علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۱۵، شماره ۱، ۱۳۹۶
صفحات ۱۰- ۲۱
مقاله عادی

نشریه علوم رایانش و فناوری اطلاعات

روش عمومی قابل‌اعتماد موقعیت‌یابی متن در تصاویر طبیعی

امین‌اله مه‌آبادی علیرضا زارعی

دانشکده فنی و مهندسی، دانشگاه شاهد، تهران، ایران

چکیده

این مقاله یک روش عمومی دقیق و سریع خودکار تشخیص متون چندزبانی از تصاویر رنگی دوربین و ویدیو با زمینه پیچیده براساس مؤلفهٔ همبند مبتنی بر لبه را در سطح بلاک، کلمه و خط متن ارایه می‌دهد. این روش نسبت به رنگ، ابعاد متن، زاویهٔ دوربین، انحنای سطح تصویر و نورپردازی ناهموار مقاوم است و در تصاویر دارای پیچیدگی زمینه و زبان‌های مختلف به خوبی عمل می‌کند. نتایج دقیق آزمایشات تجربی در تصاویر متنوع طبیعی با داده‌های استاندارد، دارای نرخ فراخوانی ۰/۹۰، دقت ۰/۸۵، شاخص F ۰/۸۷و MDR برابر با ۱۳٪ در سطح خط متن و نرخ فراخوانی ۷۰٪، دقت ۷۴٪ و شاخص F برابر با ۷۱٪ در سطح کلمه است. این روش نسبت به آخرین روش‌ها علمی دارای بهبود نرخ فراخوانی، دقت، شاخص F و MDR در سطوح خط متن و کلمه همچنین برخوردار از قدرت تشخیص داده‌های متنوع جهت پشتیبانی از داده‌های عظیم تصویری است.

کلمات کلیدی: پردازش تصویر، تصاویر رنگی، تصاویر ویدیویی، موقعیت‌یابی متن، روش مبتنی بر مولفه‌های متصل، روش قابل‌اعتماد.

۱- مقدمه

امروزه دسترسی همگان به دوربین‌های دیجیتال ارزان قیمت، سبب ایجاد حجم عظیمی از تصاویر شده است. دستیابی به اطلاعات سودمند از این حجم عظیم داده، نیازمند پردازش‌های دقیق و با کارایی مناسب است. یکی از این پردازش‌ها، تشخیص متن در تصویر است [۱]. یافتن متن در تصاویر، کاربردهای زیادی در شناسایی علایم، خواندن پلاک خودروها، بایگانی اسناد و مانند آن دارد. امروزه یافتن محل متن به‌عنوان پردازش اصلی تبدیل تصویر به نوشتار، مورد استفاده قرار می‌گیرد. بررسی‌ها بیانگر آن است که کارایی تبدیل تصویر به نوشتار، وابستگی بسیاری به‌کارایی، دقت و سرعت روش یافتن متن مخصوصا در تصاویر مناظر طبیعی با ابعاد بزرگ و متون متنوع زیاد در داده‌های عظیم تصویری دارد [۱۵، ۴۱]. متون موجود در تصاویر شامل دو نوع متن‌های مناظر و متن‌های گرافیکی هستند. متن مناظر جزء تصاویری هستند که توسط دوربین‌های عکس‌برداری گرفته شده‌اند.

مثلا علائم جاده‌ها، آگهی‌ها، متن موجود بر روی خودوروها و نوشته‌های روی پیراهن از جمله این تصاویر هستند. متن‌های مناظر و چشم‌اندازها در مقایسه با متن‌های گرافیکی که ساختاریافته و مرتبط با موضوع‌اند، غیرقابل پیش‌بینی هستند [۱، ۵۱]. استخراج متون مناظر کاربردهای بسیار دارد و می‌تواند مثلا توسط ربات‌های سیار برای کشف تابلوهای راهنمایی مبتنی بر متن، کشف و تشخیص پلاک خودرو، شناسایی اشیاء و مانند آن مورد استفاده قرار گیرد [۲]. یک سیستم استخراج اطلاعات متن از تصویر، یک ورودی به شکل تصویر ساکن یا فریم ویدئو را دریافت می‌کند [۳]. تصاویر آن می‌توانند دارای سطح خاکستری یا رنگی باشند.

استخراج و فهم اطلاعات متن شامل مراحل (۱) آشکارسازی متن[1] و (۲) موقعیت‌یابی متن[2]، و (۳) خواندن محتوای متن[3] است. ساختار یک سیستم استخراج اطلاعات متن شامل تشخیص متن، موقعیت‌یابی، شاخص‌کردن متن در سطح بلاک یا کلمه و حرف مانند نمایش شکل ۱ است. به‌دلیل رابطهٔ تنگاتنگ آشکارسازی و موقعیت‌یابی متن، محققان آن را آشکارسازی متن و در برخی موارد موقعیت‌یابی متن معرفی کرده‌اند [۱۲]. در مرحلهٔ آشکارسازی و موقعیت‌یابی متن، تصویر برای یافتن احتمال وجود متن کاوش می‌شود و موقعیت مکانی متون در تصویر با شناسایی و عمل قاب‌گذاری محدوده[4] متن‌ها انجام می‌شود. خروجی حاصل، تعیین قاب بر محدوده متن، ابعاد آن را جهت حذف نواحی غیرمتن

مشخص می‌کند. نهایتاً کاراکترهای متن از زمینه جدا و محتوای متن به‌روش تشخیص نوری کاراکتر قابل خواندن و فهم می‌شود. از آنجا که معمولاً نواحی متن در مناظر طبیعی در معرض نویز قرار می‌گیرند و برای پردازش از درجه تفکیک‌پذیری پایینی برخوردارند باید برای کاهش خطا، تصویر قاب متن استخراجی در سطح تشخیص بلاک، خط متن، کلمه یا کاراکتر، از ارتقاء کیفیت برخوردار شود [۵۲]. برای موقعیت‌یابی متن باید متون به‌طور کامل از زمینه که ممکن است نویزی باشد جدا شود [۵۵-۵۷]. سپس قاب استخراجی به یک تصویر باینری تبدیل و به‌منظور افزایش دقت تشخیص، بلاک حاوی آن به سطح کیفیت تصویری بهتر ارتقاء می‌یابد. الگوی موقعیت‌یابی متن دارای چالش‌هایی از نظر (۱) معماری روش تشخیص، (۲) شیوه تصویر برداری، (۳) پیچیدگی محاسبات، (۴) دقت تشخیص، (۵) وابستگی به‌زبان، (۶) تشخیص در تصاویر طبیعی و (۷) پردازش داده‌های عظیم تصویری است که به آن‌ها می‌پردازیم.

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

(۲) شیوه تصویربرداری: از نظر مشخصات تصویر ورودی به‌شرایط 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 فعلی با هدف چابک‌سازی روش فقط به حل چالش‌های این بعد اشاره می‌شود.

(۳) محاسبات پیچیده: گرچه به‌نظر نمی‌آید که انجام محاسبات یافتن متن بسیار پیچیده باشد ولی سرعت روش در کاربردهای بی‌درنگ و توجه به آن برای کاهش محاسبات یا درجه پیچیدگی هر روش جهت پشتیبانی از داده‌های عظیم لازم است. مثال آن کاهش فضای جستجو یا جستجوی موازی در فضای تصویر جهت افزایش سرعت در موقعیت‌یابی است.

