a_2} = \frac{\mu_{20}+\mu_{02}+\sqrt{(\mu_{20}+\mu_{02})^2+4\,\mu_{11}^2}}{\mu_{20}+\mu_{02}-\sqrt{(\mu_{20}+\mu_{02})^2+4\,\mu_{11}^2}} \qquad (۲)$$

$$, r_b = 2\left(\frac{\lambda_1}{|R|}\right)^{\frac{1}{2}} = 2\left(\frac{2a_2}{|R|}\right)^{\frac{1}{2}} r_a = 2\left(\frac{\lambda_1}{|R|}\right)^{\frac{1}{2}} = 2\left(\frac{2a_1}{|R|}\right)^{\frac{1}{2}}$$

$$ComponentStrength_i = \frac{ComponentArea_i - HoleArea_i}{ComponentArea_i} \qquad (۳)$$

حذف پهنای قلم: قلم را ناحیه پیوسته تصویر که پهنایش تقریبا ثابت است تعریف می‌کنند. عملگر پهنای قلم به‌وسیلهٔ تغییرات فاصله (فاصلهٔ اقلیدسی هر پیکسل پیش‌زمینه نسبت به‌نزدیک‌ترین پیکسل‌زمینه) برای هر پیکسل محاسبه و با مقدار محاسبه‌شده برچسب می‌خورد. پهنای واقعی را نمی‌دانند و آن را با لبه‌یاب‌ها محاسبه (بر اساس جهت‌گرادیان d_p برای هر پیکسل‌لبه p بیان) و سپس بازیابی می‌کنند. اگر p در راستای‌قلم باشد پس d_p عمود بر مسیر قلم است. با دنبال‌کردن شعاع $r = p + n.d_p$ تا یافتن پیکسل‌لبه q ادامه می‌دهند که سگمنت $[p, q]$ با عرض $\frac{\longrightarrow}{\|p-q\|}$ تعریف می‌شود. در برخی تصاویر مانند مارک‌ها و برچسب‌های تجاری به دلیل آن است که کل‌متن یک (یا دو) ناحیه‌بسته تلقی می‌شود لبه‌های متون هم‌پوشانی دارند. در مرحلهٔ حذف لبه‌های غیرمتن، قاب اطراف متن‌حذف نمی‌شود. لذا با عملیات مورفولوژی و منطقی، لبه‌های مؤلفه را از مؤلفه جدا سپس پهنای قلم را برای هر مؤلفه محاسبه و مؤلفه‌هایی که انحراف معیار زیادی دارند حذف می‌کنند. عملیات محاسبات پهنای قلم بسیار سنگین است و وجود قلم با پهناهای متفاوت دارای محاسبات بسیار پیچیده خواهد بود. این امر با داشتن اندازه‌ها و سبک‌های مختلف در یک زمینه پیچیده مشکل‌تر می‌شود.

۴-۵ پایش اشیاءمتنی

پس از حذف کاندیدهای اولیهٔ غیرمتن، با استفاده از ویژگی‌های متون (ارتفاع، طول و فاصله) و عملیات مورفولوژی بلاک‌های اولیه (سطح کاراکتر، کلمه و خط متن) تشکیل و بلاک‌های غیرمتن درسطوح‌مختلف و در طی چند مرحله حذف می‌شوند. پس از تشکیل بلاک‌های متن، عمل حذف غیرمتن در سطوح کلمه، خط متن و سپس کاراکتر مثلاً بر اساس اسکلت‌سازی انجام می‌شود.

حذف درسطح‌کلمه: با تشکیل بلاک‌های نامزد کلمه ما از ویژگی‌های چگالی-لبه، چگالی‌بلاک و نسبت ابعاد برای حذف بلاک‌های غیرمتن استفاده می‌کنیم. نواحی متنی طبق رابطه(۴) چگالی‌لبهٔ بیشتری نسبت به‌نواحی غیرمتنی دارند. معیار چگالی‌بلاک طبق رابطه(۵) بر این فرض استوار است که نواحی متنی دارای چگالی بلاک نزدیک به ۱ هستند. بلاک‌هایی با معکوس چگالی بلاک بزرگتر از یک حد و کمتر از حد دیگر بلاک‌های غیرمتن هستند حذف می‌شوند. بلاک‌های کلمه معمولاً به شکلمستطیل و نسبت طول آن‌ها به عرض‌شان بیشتر است. از ویژگی اختلاف ابعاد بلاک با توجه به مساحت‌کلمه نیز استفاده می‌شود. لذا بلاک‌هایی که دارای این شرایط باشند بلاک‌های غیرمتن تلقی و حذف می‌شوند.

$$EdgeDensity = \frac{\sum EdgeLength(Word_i)}{TotalWordAreaPixels} \qquad (۴)$$

$$BlockDensity(b_i) = \frac{WordArea(b_i)}{WordFrameArea(b_i)} \qquad (۵)$$

حذف‌درسطح‌کاراکتر: بعد از حذف در سطح‌کلمه، بلاک‌کلمات به‌بلاک‌های کاراکتری تقسیم و پس از حذف در سطح‌کاراکتر، بلاک‌های نهایی در سطح‌کلمه تشکیل می‌شوند. معمولاً متون در تصویر به‌صورت کاراکتر تنها ظاهر نمی‌شوند و دارای نویز و اشیاء پراکنده هستند. در نظر گرفتن "مساحت قابل‌قبول" برای کاراکتر، معیاری جهت تایید کاراکتر و حذف نویزها و اشیاء پراکنده است. مساحت‌های بلوک-های مجرد غیرمتن برای حذف در مجموعه ICDAR عبارت از (۱) مساحت‌کمتر از ۲۵۰۰؛ (۲) مساحت کمتر از ۶۰۰۰ که طول آن بیشتر از عرض‌اش باشد است. شرط اصلی وجود بلاک مجرد عدم وجود بلاک دیگر در اطراف آن است.

حذف درسطح‌خط‌متن: در سطح خط متن نیز همان ویژگی‌های سطح کلمه بکار می‌رود و با این تفاوت که لبه‌ها و مساحت‌ها در سطح خط متن در نظرگرفته می‌شوند. همچنین جهت‌حذف تشخیص‌های اشتباه از ویژگی خط متن مبتنی بر مفهوم اسکلت دارای راستا استفاده می‌شود. در ابتدا اسکلت هر مولفه، استخراج و با تشخیص نقاط انتهایی، بلاک‌های غیرمتن با معیار رابطه(۶) یعنی نسبت طول اسکلت به‌فاصله دو نقطه ابتدا-انتها حذف می‌شوند. برای متن، طول اسکلت مقداری نزدیک به فاصلهٔ خط‌مستقیم ولی برای غیرمتن بزرگتر از خط‌مستقیم است.

$$StraightRate = \frac{\sum EskeltonLength(S_i)}{TwoEndPointsLength} \qquad (۶)$$

$$StraightRate = \begin{cases} > 1 & TextLine \\ \leq 1 & Non\ TextLine \end{cases}$$

۴-۶ سنجش اشیاء

پس از این مرحله باید اشیاء یافت‌شده را در سطوح مختلف خط، کلمه و کاراکتر برای تجزیه شی خط به اشیاء کلمه و شیء کلمه را به اشیاء کاراکتری کلاس‌بندی کرد و برچسب زد. با این کار می‌توان گرافی از اشیاء مختلف متون و محل دقیق آن‌ها تعیین کرد که کاربرد بسیاری برای پردازشگرهای متن جهت پردازش‌هایی مانند صفحه‌بندی یا حذف فضاهای‌خالی تصاویر خواهد داشت. استفاده از ساختار بات‌های پردازشگر و ساختمان داده‌های مناسب تشکیل این گراف و حجم آن از چالش‌های داده‌های عظیم تصویری خواهد بود.

۴-۷ به‌سازی اشیاء

معمولاً متون به تنهایی در تصویر ظاهر نمی‌شوند و یا مشکلاتی در تصویر وجود دارد که برای خوانایی بیشتر، اشیاء را نیازمند به‌سازی می‌کند. این امر به تشکیل *بلاک‌های نهایی* پس از حذف نواحی غیر ضروری در سطح خط متن و کلمه به‌صورت جداگانه، کمک می‌کند تا نتایج آن‌ها با هم ادغام و بلاک‌های قابل اعتماد برای خواندن در سطح خط متن و کلمه ایجاد شوند.

۴-۸ تحلیل‌گر استراتـژیک تصویر

در انتهای ارائه اشیاء مختلف یک داده عظیم تصویری، تحلیل‌گر استراتژیک تصاویر با ارسال بات‌های خود طبق گراف وظیفه۹۹ مربوطه به‌صورت موازی وپی-درپی عملیات موردنیاز پردازش و مبادله داده‌های اصلی یا میانی را انجام می‌دهد. این تحلیل‌گر باید بتواند ضمن ارائه مشخصات واقعی داده‌های متنی داخل تصویر نسبت به ارائه متا داده‌های هر تصویر عظیم اقدام کند. همچنین تحلیل‌گر از نظر ابعاد مختلف داده عظیم تصویری باید مشخصات مناسب تصاویر را استخراج کند تا داده-های میانی در پردازش‌های بعدی به انجام محاسبات لازم و تخصیص منابع مناسب جهت زمان‌بندی عملیات با استراتژی‌های مختلف کمک کند. این عمل در بعضی فعالیت‌ها مانند عملیات پردازشی تصاویر پزشکی لازم و حیاتی است و استفاده مجدد آن‌ها را تسهیل می‌کند و پردازش آن‌ها را بهبود می‌دهد. ایجاد متا داده برای تصاویر حجیم و عظیم امری اجتناب‌ناپذیر است. قطعا این پردازشگر یک سیستم یادگیر است که امکان عملیات موازی در طی فرآیند شیءیابی را انجام می‌دهد. این تحلیل‌گر می‌تواند نسبت به انتخاب استراتژی‌های مناسب در شرایط مختلف تصویربرداری و حالت‌های مهم محیطی بر اساس علایق عامل‌ها و مطابق هر کاربرد برای تصمیم-سازی اقدام کند. این استراتژی‌ها می‌تواند مطلوب یا بیشینه باشد و بر اساس علایق عامل‌های هوشمند یا بات‌ها، مطابق گراف وظیفه تخصیص معماری سخت‌افزار (چندهسته‌ای، چند پردازندگی، جی‌پی‌یو، ابری و غباری مجازی) معیارهای متفاوتی برای ارزیابی را انتخاب کند. به بیان دیگر، هر کاربرد۱۰۰ باید دارای جریان کار خود باشد و بتواند از جریان بات‌های و عملیات آن‌ها بهره ببرد.

۵- مجموعه‌داده‌ها و معیارها

فاکتورهای موقعیت‌یابی متن عبارت از سبک فونت، اندازه (عرض و ارتفاع)، پهنای قلم، زمینه تصویر متن، وضعیت دوربین، جهت متن، زبان متن، کنتراست متن، رزولوشن تصویر، شدت روشنایی، هم‌ترازی، رنگ متن و زمینه، متن و غیرمتن، نویز و بلور است. لذا برای آزمون این موارد در مقیاس‌ها و زبان‌های مختلف نیازمند داده‌های مناسب هستیم. مجموعه داده‌های موجود موقعیت‌یابی در تصاویر به چند دسته برای استفاده در تصاویر ثابت و ویدئویی تقسیم می‌شوند. داده‌های متداول در مقیاس داده‌های معمول قرار دارند ولی داده‌های عظیم تصویری هنوز دارای تعریف خاص نیستند و وجود خارجی ندارند. بسط آن‌ها در سطوح بلاک، خط متن، کلمه و حرف به پیچیدگی مسئله می‌افزاید و در کنار معیارهای ارزیابی مطرح می‌شوند. در این بخش به بررسی داده‌های متداول و پیشنهاد داده‌های عظیم می‌پردازیم.

۵-۱ مجموعه داده‌ها

مجموعه داده‌های موجود را در دودسته داده‌های استاندارد و دسته داده‌های خاص برای تصاویر ثابت و ویدئویی تقسیم می‌کنیم و به بیان ساختار، ویژگی‌ها و کاربردشان می‌پردازیم [۸۶].

داده‌های استاندارد: از مجموعه‌های استاندارد ICDAR2003 [۲۲]، ICDAR2005 [۲۳] ، ICDAR2011 [۱۰۷]، OSTD [۱۰۷] ، MSRA-TD500 [۱۰۷]، SVT [۱۰۷] ، NEOCR [۱۰۸]، KAIST [۱۰۸]، Chars74K [۱۰۹]، SVHN [۱۱۰] و IIIT 5K-Word [۱۰۷] برای آزمون استفاده می‌شود. در تصاویر ثابت، ۵۰۹ تصویر از مجموعه استاندارد ICDAR2003 شامل۲۵۱ تصویر متن‌صحنه و ICDAR2005 شامل ۲۵۸ تصویر مختلف در ابعاد ۹۳×۳۰۷ تا ۹۶۰×۱۲۸۰ وجود دارد. تصاویر از مناظر طبیعی، مارک‌ها و برچسب‌های تجاری، اعداد و علایم جاده‌ای هستند. این مجموعه دارای حقیقت‌پایه است که محدودهٔ دقیق کلمات موجود در هر تصویر به‌صورت جداگانه و برای تمامی‌تصاویر مشخص و برای ارزیابی نتایج به‌کار می‌رود. مجموعه داده ICDAR 2017 شامل ۱۲۵۵ تصویر آموزشی و ۳۰۰ تصویر آزمون دارای عمده چالشهای مطرح به جز بحث داده‌های عظیم تصویری است.

مجموعه دادهٔ OSTD (مجموعه داده‌های متون مناظر با جهات مختلف)[۵۲] شامل ۸۹ تصویر از آرم‌ها، مناظر داخل خانه و نمای خیابان است که می‌تواند برای ارزیابی الگوریتم‌های آشکارسازی متن در جهات مختلف در تصاویر طبیعی مورد استفاده قرار بگیرد. مجموعه داده MSRA-TD500 [۱۳]، به‌منظور ارزیابی الگوریتم-های آشکارسازی متن برای متونی که در تصاویر مناظر در جهات مختلف قرار دارند ایجاد شده است. این مجموعه داده شامل ۵۰۰ تصویر با متون افقی و همچنین مورب در مناظر پیچیدهٔ طبیعی می‌باشد. مجموعه دادهٔ SVT (متون مناظر خیابانی) [۵۶]، مجموعه‌ای از تصاویر در فضای باز با متونی با تنوع بالا می‌باشد.SVT شامل ۳۵۰ تصویر که می‌تواند هم برای تشخیص کلمه و هم برای آشکارسازی و تشخیص متن به کار رود. مجموعه داده‌ای NEOCR [۱۰۲] شامل عکس‌هایی از مناظر طبیعی با متون در جهت‌های مختلف می‌باشد این مجموعهٔ داده شامل ۶۵۹ تصویر از جهان واقعی با ۵۲۳۸ قاب متن می‌باشد؛ که دارای تصاویری با متون‌هایی با زبان‌های مختلف از جمله انگلیسی، مجارستانی، روسی، ترکی و جمهوری چک می‌باشد.