(۴) دقت تشخیص: در بعضی کاربردها مانند تصاویر کتب چاپی از نظر دقت و صحت تشخیص، روش‌ها از معیارهای اندازه‌گیری خاص خود بهره می‌برند که در تغییر تنوع متن از دقت تشخیص کاسته می‌شود و بعضاً آن روش کاربردی ندارد [۵۳]. بیشتر روش‌های موجود به‌دلیل تکیه بر ویژگی‌های خاص تصویر، قادر به‌تشخیص متون خاص مانند کتاب یا صفحات تایپی وب هستند و ممکن است متون دیگر را تشخیص ندهند. لذا عموما به افزایش دقت در کنار تشخیص تصاویر عامه توجه می‌شود.

(۵) وابستگی به زبان: در بعضی کشورها مانند کره، چین و ژاپن برای راهنمایی خارجی‌ها استفاده از چندین زبان در حمل و نقل، تابلوهای ترافیکی و مانند آن مرسوم است. بیشتر روش‌های تشخیص برای کاربرد زبان خاص طراحی می‌شوند و قادر به تشخیص همزمان زبان‌های دیگر نیستند [۳۶-۳۷]. تشخیص متون چندزبانی برای یک سیستم قابل‌اعتماد لازم و با افزایش تعداد زبان، پیاده‌سازی آن بسیار مشکل است.

(۶) تصاویر متنوع طبیعی: چالش تشخیص متن در تصاویر طبیعی به‌دلیل شرایط کنترل نشده، زمینه پیچیده و تغییرات شدید الگوهای متن مانند نوع فونت، زبان، رنگ، مقیاس و جهت آن است که هنوز در تشخیص بی‌درنگ و متون چندزبانی بدون راه‌حل است [۳۶، ۵۴]. شرایط کنترل شده در تشخیص برای یک ماشین ربات وجود ندارد لذا در بسیاری کاربردها مانند ربات‌های خدمات‌رسان پردازش تصاویر طبیعی لازم است.

(۷) داده‌های عظیم تصویری: با افزایش سریع کاربردهای تصویر توسعه داده‌های عظیم تصویری[۵] این چالش از نظر پردازش در تشخیص سطوح مختلف خط متن، کلمه و کاراکترها به‌دلیل حجم زیاد و تنوع متون داده‌های عظیم تصویری از مشکلات دنیای آینده است و روش تشخیص باید بسیار سریع و مقیاس‌پذیر باشد. در آینده نه‌چندان دور از مشکلات بشریت خواهد بود لذا طراحی روش‌های دارای پردازش سریع و مقیاس‌پذیر حیاتی است.

از نظر الگوی متن‌یابی روش‌های موجود در سه گروه مبتنی بر بافت[۶]، مبتنی بر مؤلفه‌های همبند[۷] و روش‌های ترکیبی[۸] 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 (مثلا با ماشین‌های یادگیری یا رده‌بند) می‌یابند [۳۵]. به‌طور کلّی در زمینه‌های پیچیده، مدل‌های مبتنی بربافت از مدل‌های مبتنی بر مؤلفه‌های همبند پایدارترند و بیشتر به سرعت پردازش توجه دارند. با توجه به آن‌که تصویر دارای ویژگی متفاوت بافت با متن است روش‌های مبتنی برباقت می‌توانند به‌راحتی متن را از زمینه تفکیک کنند. عیب اصلی مدل‌های مبتنی برباقت، بخش‌بندی بسیار 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[۹]، طیفی و ساختاری[۱۰] بنا شده‌اند.

شکل ۱- تشخیص متون متنوع در تصاویر مختلف

مدل‌های مبتنی بر مؤلفه‌های همبند براساس 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 که برای متون زیرنویس و با رنگ یکنواخت مناسب ولی برای متون با کاراکترهای چندرنگی و پس‌زمینهٔ‌های پیچیده در ابعاد بزرگ کارآیی ندارند. بیشتر توجه آنها به‌دقت و فراخوانی تشخیص است.

دسته‌های ترکیبی معمولا ترکیبی از هر دو روش فوق است که از مزایای هر دو بهره می‌برند. در سال‌های اخیر به‌دلیل سختی کار، پیچیدگی‌های پیاده‌سازی، دقت نامناسب و سرعت کم در دو بعد کاراکتر و سطح کلمه روش ترکیبی به‌ندرت مورد استفاده قرار گرفته است [۴۲].

پس از مطالعه آخرین روش‌های علمی و مشاهدهٔ مشکلات موجود آن‌ها برای تصاویر ثابت [۳۸-۴۴] و ویدیویی [۲۶-۳۰]، فقدان روش عمومی دقیق که ضمن تشخیص در سطوح خط متن و کلمه از سرعت کافی در تصاویر با متون و ابعاد زیاد مناسب داده‌های بزرگ باشد احساس می‌شود. روش‌های موجود نتوانسته‌اند در دو جهت عمومیت روش و دقت کافی تشخیص خط متن و کلمه برخوردار باشند. مثلا بعضی الگوریتم‌ها در سطح خط متن نتایج نسبتاً خوبی دارند ولی در سطح کلمه مطلوب نیستند [۵۵].

روش‌های جدید درحال تمرکز بر کاربرد در داده‌های عظیم تصویری هستند [۴۱-۴۲]. دو مشخصه داده‌های عظیم عبارت از تنوع داده و حجم بسیار زیاد آن است. لذا یک سیستم قابل کاربرد برای داده‌های عظیم 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[11] در تصویر همراه با افزایش سرعت آن لازم است.

انگیزه ما با مطالعه و پیاده‌سازی مقالات مهم، انتخاب موقعیت‌یابی مبتنی بر مولفه‌های همبند، ارایه روش عمومی قابل‌اعتماد و سریع تشخیص متون متنوع در تصاویر طبیعی است که بتواند از مزایای تمامی آن‌ها در سطوح تشخیص بلاک، خط متن و کلمه برخوردار باشد و با غلبه بر مشکل پیچیدگی ابعاد تصویر ورودی، نتایج آن با مقالات نوین رقابت کند. همچنین ضمن دارا بودن ساختار سلسله مراتبی و قابلیت پردازش موازی عملیات، با کاهش پیچیدگی‌های پیاده‌سازی بتواند از دقت و سرعت مناسب در سطوح تشخیص بلاک، کلمه و کاراکتر برخوردار باشد. در برابر چالش‌های نورپردازی ناهموار غیریکنواخت، سطوح ناهموار، تغییرات شدید رنگ و مقیاس فونت، چندزبانی و پیچیدگی پس‌زمینهٔ تصویر مقاوم باشد. همچنین با هدف حرکت به سمت پشتیبانی از داده‌های عظیم تصویری، ارایهٔ روش مبتنی بر مؤلفه‌های همبند برای رفع چالش‌های فوق در تصاویر مناظر طبیعی ضروری است. برای کاربرد وسیع آن باید از قابلیت شناسایی تصاویر دوربین و ویدیویی و همچنین قابلیت پیاده‌سازی بر ساختارهای موازی برخوردار گردد. لذا تمرکز اصلی این روش باید بر محور تشخیص کانتورهای بسته به‌منظور تشخیص کاراکترها، کلمات و بلاک‌های متن باشد.