مجموعه دادهٔ متون مناظر KAIST [۱۰۳] شامل ۳۰۰۰ تصویر است که از مناظر درون خانه و بیرون، تحت شرایط روشنایی مختلف گرفته شده است. این مجموعه داده نیز همچنین دارای معیارهای چند زبانه که شامل متون کره‌ای و انگلیسی است علاوه بر این، این مجموعه داده برای هر کاراکتر در تصویر، یک ماسک باینری فراهم می‌کند بنابراین این مجموعهٔ داده را می‌توان در هر دو قسمت موقعیت‌یابی و ناحیه-بندی متن مورد استفاده قرار بگیرد. پایگاه دادهٔ Chars74k [۱۰۴]، برای ارزیابی الگوریتم‌های تشخیص متن برای ناحیه‌بندی کاراکترها در تصاویر طبیعی منتشر شد. این مجموعهٔ داده شامل ۶۳۶ تصویر است که در آن بخشی از حروف لاتین و اعداد عربی نیز وجود دارد. مجموعهٔ دادهٔ SVHN [۱۰۵] یک مجموعه دادهٔ بزرگ از جهان واقعی است که شامل بیش از ۶۰۰ هزار عدد در تصاویر مناظر طبیعی می‌باشد. این مجموعهٔ داده برای توسعه و ارزیابی الگوریتم‌های تشخیص رقم مورد استفاده قرار می‌گیرد.

مجموعه داده‌های IIIT5K word [۱۰۶] از بزرگ‌ترین و چالش برانگیزترین محک‌ها در زمینهٔ آشکارسازی و تشخیص متن تا به امروز می‌باشد این پایگاه داده شامل ۵ هزار تصویر با متن در تصاویر مناظر طبیعی و تصاویر تولیدات دیجیتالی می‌باشد. این پایگاه داده از آن جهت چالش برانگیز است که دارای تنوع در فونت، رنگ، اندازه، نویز و تار شدن می‌باشد. در این مجموعه ۲ هزار عکس برای آموزش و ۳ هزار عکس برای آزمون استفاده شده است. مجموعه داده COCO Text [۸۶] در حال حاضر بزرگ‌ترین و کامل‌ترین مجموعه‌های داده مشتمل بر۶۳۶۸۶ تصویر برای بخش بندی و تشخیص متن می‌باشد، تصاویر موجود انواع پیچیدگی‌های مطرح در حوزه تشخیص متن از جمله تنوع در فونت، رنگ، اندازه، نویز و غیره را شامل می-شوند.

داده‌های خـاص: برای ارزیابی بهتر روش‌ها و اعتماد به سیستم در تصاویر ویدئویی از داده‌های خاص زبان‌های مختلف (مانند فارسی و عربی)، نگارش‌ها متفاوت (تایپی و دست‌نویس)، تحت ابعاد (رزولوشن‌ها و کنتراست) و شرایط مختلف محیطی (کنترل‌شده و کنترل نشده) باید در ساختارهای تصاویر ثابت و ویدئویی استفاده شود. این داده‌ها در هر سیستم مکان‌یابی متن باید به‌تناسب مسئله و زبان پردازشی جمع‌آوری و آزمون شود تا سیستم قابلیت عمومی بودن برای پردازش تصاویر طبیعی

را نشان دهد. همچنین باید برای مکان‌یابی متون در کاربردهای خاص (مانند بینایی ماشین در ربات) ضوابط و استانداردهای مدون داده‌های عظیم تصویری طراحی شود.

کلان داده‌ها: داده‌های متداول هنوز تعریفی برای داده‌های عظیم تصویری استاندارد ندارد. باید این داده‌ها با توجه به ابعاد گوناگون آن برای آزمون تعریف شوند. در تصاویر پزشکی از بعد حجم داده‌های عظیم موجودند ولی برای مسئله‌های دیگر مانند مکان‌یابی تصویر در ابعاد مختلف باید به‌صورت دستی تهیه گردد. باید داده‌های استاندارد مختلف از نظر ابعاد حجم، تنوع، ارزش و پیچیدگی برای داده‌های عظیم تصویری ارائه شود.

متاداده‌ها: برای دنیای آینده نیازمند تعریف و تولید متا داده‌های مهم برای تصمیم‌گیری در سطوح مختلف مختلف پردازش هستیم. این داده‌ها پس از تولید باید برای کاربردهای خاص از ساختمان داده‌های مفید ذخیره دانش و ساختار متا داده‌ها برخوردار باشد. از ابزارهای موردنیاز و طراحی مفهومی مناسب برای این داده-های میانی مثلاً در قالب ساختارهای داده‌ای توزیعی بلاک چین باید استفاده شود.

۵-۲ معیارهای ارزیابی

معیارهای ارزیابی موقعیت‌یابی در تصاویر به دو دسته معیارهای ارزیابی متداول در مقیاس داده‌های معمول و معیارهای کلان داده‌ها در سطوح پاراگراف، خط متن، کلمه و حرف تقسیم می‌شوند که در زیر به شرح آن می‌پردازیم.

۵-۲-۱ معیارهای متداول

در این بخش به معیارهای متداول برای ارزیابی روش‌های سنتی متداول می-پردازیم و ضمن تعریف آن‌های در سطوح مختلف ارزیابی، کاربرد و استفاده درست آن را بیان می‌کنیم هرچند با توجه به کاربردی که این مسئله در حوزه‌های مختلف دارد برخی معیارها مهم‌تر و اساسی‌تر می‌شود و کارایی مسئله بر اساس همان معیارها (و نه همه معیارها) سنجیده می‌شود.

معیارهای خط متن: این معیارهای مرسوم عملکرد بلاک کاندیدای متن را ارزیابی می‌کنند [۸-۳۰، ۲۱، ۲۶]. پارامترهای اصلی ارزیابی بلاک‌های موقعیت‌یابی شده از قبیل تشخیص بلاک صحیح، بلاک اشتباه و بلاک ناقص عبارتند از:

- TDB: تشخیص صحیح بلاک[۱۰۱]، بلاک با حداقل یک کاراکتر که ممکن است شامل هیچ متنی نباشد.
- FDB: تشخیص غلط بلاک[۱۰۲]، این بلاک شامل هیچ متنی نیست.
- MDB: بلاک ناقص[۱۰۳]، بلاکی که بیش از ۲۰٪ کاراکترهای متن را تشکیل نمی-دهد. بلاکی که حداقل ۸۰٪ متن را تشکیل دهد مورد تأیید است [۳۰].