براساس دانش ما در راستای ارایه یک سیستم برای داده‌های عظیم تصویری، روش عمومی تشخیص مبتنی بر مؤلفه‌های همبند در تصاویر طبیعی که بتواند همزمان با داده‌های متنوع و زبان‌های مختلف در سطوح خط متن و کلمه نتایج دقیقی ارایه دهد و بتواند در برابر چالش‌های تصاویر پیچیده مقاومت کند و از کاهش زمان محاسبات برخوردار باشد مشاهده نشد. ما برای حرکت به‌سمت این سیستم‌ها در روش پیشنهادی این مقاله، اولین هدف را پشتیبانی از تنوع داده‌ها در ابعاد مختلف در نظر گرفته‌ایم تا بسیاری از چالش‌ها فوق را رفع کنیم. به‌طور خلاصه نوآوری‌های ما عبارت از ارایه

- روش عمومی دقیق و سریع تشخیص در تصاویر طبیعی در سطوح بلاک، خط متن و کلمه دارای قدرت پیش‌بینی و استخراج مؤلفه‌های همبند شبیه متن، با فراخوانی بالا،
- کلاسه‌بندی مناسب ویژگی‌های تصویر جهت حذف نواحی بستهٔ غیرواقعی مشابه متن برای افزایش دقت و تشخیص هم‌زمان متون چند مقیاسی،
- ارایه معیار اندازه‌گیری چگالی بلاک جهت حذف بلاک‌های غیرمتن و افزایش درجهٔ تفکیک‌پذیری کلمات، و
- قابلیت تشخیص هم‌زمان‌داده‌های متنوع، زبان‌های مختلف و متون با مقیاس متفاوت در تصاویر دوربین و ویدئویی در راستای پشتیبانی ازداده‌های عظیم تصویری است.

در ادامه مقاله و در بخش ۲ مهم‌ترین کارهای علمی مرتبط با استخراج متن از تصویر را شرح می‌دهیم. در بخش ۳ ضمن تعریف مساله موقعیت‌یابی در تصاویر رنگی طبیعی، به مدل‌سازی و ارایه روش پیشنهادی می‌پردازیم. در بخش ۴ نتایج آزمایش‌های تجربی و تحلیل آن بیان و نهایتاً در بخش ۵، نتیجه‌گیری و کارهای آینده ارایه می‌شود.

۲- کارهای مرتبط

به‌دلیل توجه زیاد محققان به‌روش‌های مبتنی بر مؤلفه‌های همبند برای مکان‌یابی متن در سال‌های اخیر، اکثر کارهای موجود در این حوزه است [۹-۲۰]. روش ارایه شده جین و همکاران [۱۰] تصاویر را توسط خوشه‌بندی رنگی، مؤلفه‌های گروه‌بندی شده در خط متن، تجزیه تحلیل مؤلفه‌ها و حذف مؤلفه‌های غیرمتنی براساس 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[12] نزدیک به‌هم کاراکترها توسط اپشتاین با ارایه عملگر جدیدی به‌نام تغییر پهنای قلم[13] انجام شد [۱۱]. روشی آسان برای بازیابی پهنای قلم کاراکترها در نقشهٔ لبه 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 افقی متن توجه دارد. نویمانو همکاران [۱۲]، الگوریتم تشخیص متن براساس نواحی حدی بیشینه پایدار[14] 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. به‌دلیل 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 [۱۱-۱۲] که در بسیاری از آثار از جمله [۱۳-۱۷] استفاده شده و عامل موفقیت‌های بزرگی در تشخیص چهره [۱۸] و حذف نویز تصویر [۱۹] به شمار می‌آیند. ژائوو همکاران [۲۰]، یک فرهنگ لغت از نمونه‌های آموزشی ساختند و از آن به‌عنوان تصمیم گیرنده درمورد 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-mean را در تشخیص محل کاندیداهای متن برمبنای بیشینهٔ اختلاف بکار بردند. برای جداسازی رشته‌های متنی از یکدیگر از اسکلت هر مؤلفهٔ همبند استفاده کردند. در نهایت میزان صافی طول[15] و چگالی لبه رشتهٔ متنی را برای حذف نواحی اشتباه[16] تشخیص داده شده 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[17] 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 نهایتا با گروه‌بندی و حذف مولفه‌های غیرمتن، مولفه‌های 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 ارایه شد که باعث بهبود تشخیص گردید [۳۱-۳۵، ۳۸-۴۰]. افزایش عملکرد این روش‌ها تا حد زیادی به‌دلیل آموزش و بکارگیری داده‌های بزرگ آموزشی است که در دسترس عموم نیست. از آنجا که کلید مؤثر آشکارسازی و تشخیص متن در ساختار تقسیم‌بندی کاراکترها و روابط بین آن‌ها نهفته است این روش را از یک روش‌عمومی خارج و به حوزه خاص محصور می‌کند.

در آخرین روش‌های علمی تعداد کمی از تحقیقات بر مباحث قبلی توجه کرده‌اند [۳۸-۴۰]. روش یانگو 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 کارآیی خوب است فقط مشکل کندی و یادگیری شبکه دارد. روش واسی پولوس [۴۰] از روش‌های مبتنی بر ناحیه است که تغییرات رنگ زیاد در متن را قبول نمی‌کند و کنتر است متن و زمینه باید به اندازه کافی زیاد باشد. همچنین نیازمند دانش از شکل کاراکترها ولی نسبت به نوع فونت و ابعاد آن مقاوم و بیشتر مناسب متون کتاب و مجلات است. تحقیقات جدید با تمرکز بر پشتیبانی از داده‌های عظیم تصویری سعی بر ارایه روش‌های مقیاس‌پذیر دارند [۴۱-۴۲]. این روش‌ها علاوه رفع چالش‌های قبلی باید از قابلیت کاهش فضای جستجوی برخوردار باشند، از تنوع داده‌های مختلف در ابعاد متفاوت پشتیبانی کنند و از

سرعت لازم برای تشخیص سریع بهره‌مند شوند تا مناسب دنیای آینده گردند. لذا ارایه روش عمومیِ تشخیص تصاویر متنوع طبیعی با زبان‌های مختلف در سطوح بلاک، خط متن، کلمه و حرف برای نتایج دقیق در داده‌های عظیم ضروری است تا ضمن حل چالش‌های قبلی از کاهش زمان جستجو و پردازش برخوردار گردد. سیستم‌های آینده دارای دو چالش اساسی پشتیبانی از داده‌های متنوع و مقیاس‌پذیری روش هستند. خلاصه بررسی نقاط قوت و ضعف روش‌های مهم در جدول ۱ آمده است.