برای هر تصویر به‌صورت دقیق پیکسل‌های بلاک‌های متن در فایل مجموعه داده معلوم است. معیارهای ارزیابی فراخوانی رابطه (۷)، دقت رابطه (۸)، معیار اف رابطه (۹) و نرخ عدم تشخیص رابطه (۱۰) است.

$$\text{Recall} = R = \text{TDB}/\text{ATB} \quad (۷)$$

$$\text{Precision} = P = \text{TDB}/(\text{TDB} + \text{FDB}) \quad (۸)$$

$$\text{F}_{\text{measure}} = F = 2\text{PR}/(\text{P} + \text{R}) \quad (۹)$$

$$\text{Miss Detection Rate} = MDR = \text{MDB}/\text{TDB} \quad (۱۰)$$

معیارهای کلمه و کاراکتر: مجموعه‌ای از مستطیل‌هایی که توسط هر الگوریتم تخمین زده می‌شود با مستطیل‌های حقیقت‌پایه موجود در مجموعه ICDAR مقایسه می‌شوند. بهترین میزان تطبیق m بین این دو مستطیل (مساحت تقاطع دو مستطیل برابر با مینیمم مستطیل شامل هر دو) بیشترین ارزش است. ازاین‌رو بهترین تطبیق برای مستطیل r در هر گروه از مستطیل‌های R از رابطه (۱۱) تعیین می‌شود. دقت از رابطه (۱۲) و فراخوانی از رابطه (۱۳) در سطح کلمه به دست می‌آید که در آن‌ها E و T به ترتیب مستطیل‌های حقیقت‌پایه و مستطیل‌های تخمین‌زده‌شده هستند. f رابطه(۱۴)، ترکیبی از دو معیار دقت و فراخوانی با وزن‌نسبی پارامتر α (معمولاً $\alpha = 0.5$) تعیین می‌شود. همچنین معیار میانگین‌زمان‌پردازش[۱۰۴] (APT) رابطه(۱۶) برای مقایسه‌سرعت روش‌ها بازای k بار اجرای T_i در تصاویر با پیچیدگی‌های مختلف به‌کار می‌رود.

$$m(r, R) = \max\{m(r, \acute{r}) | \acute{r} \in R\} \quad (۱۱)$$

$$Precision = \frac{\sum_{r_e \in E} m(r_e; T)}{|E|} \quad (۱۲)$$

$$Recall = \frac{\sum_{r_t \in T} m(r_t; E)}{|T|} \quad (۱۳)$$

$$f = \frac{1}{\frac{a}{precision} + \frac{1-a}{recall}} \quad (۱۴)$$

$$\text{ATP} = \frac{\sum_{i=1}^{k} \text{T}_i}{\text{k}} \quad (۱۵)$$

۵-۲-۲ معیارهای داده عظیم

داده‌های کلان نیازمند معیارهای ارزیابی خود هستند و هنوز ابعاد مسئله آن به طور کامل مشخص نیست. بیان معیارهای موردنیاز و ارائه‌شده موجود برای سیستم موقعیت‌یابی می‌تواند به طرح مسئله کمک کند. با توجه به گسترش کلان‌داده تصویری نیاز به معیارهای نوین برای ابعاد حجم، پیچیدگی، تنوع و ارزش لازم است. در بعد حجم برای ذخیره‌سازی در حافظه اصلی و میزان پردازش‌های سریال و موازی معیارهای خاص نیازمندیم و روش آن نیز باید توزیعی و مقیاس‌پذیر باشد. در بعد تنوع معیارهای متداول می‌تواند ملاک ارزیابی قرار گیرد ولی با توجه به ابعاد مختلف آن نیازمند تجدید نظر است. زیرا روش‌های فعلی در داده‌های کلان فاقد رفتار و کارایی مناسب هستند. در بعد ارزش با توجه ارائه معیارهای جدید مانند زمان (-ذخیره و بازیابی داده‌ها و انجام پردازش فرآیندها و مبادلات داده‌ای)، هزینه (پردازش و توان مصرفی)، منابع (موردنیاز مانند حافظه) و نوع ارزش‌گذاری پردازش‌ها نیازمند ملاک‌ها و معیارهای نوین در هر بعد هستیم. در بعد پیچیدگی (وجود نویز، رزولوشن زیاد، زمینه شلوغ و متون درهم) نیازمند معیارهای خاص تشخیص میزان پیچیدگی تصویر است تا به تناسب نوع و روش و میزان منابع موردنیاز ارزیابی شود. در ابعاد مختلف نیازمند انجام سریع ارزیابی، تشخیص و موقعیت‌یابی متون با کمترین هزینه، منابع مصرفی و حداقل داده‌های موردنیاز هستیم. به دلیل ابعاد بزرگ مسئله تعریف معیارهای خطا در سطوح مختلف ماکرو و میکرو و نسبت آن با ریسک تصمیم‌سازی احساس می‌شود.

معیارهای خوشه‌بندی: این معیارها مانند شاخص خلوص رابطه (۱۶)، شاخص جاکارد رابطه (۱۷)، شاخص دون رابطه (۱۸)، شاخص سیلوئت رابطه (۱۹) بیان‌شده که در آینده باید بازنگری شود[۶۳].

$$P = \frac{1}{N} \sum_{m \in M} Max\, |m \cap d| \quad (۱۶)$$

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN} \quad (۱۷)$$

$$V_D = \left[\frac{\min\limits_{i=1 \leqslant j \leqslant k} D(C_i, C_j)}{\max\limits_{1 \leqslant l \leqslant k} diam(C_l)} \right] \quad (۱۸)$$

$$s(i) = \frac{b(i) - a(i)}{max(b(i), a(i))} \quad (۱۹)$$

معیارهای پردازش موازی: این معیار در سطح عملیات موازی و بر اساس ساختار پردازشی مانند جی‌پی‌یو بیان می‌شود که برای یافتن جواب مطلوب هر تابع هدف خوب محسوب گردد.

معیارهای بی‌درنگی: این معیار در عملیات بی‌درنگ بر اساس ضرب‌الاجل هر تسک یا عملیات تعریف و باید برآورده شود تا بتوان هر عمل را در زمان خود انجام داده و نتایج در زمان غیرواقعی به دست نیاید.

معیارهای توان مصرفی: این معیار در عملیاتی که هدف کمینه‌سازی مصرف توان است مانند رابطه (۲۰) بکار می‌رود [۶۴].

$$Score = \frac{mAP}{Total\ Energy\ Consumtion} \quad (۲۰)$$

معیارهای ارزیابی تهدیدات: باید اهداف پردازش در هر کاربرد تعریف شود و بر اساس آن خطاهای موردنیاز از تصویربرداری تا خطاهای انسانی که در عملیات وقفه ایجاد می‌کند مانند رابطه (۲۱) تعیین گردد. این امر به ردیابی و کنترل تهدیدات ممکن در عملیات کمک می‌کند تا به کاهش ریسک عملیات در کاربردهای بی‌درنگ مانند خطر ابعاد عملیات در جراحی‌های پزشکی منتهی شود. این امر نگاه سلسله مراتبی عملیات تا تغییر استراتژی و بالعکس را می‌طلبد. باید با مشخص‌کردن مدل مناسب برای تهدیدات به معیارهای مناسب ارزیابی آن‌ها توجه شود.

$$thr(B) = min\left(0.5, \frac{wh}{(w+10)(h+10)}\right) \quad (۲۱)$$

معیارهای تصمیم‌سازی: در عملیات با وضعیت عدم اطمینان باید به میزان ریسک‌گرایی و ریسک‌گریزی در تصمیم‌سازی‌ها مخصوصاً حالت آنلاین دریافت و پردازش تصاویر توجه شود. شاید در هر پریود پردازش، تصمیم‌گیری ریسک‌گریز امکان نداشته باشد و یا باید بده و بستانی در سیستم صورت بگیرد تا بعضی منافع فدای آن شود. می‌توان با استفاده از تئوری سودمندی[۱۰۵]، مدل‌سازی و تصمیم‌سازی در حالت ریسک‌گریزی و مدل‌های دیگر در ریسک‌گرایی به انجام مفید عملیات کمک کرد. برای این وضعیت باید ضمن توجه به معیارهای مناسب بتوان توابع مفید تولید کرد. از آنجا که همیشه در حالت مطمئن قرار نداریم تعریف این معیارها در تمامی سطوح شکل۱ برای تصمیم‌سازی در وضعیت‌های ریسک‌گریزی و ریسک‌گرایی لازم است.