جدول ۱- مقایسه الگوریتم‌های آشکارسازی و مکان‌یابی متن

الگوریتم	نقاط قوت	نقاط ضعف
ژانگ [۵۰]	- به‌مناسب تصویر با زمینه ساده. - عمل خوب در تصاویر طبیعی.	- ضعیف در تصاویر با زمینه پیچیده. - فقدان مقیاس‌پذیری.
جین [۱۰]	-مناسب تصویر با زمینه ساده. - عمل خوب در تصاویر طبیعی.	- ضعیف در تصاویر با زمینه پیچیده. - متکی به قوانین تعریف شده دستی. - فقدان مقیاس‌پذیری.
لی [۵۱]	- مناسب آشکارسازی متون متحرک ویدئویی.	- تنها مناسب متون افقی.
کیم [۵۳]	- بمناسب تصویر با زمینه ساده. - عمل خوب در تصاویر طبیعی و ویدئویی.	- ضعیف در تصاویر با زمینه پیچیده. - مناسب متون افقی. - فقدان مقیاس‌پذیری.
چن [۵۷]	- آشکارسازی تصاویر با زمینه پیچیده. - عملکرد سریع.	- مناسب متون افقی. - فقدان مقیاس‌پذیری.
لیو [۵۴]	- مناسب آشکارسازی فریم‌های ویدئویی. - آشکارسازی چند زبان مختلف.	- تنها مناسب متون افقی. - فقدان مقیاس‌پذیری.
لیو [۵۵]	- عمل خوب در تصاویر طبیعی.	- مناسب متون افقی.
وانگ [۵۶]	- آشکارسازی در تصاویر با زمینه پیچیده.	- مناسب متون افقی. - نیازمند لغت‌نامه برای تصویر. - فقدان مقیاس‌پذیری.
اپشتاین [۱۱]	- آشکارسازی در تصاویر با زمینه پیچیده. - آشکارسازی در چند زبان مختلف. - عملکرد سریع.	- تنها مناسب متون افقی. - متکی به قوانین تعریف شده دستی. - فقدان مقیاس‌پذیری.
نیومن [۱۲]	- آشکارسازی در تصاویر با زمینه پیچیده. - عملکرد سریع.	- مناسب متون افقی. - فقدان مقیاس‌پذیری.
یی [۵۲]	- آشکارسازی در جهات مختلف. - آشکارسازی چند زبان مختلف.	- متکی به قوانین تعریفی‌شده. - ضعیف در تصاویر با زمینه پیچیده.
شیواکومار [۲۱]	- آشکارسازی در جهات مختلف. - مناسب خط متن.	- متکی به قوانین تعریف شده. - فقدان مقیاس‌پذیری.
یائو [۱۳]	- آشکارسازی سریع در جهات مختلف. - آشکارسازی چند زبان مختلف.	- متکی به قوانین تعریف شده. - فقدان مقیاس‌پذیری.
هوانگ [۴۸]	- آشکارسازی در تصاویر با زمینه پیچیده.	- مناسب متون افقی. - متکی به قوانین تعریف شده. - فقدان مقیاس‌پذیری.
هوانگ [۴۹]	- آشکارسازی در تصاویر طبیعی. - عملکرد خوب.	- تنها مناسب متون افقی. - فقدان مقیاس‌پذیری
یانگ [۵۸]	- آشکارسازی در تصاویر طبیعی. - تشخیص فقط خط متن. - روش کند با عملکرد خوب.	- یادگیری شبکه عصبی. - مناسب متون افقی. - فقدان مقیاس‌پذیری
واسیلو پلوس [۴۰]	- روش مبتنی بر ناحیه و تصویر دوربین. - متون هم‌رنگ غیرمتنوع. - تفاوت کنتراست زیاد متن و زمینه.	- نیازمند دانش شکل کاراکترها. - عدم تنوع رنگ متون. - فقدان مقیاس‌پذیری.
روش پیشنهادی	- تصاویر دوربین و ویدئویی. - تشخیص داده‌های متنوع در تصاویر پیچیده. - آشکارسازی سریع چندین زبان مختلف. - تشخیص هم‌زمان بلاک، خط متن و کلمه.	- مناسب متون افقی با چرخش کم. - مناسب تقعر و تحدب متوسط متن. - فقدان مقیاس‌پذیری.

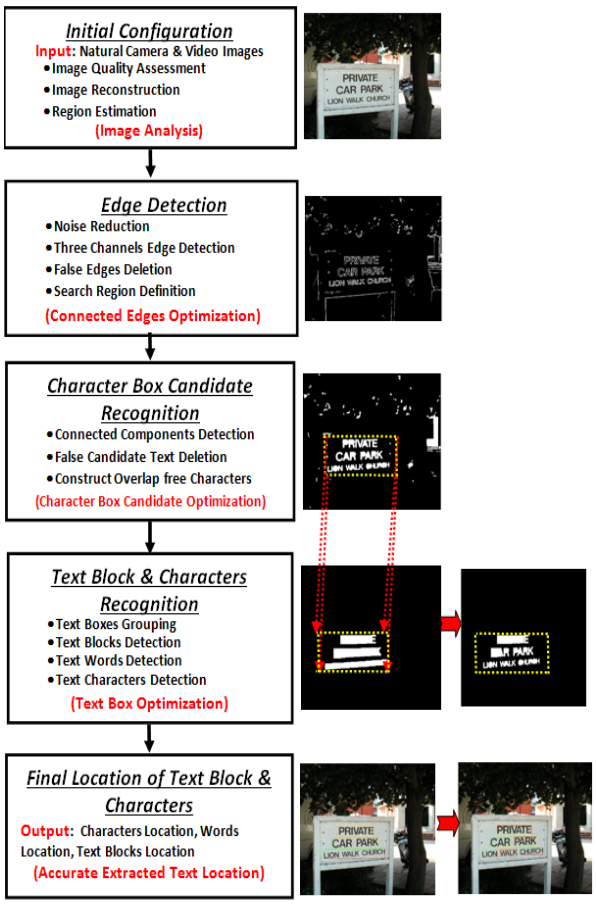
۳- روش پیشنهادی

دو ویژگی بارز در متون تصاویر وجود یکپارچگی و کنتر است خوب متن با زمینه و شکل نواحی بسته لبه‌های متون است. با توجه به این ویژگی‌ها، ما در روش پیشنهادی از استخراج لبه و ویژگی بسته بودن کاراکترهای متن جهت شناخت کاندیداهای اولیه متون استفاده و برای افزایش سرعت، آن را مقاوم‌سازی می‌کنیم. شخیص 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، وضعیت ضخامت متن، زاویه دید و بستر مسطح یا غیرمسطح تصویر دارد.

شخیص 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 نورمحیط، وضعیت ضخامت متن، زاویه دید و بسترمسطح یا غیرمسطح تصویر دارد. ما روشی ارایه دادیم که با دقت بالای تشخیص در سطوح بلاک و کلمه مناظر طبیعی در پس‌زمینهٔ پیچیده دارای قابلیت شناسایی تصاویر ویدیویی نیز هست. مطابق دیاگرام شکل ۲، روش پیشنهادی ابتدا لبه‌های تصویر را تعیین و سپس با تقسیم تصویر، کاراکترهای کاندیدا را تشخیص می‌دهد. نهایتا محل دقیق کاراکترها و کلمات و بلاک‌های متنی را تعیین و قاب‌گذاری می‌کند که در ادامه به شرح جزئیات الگوریتم‌های روش پیشنهادی می‌پردازیم.

۳-۱- لبه‌یابی

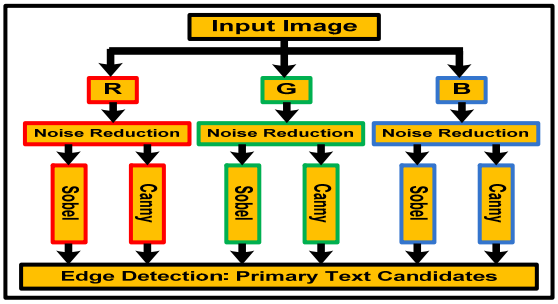
شکل ۳ مراحل لبه‌یابی را نشان می‌دهد که از دو لبه‌یاب همزمان کنی و سوبل برای تشکیل لبه‌های کاندیدهای اولیهٔ متن و یافتن مرز اشیاء موجود استفاده می‌شود.



شکل ۲- دیاگرام بلاکی روش پیشنهادی

با توجه به نقاط ضعف و قوت لبه‌یاب‌های سوبل و کنی (شکل ۴ و ۵) ما بـرای بهره‌برداری و قدرت هر دو لبه‌یاب و برای یافتن لبه‌های متن به‌شـکل محـیط‌هـای بسته از ترکیب هم‌زمان این دو استفاده کرده‌ایم. یعنی پس از حذف نویز در هر سه کانال رنگی، با لبه‌یاب کنی با آستانه تجربـی [۰/۲-۰/۱] و لبـه‌یـاب سـوبل در هـر کانـال لبـهٔ بـا بهـره از رابطـه (۱) در تصـاویر عمـل مـی‌کنـیم. کـه در آن رابطـه E_G ، E_R و E_B لبه‌های تصویر سه‌کانال رنگ، E_{BS}، E_{GS} و E_{RS} لبه‌های تصویر سه کانال رنگ با لبه‌یاب سوبل، E_{BC}، E_{GC} و E_{RC} لبه‌های تصویر متناظر با سه کانـال رنگ با لبه‌یاب کنی هستند. عملگر همـان عملگـر منطقـی OR اسـت. اسـتفاده از ترکیب این دو لبه‌یاب توانسته نقاط ضعف هر دو لبه‌یاب را از بین برده (شکل ۶) و لبه‌های متن را به صورت محیط‌های کاملاً بسته آشکار سازد. هدف آن آماده‌سـازی تصویر برای یافتن و تقلیل نواحی متن است.

$$E=E_R E_G E_B$$
$$E_R=E_{RS} E_{RC}$$
$$E_G=E_{GS} E_{GC} \quad (۱)$$
$$E_B=E_{BS} E_{BC}$$



شکل ۳- مرحله لبه‌یابی در روش پیشنهادی

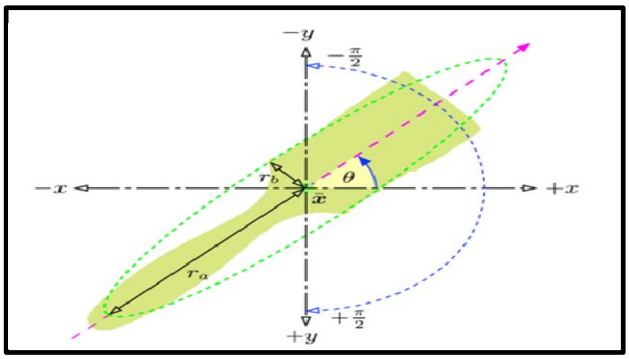
(الف) (ب) (ج)

شکل ۴- تصاویر لبه (الف) تصویر اصلی (ب) تصویر لبهٔ سوبل (ج) تصویر لبهٔکنی

(الف) (ب) (ج)

شکل ۵- تصاویر لبه (الف) تصویر اصلی (ب) تصویر لبهٔسوبل (ج) تصویر لبهٔکنی

شکل ۶- تصاویر لبهٔ شکل‌های ۴ (الف) و ۵ (الف) در روش پیشنهادی



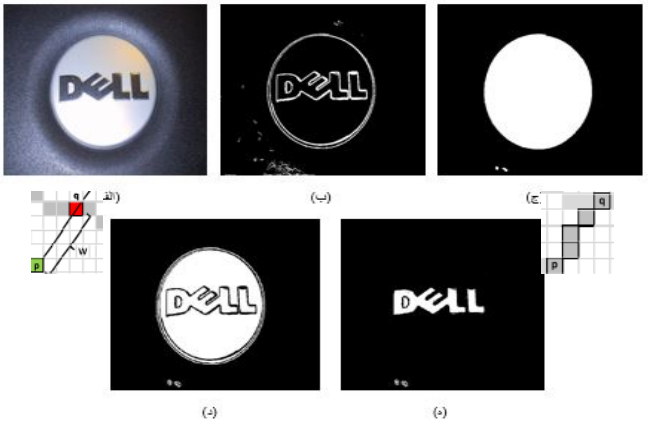
شکل ۷- محورهای فهم ویژگی گریز از مرکز

(ب) (الف)

شکل ۸- بلاک‌های کاندیدای متن (الف) بلاک‌ها قبل حذف هندسی، (ب) بلاک‌ها بعدحذف هندسی

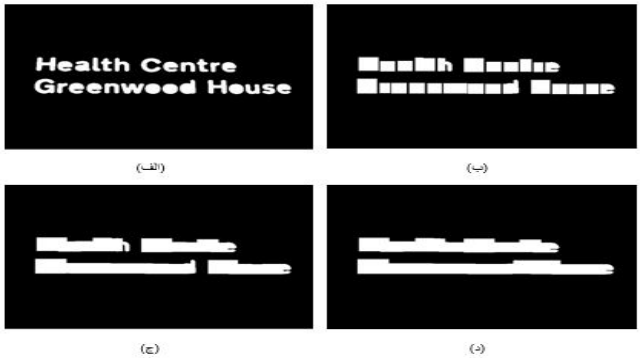
(الف) (ب) (ج) (د)

شکل ۹- تشکیل کاندیدهای اولیهٔ متن در روش پیشنهادی (الف) تصویر اصلی (ب) تصویر لبه (ج) حذف لبه‌های غیرمتن (د) پرکردن گودال‌ها



(الف) (ب) (ج) (د) (ه)

شکل ۱۰- بکارگیری معیار پهنای قلم (الـف) تصـویر اصـلی (ب) تصـویر لبـه (ج) خروجی حذف لبه و حذف هندسی (د) جداسازی لبه‌ها (ه) حذف با معیـار پهنـای قلم



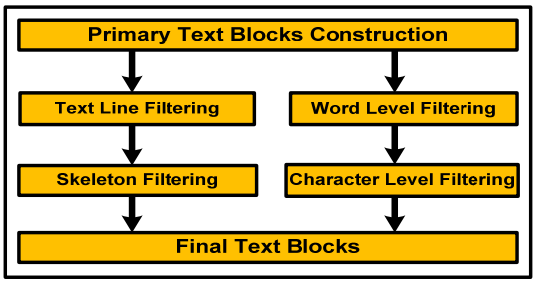
شکل ۱۱- بلاک‌های اولیه کاندیدای متن (الف) متن باینری (ب) سطح کاراکتر (ج) سطح کلمه (د) سطح خط متن

۳-۲- حذف کاندیداهای غیرمتن

حذف کاندیدهای غیرمتن در سه مرحله برای برخورداری از دقت بالای تشخیص صورت می‌گیرد. این مراحل عبارت از حذف لبه‌های غیرمتن، حذف براساس مشخصات هندسی و براساس پهنای قلم است که در ادامه به شرح آن‌ها می‌پردازیم.

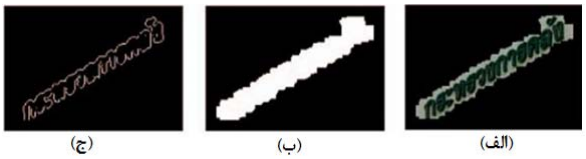
حذف لبه‌های غیرمتن: آشکارسازی لبه با برچسب‌گذاری در هشت همسایگی مؤلفه‌های همبند و اطلاعات بلاک‌های مجاور صورت می‌گیرد. هر عنصر بدست آمده را بلاک لبه[۱۸] EB می‌نامیم. ما فرضیات تجربی و علمی برای حذف محل‌های غیرمتن اعمال می‌کنیم. نسبت ابعاد بلاک‌های بین ۰/۱ و۱۰ را برای بررسی نواحی باریک در نظر می‌گیریم که ابعاد EB باید بزرگتر از ۱۵پیکسل و کوچکتر از ۰/۲ ابعاد تصویر باشد.از آنجا که لبهٔ آشکار شده، 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 می‌شود. لذا احتمال دارد که هر EB کامل، یک یا چند EB را احاطه کند. مثلا 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 وEB خارجی حذف می‌شود. بنابر لبه‌های مؤلفه‌های غیرمتن با در نظر گرفتن محدودیت هر لبهٔ مؤلفه باشرط زیر حذف می‌شوند.