۶- نتیجه‌گیری

مشکلات ارائه روش عمومی مؤثر جهت موقعیت‌یابی و تشخیص متون در تصاویر، وابستگی بسیاری به ابعاد متون، جهت آنها، رنگ متون، میزان رنگ‌آمیزی، پیچیدگی زمینه، نورپردازی محیطی، میزان تقعر و تحدب زمینه متن، رزولوشن تصویر، پیچیدگی زبان متن، و تنوع زبان برای برخورداری از سرعت و دقت مناسب دارد. در کنار این مسئله *تنوع*، حجم، ارزش و پیچیدگی تصویر یا فریم ویدئویی به‌عنوان چالش مهم در آینده به نمایش در خواهد آمد. روش‌های سنتی موقعیت‌یابی متن نسبت به مشکلات آینده یعنی دنیای داده‌های عظیم تصویری مقاوم نخواهند بود و بی‌درنگ عمل نخواهند کرد. آن‌ها فاقد انتخاب استراتژی‌های مناسب بر اساس علایق عامل‌های هوشمند برای استفاده از معیارهای کلان داده‌ها هستند. از مشکلات مهم این روش‌ها فقدان استراتژی مناسب تشخیص هم‌زمان در زوایای مختلف در تصاویر و عدم مقیاس‌پذیری در راستای داده‌های تصویری عظیم است. پشتیبانی از تنوع داده‌ها در ابعاد مختلف برای رفع بسیاری از چالش‌های تصویر است و برای کاربرد داده‌های عظیم تصویری نیازمند ساختمان داده‌های توزیعی موازی مانند بلاک چین، تولید متاداده‌ها برای حفظ اطلاعات سراسری، الگوریتم‌های موازی و توزیعی مقیاس‌پذیر، استفاده از داده‌های محلی و سراسری در تصمیم‌گیری، استفاده از تحلیل‌گر استراتژیک برخط، طراحی و پیاده‌سازی ابزارهای عامل‌گرا و بات‌گرا، و معیارهای جدید برای اندازه‌گیری داده‌های استخراجی و بررسی و تحلیل الگوریتم هستیم. در کارهای آینده باید ضمن افزایش سرعت، روش‌های بی‌درنگ برای داده‌های بزرگ در کاربردهای خاص مانند متون پیچیده در جهت‌های مختلف ارائه شود. می‌توان از مدل ماده سیاه[۱۰۶] برای کشف سریع نواحی غیرمتنی تصاویر و تئوری مجموعه‌های خشن برای بخش‌بندی سریع و تعیین نواحی متنی استفاده کرد. مسئله مورد توجه دیگر ارائه معیارها و توابع موردنیاز موقعیت‌یابی در وضعیت اطمینان یا عدم اطمینان و تصمیم‌گیری ریسک‌گرایی یا ریسک‌گریزی در حالت عدم اطمینان است.

۷- مراجع

[1] Z. Yingying, C. Yao, and X. Bai, "Scene text detection and recognition: Recent advances and future trends," *Frontiers of Computer Science*, vol.10, no.1, pp. 19-36, 2016.

[2] L. Rainer, and A. Wernicke, "Localizing and segmenting text in images and videos," *Circuits and Systems for Video Technology*, vol.12, no.4, pp.256-268, 2002.

[3] W. Edward, and M. Chen, "A new robust algorithm for video text extraction," *Pattern Recognition,* vol.36, no.6, pp.1397-1406, 2003.

[4] C. Min, J. Song, and M. R. Lyu, "A new approach for video text detection," *IEEE International Conference on Image Processing,* vol.1, pp.1-17, 2002.

[5] J A. Jamil, I. Siddiqi, F. Arif, and A. Raza, "Edge-based features for localization of artificial Urdu text in video images," *IEEE International Conference* on *Document Analysis and Recognition (ICDAR),* pp.1120-1124, 2011.

[6] A. Marios, B. Gatos, I. Pratikakis, "A two-stage scheme for text detection in video images," *Image and Vision Computing*, vol.28, no.9, pp.1413-1426, 2010.

[7] P. Xujun, H. Cao, R. Prasad, and P. Natarajan, "Text extraction from video using conditional random fields," *IEEE International Conference on Document Analysis and Recognition (ICDAR),* pp.1029-1033, 2011.

[8] S. Palaiahnakote, T. Phan, S. Lu, and C. Lim Tan, "Gradient vector flow and grouping-based method for arbitrarily oriented scene text detection in video images,"*Circuits and Systems for Video Technology,* vol.23, no.10, pp.1729-1739, 2013.

[9] P. Yi-Feng, X. Hou, and C. Liu, "A hybrid approach to detect and localize texts in natural scene images," *Image Processing,* vol.20, no.3, pp.800-813, 2011.

[10] A. K. Jain, and B. Yu. "Automatic text location in images and video frames," *Pattern recognition*, vol.31, no.12, pp.2055-2076, 1998.

[11] E. Boris, E. Ofek, Y. Wexler, "Detecting text in natural scenes with stroke width transform," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp.2963-2970, 2010.

[12] N. Lukas, and J. Matas, "A method for text localization and recognition in real-world images," *Asian Conference on Computer Vision*, pp.770-783, 2011.

[13] Y. Cong, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp.1083-1090, 2012.

[14] H. Weilin, Z. Lin, J. Yang, and J. Wang, "Text localization in natural images using stroke feature transform and text covariance descriptors," *IEEE international conference on Computer Vision (ICCV),* pp. 1241-1248 ,2013.

[15] T. Novikova, O. Barinova, P. Kohli, and V. Lempitsky, "Large-lexicon attribute-consistent text recognition in natural images," *European conference on Computer Vision–ECCV*, pp.752-765, 2012.

[16] Y. Cong, X. Bai, and W. Liu, "A unified framework for multioriented text detection and recognition," *Image Processing,* vol.23, no.11, pp.4737-4749, 2014.

[17] Y. Xu-Cheng, X. Yin, K. Huang, and H. Hao, "Robust text detection in natural scene images," *Pattern Analysis and Machine Intelligence*, vol.36, no.5, pp.970-983, 2014.

[18] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. W. Ma, ''Robust face recognition via sparse representation,'' *Pattern Analysis and Machine Intelligence*, vol.31, no. 2, pp.210-227, 2008.

[19] E. Michael, and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing*, vol.15, no.12, pp.3736-3745, 2006.

[20] Z. Ming, S. Li, and J. Kwok, "Text detection in images using sparse representation with discriminative dictionaries," *Image and Vision Computing*, vol.28, no.12, pp.1590-1599, 2010.
[21] S. Palaiahnakote, T. QuyPhan, and C. L. Tan, "A laplacian approach to multi-oriented text detection in video," *Pattern Analysis and Machine Intelligence*, vol. 33, no.2, pp.412-419, 2011.
[22] L. Simon, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, "ICDAR 2003 robust reading competitions," *International Conference on Document Analysis and Recognition (ICDAR),* pp.682-678, 2003.
[23] L. Simon, "ICDAR 2005 text locating competition results, " *International Conference on Document Analysis and Recognition (ICDAR)*, pp.80-84, 2005.
[24] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," *IEEE International Conference on Image Processing (ICIP)*, pp.2609-2612, 2011.
[25] L. Su, and K. E. Barner, "Weighted DCT coefficient based text detection," *International Conference on Acoustics, Speech and Signal Processing ICASSP*, pp.1341-1344, 2008.
[26] W. Edward, and M. Chen, "A new robust algorithm for video text extraction," *Pattern Recognition*, vol.36, no.6, pp. 1397-1406, 2003.
[27] C. Min, J. Song, and M. R. Lyu, "A new approach for video text detection," International Conference on Image Processing Proceedings, vol.1, pp.110-117, 2002.
[28] Y. Qixiang, Q. Huang, W. Gao, and D. Zhao, "Fast and robust text detection in images and video frames," *Image and Vision Computing*, vol.23, no.6, pp.565-576, 2005.
[29] L. C. Woo, K. Jung, and H. J. Kim, "Automatic text detection and removal in video sequences," *Pattern Recognition Letters*, vol.24, no.15, pp.2607-2623, 2003.
[30] C. Datong, J. Odobez, and J. Thiran, "A localization/verification scheme for finding text in images and video frames based on contrast independent features and machine learning methods," *Signal Processing: Image Communication*, vol.19, no.3, pp.205-217, 2004.
[31] W. Tao, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," *International Conference on Pattern Recognition (ICPR)*, pp.3304-3308, 2012.
[32] J. Max, K. Simonyan, A. Vedaldi, A. Zisserman, "Synthetic data and artificial neural networks for natural scene text recognition," *arXiv preprint arXiv:1406.2227* (2014).
[33] S. Bolan, and S. Lu, "Accurate scene text recognition based on recurrent neural network," *Computer Vision-ACCV 2014, Springer International Publishing*, pp.35-48, 2015.
[34] J. Max, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading text in the wild with convolutional neural networks," *International Journal of Computer Vision*, vol.116, no.1, pp.1-20, 2014.
[35] J. Max, K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep structured output learning for unconstrained text recognition," *arXiv preprint arXiv:1412.5903*, 2014.
[36] J. Munho, and K. Jo, "Multi language text detection using fast stroke width transform," *Korea-Japan Joint Workshop onFrontiers of Computer Vision (FCV),* pp.1-4, 2015.
[37] T. Kobchaisawat, and H. C. Thanarat, "A method for multi-oriented Thai text localization in natural scene images using Convolutional Neural Network," *International Conference on Signal and Image Processing Applications (ICSIPA),* pp.220-225, 2015.
[38] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, A. Bagdanov, M. Iwamura, J. Matas, L. Neumann, V. R. Chandrasekhar, S. Lu, and F. Shafait, "Icdar 2015 competition on robust reading," *International Conference on Document Analysis and Recognition (ICDAR),* pp.1156–1160, 2015.
[39] Y. Zheng, Q. Li, J. Liu, H. Liu, G. Li, and S. Zhang, "A cascaded method for text detection in natural scene images," *Neurocomputing*, vol.238, pp.307-315, 2017.
[40] N. Vasilopoulos, and E. Kavallieratou, "Unified layout analysis and text localization framework, " *Electronic Imaging*, vol. 26, no.1, 2017.
[41] A. B. Ayed, M. B. Halima, and A. M. Alimi, "MapReduce Based Text Detection in Big Data Natural Scene Videos," *INNS Conference on Big Data,* vol.53, pp.216-223, 2015.
[42] S. Ali S, K. Iqbal, S. Khan, Q. Z. Aqil, and R. Tariq, "A Review on Text Detection Techniques," *VFAST Transactions on Software Engineering*, vo.78, pp.4-3, 2016.
[43] P. Shivakumara, R. P. Sreedhar, T. Q. Phan, S. Lu, and C. L. Tan, "Multioriented video scene text detection through bayesian classification and boundary growing," *Circuits and Systems for Video Technology*, vol.22, no.8, pp.1227-1235, 2012.
[44] X. Wang, Y. Jiang, Z. Luo, C. L. Liu, H. Choi, and S. Kim, "Arbitrary shape scene text detection with adaptive text region representation," *IEEE Conference on Computer Vision and Pattern Recognition*, pp.6449-6458, 2019.
[45] J. Zhou, L. Xu, B. Xiao, and R. Dai, "A robust system for text extraction in video," *IEEE International Conference on Machine Vision*, pp.119-124, 2007.
[46] S. Palaiahnakote, T. Q. Phan, and C. L. Tan, "New Fourier-statistical features in RGB space for video text detection*," Circuits and Systems for Video Technology*, vol. 20, no.11, pp.1520-1532, 2010.
[47] L. Chunmei, C. Wang, and R. Dai, "Text detection in images based on unsupervised classification of edge-based features," *International Conference on Document Analysis and Recognition*, pp.610-614, 2005.
[48] W. Huang, Z. Lin, J. Yang, and J. Wang, "Text localization in natural images using stroke feature transform and text covariance descriptors," *International Conference on Computer Vision (ICCV),* pp.1241-1248, 2013.
[49] W. Huang, Q. Yu, and X. Tang, "Robust scene text detection with convolution neural network induced mser trees*," Computer Vision–ECCV*, pp.497-511, 2014.
[50] Z. Yu, K. Karu, and A. K. Jain, "Locating text in complex color images," *International Conference on Document Analysis and Recognition*, vol.1, pp.146-149, 1995.
[51] L. Huiping, D. Doermann, and O. Kia, "Automatic text detection and tracking in digital video," *Image Processing*, vol.9, no.1, pp,147-156, 2000.
[52] Y. Chucai, and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping." *Image Processing*, vol.20, no.9, pp.2594-2605, 2011.
[53] K. Kwang, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *Pattern Analysis and Machine Intelligence*, vol.25, no.12, pp.1631-1639, 2003.
[54] L. Michael, J. Song, and M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," *Circuits and Systems for Video Technology*, vol.15, no.2, pp.243-255, 2005.
[55] Y. Liu, and T. Ikenaga, "A contour-based robust algorithm for text detection in color images," *IEICE transactions on information and systems*, vol.89, no.3, pp.1221-1230, 2006.
[56] W. Kai, and S. Belongie, "Word spotting in the wild," *European Conference on Computer Vision, Springer Berlin Heidelberg*, 2010.
[57] C. Xiangrong, and A. L. Yuille, "Detecting and reading text in natural scenes," *IEEE Conference on Computer Vision and Pattern Recognition*, vol.2, pp.II-II, 2004.
[58] W. Christian, and J. M. Jolion, "Extraction and recognition of artificial text in multimedia documents," *Formal Pattern Analysis & Applications*, vol.6, no.4, pp.309-326, 2004.
[59] Y. Bae, and J. Park, "Architecture for fast object detection supporting CPU-GPU hybrid and distributed computing, " *IEEE International Conference Consumer Electronics (ICCE),* pp.158-159, 2017.
[60] J. Pont-Tuset, P. Arbelaez, J. T. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping for image segmentation and object proposal generation, " *Pattern Analysis and Machine Intelligence*, vo.39, no.1, pp.128-140, 2017.
[61] T. Kurc, X. Qi, D. Wang, F. Wang, G. Teodoro, L. Cooper, M. Nalisnik, L. Yang, J. Saltz, and D. J. Foran, "Scalable analysis of Big pathology image data cohorts using efficient methods and high-performance computing strategies, " *BMC bioinformatics*, vol.16. no.1, pp.1-21, 2015.
[62] Z. Chen, W. Zhang, B. Hu, X. Cao, S. Liu, and D. Meng, "Retrieving Objects by Partitioning," *IEEE Transactions on Big Data*, vol.3, no.1, pp.44-54, 2017.
[63] R. Kune, P. K. Konugurthi, A. Agarwal, R. R. Chillarige, and R. Buyya, "XHAMI–extended HDFS and MapReduce interface for Big Data image processing applications in cloud computing environments, " *Software: Practice and Experience*, vol.47, no.3, pp.455-472, 2017.
[64] K. Gauen, R. Rangan, A. Mohan, Y. H. Lu, W. Liu, and A. C. Berg, "Low-power image recognition challenge, " *IEEE Conference on Design Automation (ASP-DAC),* pp.99-104, 2017.
[65] I. Lee, "Big data: Dimensions, evolution, impacts, and challenges, " *Business Horizons*, vol. 60, no.3, pp.293-303, 2017.
[66] J. Liu, Y. Huang, J. Peng, J. Yao, and L. Wang, "Fast Object Detection at Constrained Energy, " *IEEE Transactions on Emerging Topics in Computing*, vol.6, no.3, pp.409-4016, 2016.