If (N_{int}> 2) { Accept: EB_{int}, Reject: EB_{out}}



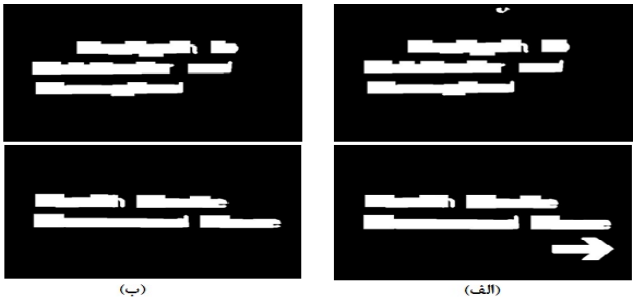
شکل ۱۲- مراحل حذف بلاک‌های غیرمتن و تشکیل بلاک‌های نهایی روش پیشنهادی

که در آن EB_{int} به EBهایی که کاملاً در EB فعلی قرار دارد و مساحت آنها بزرگتر از ۷۰۰ است و N_{int} به تعداد EB_{int} اشاره می‌کند. محدودیت‌های روی لبهٔ مؤلفه‌های کاندیدا، مؤلفه‌های غیرمتن را بطور مؤثر حذف می‌کند. گودال‌های جامانده پس از حذف لبه‌های غیرمتن با استفاده از عملیات مورفولوژی، پرُ می‌شوند و بر مبنای شکل ۹ کاندیداهای اولیه متن بدست می‌آیند.



شکل ۱۳- حذف اشتباهات بر مبنای چگالی لبه (الف) متن (ب) مؤلفهٔ همبند (ج) لبه

شکل ۱۴- بلاک‌های کاندیدای متن (الف) قبل حذف سطح کلمه (ب) بعد حذف سطح کلمه



شکل ۱۵- بلاک کاندیداهای متن (الف) قبل حذف سطح کاراکتر (ب) بعد حذف سطح کاراکتر

شکل ۱۶- اسکلت خط متن (الف) مؤلفهٔ همبند، (ب) اسکلت.بلاک کاندیدای متن

شکل ۱۷- حذف خط متن (الف) قبل حذف خط متن، (ب) بعد حذف و تشکیل خط متن

حذف هندسی: برای حذف نواحی غیرمتن ما مجموعه‌ای از ویژگی‌های هندسی انعطاف‌پذیر مساحت ناحیه، گریز از مرکز[۱۹] (نسبت طول محور اصلی به محور فرعی) مطابق شکل ۷ و رابطه (۲) و استحکام[۲۰] رابطه (۳) بر روی هر مؤلفهٔ همبند تعریف می‌کنیم. در ابتدا همهٔ مؤلفه‌های بسیارکوچک پراکنده و سپس مؤلفه‌های بزرگتر و با طول زیاد حذف می‌شوند. با انتخاب آستانهٔ محافظه کارانه می‌توان

مطمئن شد که بعضی حروف کشیده 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 هندسی تجربی (مساحت، گریز از مرکز و استحکام) جهت حذف مولفه‌های غیرمشابه متن عبارت از مؤلفه‌هایی با مساحت کمتر از ۵۰ پیکسل،گریز از مرکز بزرگتر از ۰/۹۹۸ (گریز از مرکز خط برابر ۱ و گریز از مرکز دایره برابر ۰) و استحکام کمتر از ۰/۳است. شکل ۸ نتیجه حذف هندسی را نشان می‌دهد.

$$\text{Eccentricity}_i = \frac{a_1}{a_2} = \frac{\mu_{20}+\mu_{02}+\sqrt{(\mu_{20}+\mu_{02})^2+4\,\mu_{11}^2}}{\mu_{20}+\mu_{02}-\sqrt{(\mu_{20}+\mu_{02})^2+4\,\mu_{11}^2}} \quad (۲)$$

$$r_a = 2\left(\frac{\lambda_1}{|R|}\right)^{\frac{1}{2}} = 2\left(\frac{2a_1}{|R|}\right)^{\frac{1}{2}}, r_b = 2\left(\frac{\lambda_1}{|R|}\right)^{\frac{1}{2}} = 2\left(\frac{2a_2}{|R|}\right)^{\frac{1}{2}}$$
$$\text{ComponentStrength}_i = \frac{\text{ComponentArea}_i - \text{HoleArea}_i}{\text{ComponentArea}_i} \quad (۳)$$

شکل ۱۸- بلاک‌های نهایی (الف) بلاک سطح کلمه (ب) بلاک خط متن (ج) قاب محدوده کلمه (د) قاب محدوده خط متن

۳-۳- حذف براساس پهنای قلم

عملگر پهنای قلم به‌وسیلهٔ تغییرات فاصله (فاصلهٔ اقلیدسی هر پیکسل پیش‌زمینه نسبت به نزدیک‌ترین پیکسل زمینه) برای هر پیکسل محاسبه و با مقدار محاسبه شده برچسب می‌زند. قلم را یک ناحیه پیوسته تصویر که پهنایش تقریبا ثابت است تعریف می‌کنیم. پهنای واقعی را نمی‌دانیم و آن را با لبه‌یاب‌ها محاسبه (براساس جهت گرادیان d_p برای هر پیکسل لبه p بیان) و سپس بازیابی می‌کنیم. اگر p در راستای قلم باشد پس d_p باید عمود بر مسیر قلم است. با دنبال کردن شعاع $r = p + n.d_p$ تا یافتن پیکسل لبه q ادامه می‌دهیم که سگمنت $[p,q]$ با عرض $\overrightarrow{\|p-q\|}$ تعریف می‌شود.

در برخی تصاویر مانند مارک‌ها و برچسب‌های تجاری به‌دلیل آن است که کل متن یک (یا دو) ناحیه بسته تلقی می‌شود لبه‌های متون هم‌پوشانی دارند. در مرحلهٔ حذف لبه‌های غیرمتن، قاب اطراف متن حذف نمی‌شود. لذا اول 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 مانند شکل ۱۰ حذف می‌کنیم.

۴-۳- حذف بلاک‌های غیرمتن

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) مطابق شکل ۱۲ تشکیل و بلاک‌های غیرمتن در سطوح مختلف و در طی چند مرحله حذف می‌شوند. پس از تشکیل بلاک‌های متن، عمل حذف غیرمتن در سطوح کلمه، خط متن و سپس کاراکتر و براساس اسکلت‌سازی انجام می‌شود که در ادامه به بیان آن می‌پردازیم.

حذف سطح کلمه: با تشکیل بلاک‌های کاندیدای کلمه ما از ویژگی‌های چگالی لبه، چگالی بلاک و نسبت ابعاد برای حذف بلاک‌های غیرمتن استفاده می‌کنیم. نواحی متنی چگالی لبهٔ (رابطه (۴)) بیشتری نسبت به نواحی غیرمتنی دارند (شکل ۱۳). معیار چگالی بلاک (رابطه (۵)) که ما ارایه داده‌ایم بر این فرض استوار است که نواحی متنی دارای چگالی بلاک نزدیک به ۱ هستند. بلاک‌هایی با معکوس چگالی بلاک بزرگتر از ۲ و کمتر از ۶ بلاک‌های غیرمتن هستند و حذف می‌شوند. بلاک‌های کلمه معمولا به‌شکل مستطیل و نسبت طول آنها به عرض‌شان بیشتر است. از ویژگی اختلاف ابعاد بلاک با توجه به مساحت کلمه نیز استفاده می‌کنیم. لذا بلاک‌هایی که دارای این شرایط باشند بلاک‌های غیرمتن تلقی و حذف می‌شوند: (۱) بلاک با مساحت بزرگتر از ۷۰۰۰ که اختلاف طول و عرض آن کمتر از ۲۵ پیکسل؛ (۲) بلاک با اختلاف طول و عرض کمتر از ۳۵ پیکسل و ارتفاع بزرگتر از ۲۰۰ پیکسل؛ (۳) بلاک با مساحت بزرگتر از ۴۵۰ و اختلاف طول و عرض کمتر از ۷ پیکسل؛ (۴) بلاک با مساحت بیشتر از ۵۰۰۰۰ پیکسل. شکل ۱۴ نتیجه حذف در سطح کلمه را نشان می‌دهد.

$$\text{EdgeDensity} = \frac{\sum \text{EdgeLength}(\text{Word}_i)}{\text{TotalWordAreaPixels}} \quad (۴)$$

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

حذف سطح کاراکتر: بعد از حذف در سطح کلمه، بلاک کلمات به بلاک‌های کاراکتری تقسیم و پس از حذف در سطح کاراکتر، بلاک‌های نهایی در سطح کلمه تشکیل می‌شوند. معمولاً متون در تصویر به صورت کاراکتر تنها ظاهر نمی‌شوند و دارای نویز و اشیاء پراکنده هستند. در نظر گرفتن "مساحت قابل قبول" برای کاراکتر، معیاری جهت تایید کاراکتر و حذف نویزها و اشیاء پراکنده است. مساحت‌های بلوک‌های مجرد غیرمتن برای حذف در مجموعه ICDAR عبارت است از: (۱) مساحت کمتر از ۲۵۰۰؛ (۲) مساحت کمتر از ۶۰۰۰ که طولآن بیشتر از عرض‌اش باشد. شرط اصلی وجود بلوک مجرد عدم‌وجود بلاک در اطراف آن است. شکل ۱۵ نتیجه حذف در سطح کاراکتر را نشان می‌دهد.

حذف سطح خط متن: 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 جهت حذف تشخیص‌های اشتباه از ویژگی خط متن مبتنی بر مفهوم اسکلت دارای راستا استفاده می‌شود (شکل ۱۶). در ابتدا اسکلت هر مولفه، استخراج و با تشخیص نقاط انتهایی، بلاک‌های غیرمتن با معیار رابطه (۶) یعنی نسبت طول اسکلت به فاصله دو نقطه ابتدا-انتها حذف می‌شوند. برای متن، طول اسکلت مقداری نزدیک به فاصلهٔ خط مستقیم ولی برای غیرمتن بزرگتر از خط مستقیم است (شکل ۱۷).

$$\text{StraightRate} = \frac{\sum \text{EskeltonLength}(S_i)}{\text{TwoEndPointsLength}}$$
$$\text{StraightRate} = \begin{cases} > 1 & \text{TextLine} \\ \leq 1 & \text{Non} - \text{TextLine} \end{cases} \quad (۶)$$

تشکیل بلاک‌های نهایی: پس از حذف در سطح خط متن و کلمه به صورت جداگانه، نتایج آنها با هم ادغام و بلاک‌های نهایی در سطح خط متن و در سطح کلمه ایجاد می‌شوند (شکل ۱۸).

۴- نتایج آزمایشات تجربی

در این بخش به بررسی کارآیی روش ارایه شده می‌پردازیم و نتایج آزمایشات را در پنج بخش تنظیمات آزمایشات، مجموعه‌های داده‌ای، آخرین روش‌های علمی برای مقایسه نتایج، معیارهای ارزیابی نتایج، و تحلیل نتایج تجربی بر روی تصاویر دوربین و ویدیویی ارایه می‌کنیم.

۴-۱- تنظیم آزمایشات و مجموعه داده‌ها

ما مدل و الگوریتم پیشنهادی و الگوریتم‌های مختلف مورد مقایسه را در ابزار Matlab و در ماشین Core i7-2.79GHz با 4GRAM و سیستم عامل Win7 پیاده‌سازی و ارزیابی کرده‌ایم. از دو مجموعه استاندارد ICDAR2003 [۲۲]، ICDAR2005 [۲۳] و دو مجموعه جمع‌آوری شده ZAREI-1 و ZAREI-2 و دو نوع تصاویر دوربین ثابت و ویدیویی (۱۰۶۵ فریم ویدیویی و ۲۵۱ تصویر دوربین) برای آزمون استفاده می‌شود. در تصاویر ثابت، ۵۰۹ تصویر از مجموعه استاندارد ICDAR2003 شامل ۲۵۱ تصویر متن صحنه 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 حقیقت پایه است که محدودهٔ دقیق کلمات موجود در هر تصویر به صورت جداگانه و برای تمامی تصاویر مشخص و برای ارزیابی نتایج به‌کار می‌رود. برای ارزیابی بهتر روش پیشنهادی در تصاویر ویدیویی از دو مجموعه ZAREI-1 و ZAREI-2 استفاده می‌شود. مجموعه ZAREI-1 شامل تصاویر ویدیویی در ابعاد ۴۸۰×۶۴۰، ۶۰۰×۸۰۰، ۷۲۰×۱۲۸۰ و ۷۶۸×۱۳۶۶ دارای ۲۰۰ فریم متون غیرافقی (شامل ۱۶۰ فریم متن صحنه و ۴۰ فریم متن گرافیکی)، ۸۰۰ فریم متون افقی (شامل ۱۶۰ فریم متن چینی و کره‌ای، ۱۰۰ فریم متن فارسی و ۳۰۰ فریم متن گرافیکی انگلیسی) است. همچنین دارای ۴ فیلم با زیرنویس فارسی و ۴ فیلم با زیرنویس انگلیسی است. مجموعه ZAREI-2 در ابعاد متفاوت ۲۴۰×۳۲۰، ۴۸۰×۶۴۰، ۷۲۰×۹۶۰، ۹۶۰×۱۲۸۰ و ۱۲۰۰×۱۶۰۰ برای ارزیابی سرعت به‌کار می‌رود که شامل ۶۶ تصویر، ۲۲ تصویر با پیچیدگی کم، ۲۲ تصویر با پیچیدگی متوسط و۲۲ تصویر با پیچیدگی زیاد زمینه است.

۴-۲- روش‌های علمی مورد مقایسه

برای مقایسه نتایج روش پیشنهادی را با روش [۸] اعتبارسنجی و سپس روش‌های مهم [۸]، Laplacian [۲۱]، MSER [۲۴]، DCT [۲۵] و [۳۹] را در کنار روش‌های موردنیاز دیگر پیاده‌سازی و بر مبنای داده‌های استاندارد، نتایج موردنیاز را استخراج کرده‌ایم. یک روش از روش‌های فوق مبتنی بر بافت و چهار روش مهم دیگر مبتنی بر مولفه‌های همبند هستند.