[67] R. Zhang, X. Liu, J. Hu, K. Chang, and K. Liu, "A fast method for moving object detection in video surveillance image, " *Signal, Image and Video Processing*, vol.11, no.5, pp.841-848, 2017.
[68] L. Dong, Z. Lin, Y. Liang, L. He, N. Zhang, Q. Chen, X. Cao, and E. Izquierdo, "A Hierarchical Distributed Processing Framework for Big Image Data," *IEEE Transactions on Big Data*, vol.2, no.4, pp.297-309, 2016.
[69] F. Ronald B., S. Gardner, and P. Palangpour, "Energy-efficient secure vision processing applying object detection algorithms," *U.S. Patent Application*, No. 15/227,949, 2017.
[70] G. Xiang, H. Yeh, and P. Marayong, "A high-speed color-based object detection algorithm for quayside crane operator assistance system," *Annual IEEE International Systems Conference (SysCon),* pp.1-6, 2017.
[71] D. Nguyen, L. Shijian, N. Ouarti, and M. Mokhtari, "Text-Edge-Box: An Object Proposal Approach for Scene Texts Localization," *IEEE Winter Conference on Applications of Computer Vision (WACV),* pp. 1296-1305, 2017.
[72] U. B. Karanje, and R. Dagade, "Survey on Text Detection, Segmentation and Recognition from a Natural Scene Images," *International Journal of Computer Applications*, vol.108, no.13, 2014.
[73] X. Shen, W. Liu, I. Tsang, F. Shen, and Q. S. Sun, "Compressed K-Means for Large-Scale Clustering, *" Thirty-first aaai Conference on Artificial Intelligence*, 2017.
[74] H. Kevin, and M. Golparvar-Fard, "Potential of big visual data and building information modeling for construction performance analytics: An exploratory study," *Automation in Construction*, vol. 73, pp.184-198, 2017.
[75] K. Amandeep, R. Dhir, and G. S. Lehal, "A survey on camera-captured scene text detection and extraction: towards Gurmukhi script," *International Journal of Multimedia Information Retrieval*, vol.6, no.2, pp.115-142, 2017.
[76] T. Mukesh, and R. Singhai, "A Review of Detection and Tracking of Object from Image and Video Sequences," *International Journal of Computational Intelligence Research*, vol.13, no.5, pp.745-765, 2017.
[77] L. Yang, S. Cheng, P. K. Verma, and S. Wang, "Text Search: Towards Fast Text Localization in Scene Images," *IEEE International Symposium on Multimedia (ISM)*, pp. 83-86, 2016.
[78] M. Liao, B. Shi, X. Bai, X. Wang, and W. Liu, "TextBoxes: A Fast Text Detector with a Single Deep Neural Network," *arXiv preprint arXiv:1611.06779*, 2017.
[79] G. Luís, and D. Karatzas. "Text proposals: a text-specific selective search algorithm for word spotting in the wild," *Pattern Recognition,* vol.70, pp.60-74, 2017.
[80] S. Qin, and M. Manduchi. "Cascaded Segmentation-Detection Networks for Word-Level Text Spotting," *arXiv preprint arXiv:1704.00834*, 2017.
[81] J. Zhang, G. Wu, X. Hu, and X. Wu, "A distributed cache for hadoop distributed file system in real-time cloud services*," International Conference on Grid Computing (GRID)*, pp.12-21, 2012.
[82] B. Kulis, and K. Grauman, "Kernelized locality-sensitive hashing for scalable image search*," international conference on computer vision (ICCV),* pp.2130-2137, 2017.
[83] Z. Zhang, D. S. Katz, J. M. Wozniak, A. Espinosa, and I. Foster, "Design and analysis of data management in scalable parallel scripting," *International Conference on High Performance Computing, Networking, Storage and Analysis*, pp. 1-12, 2012.
[84] M. Almeer, "Cloud Hadoop MapReduce for remote sensing image analysis," Emerging Trends in Computing and Information Sciences, vol.3, no.4, pp.637-644, 2012.
[85] L. Neumann, and J. Matas, "Real-time lexicon-free scene text localization and recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol.38, no.9, pp.1872-85, 2016.
[86] A. Veit, T. Matera, L. Neumann, J. Matas, and S. Belongie, "Coco-text: Dataset and benchmark for text detection and recognition in natural images," *arXiv preprint arXiv:1601.07140*, 2016.
[87] H. Turki, M. B. Halima, and A. M. Alimi, "Scene text detection images with pyramid image and MSER enhanced," *International Conference on Intelligent Systems Design and Applications (ISDA),* pp.301-306, 2015.
[88] S. Matko, "Scene Text Segmentation using Low Variation Extremal Regions and Sorting Based Character Grouping," *Neurocomputing*, vol.266, pp.56-65, 2017.
[89] C. Hojin, M. Sung, and B. Jun. "Canny text detector: Fast and robust scene text localization algorithm," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3566-3573, 2016.
[90] Y. Song, J. Chen, H. Xie, Z. Chen, X. Gao, and X. Chen, "Robust and parallel Uyghur text localization in complex background images," *Machine Vision and Applications*, vol.28, no.7, pp.755-69, 2017.
[91] Y. Chong, Y. Song, and Y. Zhan, "Scene text localization using edge analysis and feature pool," *Neurocomputing*, vol.175, pp. 652-661, 2016.
[92] C. Kai, F. Yin, and C. L. Liu. "Effective Candidate Component Extraction for Text Localization in Born-Digital Images by Combining Text Contours and Stroke Interior Regions," *IAPR Workshop on Document Analysis Systems (DAS),* pp. 352-357, 2016.
[93] Vidhya, K.A. and Geetha, T.V., "Rough set theory for document clustering: A review," *Journal of Intelligent & Fuzzy Systems*, vol.32, no.3, pp.2165-2185, 2017.
[94] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, and W. Ziarko,"Rough Set," *Communication of the ACM*, vol.38, no.11, pp.88–95, 1995.
[95] H. Cho, and M.K. An, "Co-clustering-based clustering and segmentation for pattern discovery from time course data," *International Journal of Information and Electronics Engineering*, vol.4, no.5, pp.358, 2014.
[96] E. Elhamifar, and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE transactions on pattern analysis and machine intelligence*, vol.35, no.11, pp.2765-2781, 2014.
[97] Z. Li, L. F. Cheong, S. Yang, and K. C. Toh, "Simultaneous Clustering and Model Selection: Algorithm, Theory and Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.40, no.8, pp.1964-78, 2017.
[98] D. Bazazian, R. G´omez, A. Nicolaou, L. Gomez, D. Karatzas, and A. D. Bagdanov, "Improving Text Proposals for Scene Images with Fully Convolutional Networks," *arXiv preprint arXiv:1702.05089*, 2017.
[99] T. He, W. Huang, Y. Qiao, and J. Yao, " Accurate Text Localization in Natural Image with Cascaded Convolutional Text Network," *arXiv preprint arXiv:1603.09423*, 2016.
[100] Wei Y, Zhang Z, Shen W, Zeng D, Fang M, and Zhou S. "Text detection in scene images based on exhaustive segmentation," *Signal Process: Image Communication,* vol.50, pp.1–8, 2017.
[101] M. Jiang, J. Cheng, M. Chen, and X. Ku, " An Improved Text Localization Method for Natural Scene Images," *Journal of physics: conference series*, vol.960, no.1, p.012027, 2018, doi:10.1088/1742-6596/960/1/012027.
[102] N. Robert, A. Dicker, and K. Meyer-Wegener, "NEOCR: A configurable dataset for natural image text recognition," *Camera-Based Document Analysis and Recognition*, pp.150-163, 2011, Springer Berlin Heidelberg.
[103] L. SeongHun, M. Su Cho, K. Jung, and J. Kim, "Scene text extraction with edge constraint and text collinearity," *international conference on pattern recognition,* pp. 3983-3986, 2010.
[104] D. Campos, T. Emídio, B. R. Babu, and M. Varma, "Character Recognition in Natural Images," *VISAPP*, vol.2, pp.273-280, 2009.
[105] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng. "Reading digits in natural images with unsupervised feature learning," *NIPS workshop on deep learning and unsupervised feature learning, Granada, Spain*, 2011.
[106] A. Mishra, A. Karteek, and C. V. Jawahar, "Scene text recognition using higher order language priors," *BMVC 23rd British Machine Vision Conference BMVA*, 2012.
[107] http://www.cvc.uab.es.
[108] http://www.iapr-tc11.org.
[109] http://wwwee.surrey.ac.uk.
[110] http://benchmarks.ai

امین‌اله مه‌آبادی تحصیلات خود را در رشته مهندسی برق سخت‌افزار و معماری کامپیوتر به انجام رسانده و استادیار گروه مهندسی کامپیوتر و فناوری اطلاعات دانشگاه شاهد است. تحقیقات مورد علاقه نامبرده پردازش داده‌های عظیم تصویری، سیستم‌های بی‌درنگ نهفته، پردازش موازی و ابزارهای هوشمند است.

آدرس پست الکترونیکی ایشان عبارت است از:

mahabadi@shahed.ac.ir

[1] Real-time Embeded System
[2] Cloud Computing
[3] Big Data
[4] Bot
[5] Big Image Data (BID)
[6] Agent
[7] Local Information
[8] Global Information
[9] Bot Net
[10] Smart Analyzer
[11] Strategic Planning
[12] Hierachical
[13] Text Localization
[14] Paragraph
[15] Sentence
[16] Word
[17] Character
[18] Variety
[19] Volume
[20] Complexity
[21] Camera Location
[22] Resolution
[23] Natural Noise
[24] Aritificial Noise
[25] Velocity
[26] Value
[27] Detection
[28] Localization
[29] Optical Character Recognition (OCR)
[30] Tracking
[31] Image Framing (IF)
[32] Agent-oriented
[33] Scalable
[34] Search Space
[35] Modeling
[36] Architecture
[37] Imaging
[38] Computational Complexity
[39] Precision
[40] Language dependency
[41] Natural Image
[42] Resource Optimization
[43] Nano technology
[44] Static
[45] Dynamic
[46] Centralized
[47] Decentralized
[48] Distributed
[49] Negotiation
[50] Agile
[51] Multi-scale
[52] Robust
[53] Object Detection (OD)
[54] Font
[55] Language
[56] Color
[57] Scale
[58] Direction
[59] Server Bots
[60] GPU
[61] Knowledge
[62] Meta Data
[63] Dark Silicon
[64] Uncertainty
[65] Decision Making
[66] Risk
[67] Risk-averse
[68] Risk-aware
[69] Threats
[70] Ground Truth
[71] Stroke Width (SW)
[72] Stroke Width Transform (SWT)
[73] Maximally Stable External Regions (MSER)
[74] Straightness
[75] False Positive (FP)
[76] Gradient Vector Flow (GVF)
[77] Map Reduce
[78] Hadoop
[79] Texture-based Methods (TBM)
[80] Connected Component-based Methods (CCM)
[81] Hybrid Methods (HM)
[82] Statistical Method
[83] Structural and Spectral (SS)
[84] Image Serach Space (ISS)
[85] Pipeline
[86] Wavelet
[87] Rough Set Theory
[88] Co-clustering
[89] Log
[90] Roberts
[91] Zerocross
[92] Canny
[93] Sobel
[94] Laplacian
[95] Mar-Hildreth
[96] Edge Block (EB)
[97] Eccentricity
[98] Strength
[99] Task Graph
[100] Application
[101] Truly Detected Block (TDB)
[102] Falsely Detected Block (FDB)
[103] Text Block with Missing Data (TBMD)
[104] Average Processing Time (APT)
[105] Utility Theory
[106] Dark Matter

The CSI Journal on Computing Science and Information Technology

The CSI Journal on
Computing Science and
Information Technology
Vol. 18, No.1, 2020

A Text Localization Model in Big Imade Data

Aminollah Mahabadi

Computer Engineering Department, Shahed University, Tehran, Iran

Acoustic Research Center, Shahed University, Tehran, Iran

Abstract

Conventional text localization methods in natural images have the challenge of a large search space for accurate, correct, fast, and efficient detection and become a hard problem to support big image data. Reducing complex computations can be designed to control search space, resource consuming, and cost by using software techniques in multicore hardware, grid structures, cloud, and fog compution. localizing and reading texts at different levels of paragraph, text, word, and character in big image data terribly adds to its complexity. In this paper, non-learning related works of text localization in small images and common methods of natural image data are presented and a suitable model with strategic image analyzer using smart agents and smart bots for processing big image data to localize the text are presented and explained by cyber bots. Also, the existing datasets and required metrics for general image and big image data are proposed.

Keywords: Text Localization, Big Image Data, Video images, Image Processing, Natural Images.