۴-۳- معیارهای ارزیابی

معیارهای ارزیابی موقعیت‌یابی در تصاویر به دو دسته معیارهای ارزیابی در سطح خط متن و معیارهای ارزیابی در سطح کلمه تقسیم می‌شوند که در زیر به‌شرح آن می‌پردازیم.

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

- TDB: تشخیص صحیح بلاک[۲۱]، بلاک با حداقل یک کاراکتر که ممکن است شامل هیچ متنی نباشد.
- FDB: تشخیص غلط بلاک[۲۲]، این بلاک شامل هیچ متنی نیست.
- MDB: بلاک ناقص[۲۳]، بلاکی که بیش از ۲۰٪ کاراکترهای متن را تشکیل نمی‌دهد. بلاکی که حداقل ۸۰٪ متن را تشکیل دهد مورد تایید است [۳۰].

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

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

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

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

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

معیارهای کلمه و کاراکتر: مجموعه‌ای از مستطیل‌هایی که توسط هر الگوریتم تخمین‌زده می‌شود، با مستطیل‌های حقیقت پایه موجود در مجموعه ICDAR یعنی ملاک درستی تشخیص مقایسه می‌شوند. میزان تطبیق m بین این دو مستطیل (مساحت تقاطع دو مستطیل برابر با مینیمم مستطیل شامل هر دو) با بیشترین ارزش است. از این رو 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 به‌دست می‌آید که در آن‌ها E و T به‌ترتیب مستطیل‌های حقیقت پایه و مستطیل‌های تخمین زده شده هستند.

جدول ۲- مقایسهٔ نتایج روش‌های مهم در سطح خط متن

Method	R	P	F	MDR
Proposed Method	0.90	0.85	0.87	0.13
GVF 2013 [8]	0.91	0.76	0.83	0.13
Bayesian [43]	0.87	0.72	0.78	0.14
Laplacian [21]	0.86	0.76	0.81	0.13
Zhou et al. [45]	0.66	0.83	0.73	0.26
Fourier-RGB [46]	0.80	0.66	0.72	0.04
Liu et al. [47]	0.53	0.61	0.57	0.24
Wong & Chen [3]	0.52	0.83	0.64	0.08
Cai et al. [4]	0.67	0.33	0.44	0.43
Yang et al. [39]	0.77	0.85	0.83	0.13

جدول ۳- مقایسهٔ با نتایج روش‌های مهم در سطح کلمه

Methods	R	P	F
Proposed Method	0.70	0.74	0.71
GVF 2013 [8]	0.42	0.36	0.35
MSER [24]	0.60	0.73	0.66
Hinner Becker [23]	0.67	0.62	0.62
Alex Chen [23]	0.60	0.60	0.58
Qiang Zhu [23]	0.40	0.33	0.33

f رابطه (۱۴)، ترکیبی از دو معیار دقت و فراخوانی با وزن نسبی پارامتر α (معمولاً $\alpha = 0.5$) تعیین می‌شود. همچنین معیار میانگین زمان پردازش[۲۴] (APT) رابطه (۱۶) برای مقایسه سرعت روش‌ها به‌ازای k بار اجرای T_i در تصاویر با پیچیدگی‌های مختلف به‌کار می‌رود.

جدول ۴- مقایسهٔ روش پیشنهادی با روش‌های مهم در تصاویر ویدیویی

Method	Language	R%	P%	F%	MDR	APT
Laplacian	Persian	98	62	75	6%	36.1
	English	98	64	77	5%	35.2
MSER	Persian	86	52	64	14%	10.1
	English	89	58	70	12%	9.8
DCT	Persian	84	50	62	34%	1.1
	English	86	73	78	31%	1.1
ProposedMethod	Persian	94	91	92	8%	4.4
	English	98	95	96	5%	4.3

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

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

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

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

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

۴-۴- تحلیل نتایج تجربی

تحلیل نتایج تجربی در دو بخش برای تصاویر ثابت دوربین و تصاویر ویدیویی با زیرنویس فارسی و انگلیسی بیان می‌شود. نمونه موقعیت‌یابی روش پیشنهادی در شکل‌های ۱۹، ۲۰، ۲۱ و ۲۲ ارایه شده است. شکل ۲۱ مبین تصاویر طبیعی دارای چالش‌های مختلف مانند تنوع زبان و فونت، نورپردازی، ابعاد متون، کیفیت و رنگ، پیچیدگی زمینه و کنتراست تصویر است که روش پیشنهادی موقعیت‌یابی آن را دقیق انجام داده است.

شکل ۱۹- نتایج موقعیت‌یابی روش پیشنهادی در تصاویر با زیرنویس فارسی و انگلیسی

نتایج تصاویر دوربین: طبق نتایج جدول ۲ روش پیشنهادی در مجموعه‌های ICDAR2003 و ICDAR2003 از نرخ فراخوانی ۰/۹۰، دقت ۰/۸۵، شاخص F ۰/۸۷ و MDR برابر با ۰/۱۳ برخوردار و نسبت به نتایج روش [۸] دارای دومین نرخ فراخوانی و MDR، بالاترین نرخ دقت و شاخص F است. طبق نتایج جدول ۳ روش پیشنهادی درمجموعه ICDAR2003 و ICDAR2005 در سطح کلمه دارای نرخ فراخوانی ۰/۷۰، دقت ۰/۷۴ و شاخص F برابر با ۰/۷۱ یعنی بالاترین نرخ مبتنی بر فراخوانی، دقت وشاخص F است.

شکل ۲۰- موقعیت‌یابی روش پیشنهادی با چالش‌های نورپردازی ناهموار، سطوح غیرمسطح، تنوع فونت، پیچیدگی زمینه و تغییر رنگ فونت در تصاویر متنوع طبیعی

شکل ۲۱- نمونه تصاویر موقعیت‌یابی تسط روش پیشنهادی

شکل ۲۲- نتایج روش‌های مکان‌یابی (الف) تصویر اصلی (ب) روش Laplacian (ج) روش MSER (د) روش DCT (ه) روش پیشنهادی

نتایج تصاویر ویدیویی: از ویژگی‌های بارز مجموعه‌های ICDAR2003 و ICDAR2003 وجود تنوع متون و زمینه‌های تصاویر است به‌طوری که تمامی چالش‌های مربوط به مکان‌یابی متن را می‌توان در این مجموعه یافت. شکل ۱۹ مبین موقعیت‌یابی در تصاویر ویدئویی با زیرنویس فارسی و انگلیسی در مجموعه‌های ZAREI-1 و ZAREI-2 و موید عملکرد عالی روش پیشنهادی در شرایط مختلف است. در شکل ۲۰ برخی تصاویر چالش‌برانگیز این مجموعه و نتایج روش پیشنهادی ارائه شده است. در جدول ۴ نتایج مقایسهٔ روش پیشنهادی با سه

روش Laplacian [۲۱]، MSER [۲۴] و DCT [۲۵] در مجموعه دادهٔ ZAREI-1 آمده است. مطابق جدول ۴ و در مجموعه ZAREI-1 و ZAREI-2 در تصاویر با زیرنویس فارسی روش پیشنهادی توانسته از بالاترین نرخ دقت و شاخص F و دومین نرخ فراخوانی و MDR برخوردار باشد. روش مبتنی برباقت یعنی روش DCT سریعتر از روش پیشنهادی عمل می‌کند ولی دارای نرخ MDR بالا و نرخ دقت، فراخوانی و شاخص F کمتری است. روش Laplacian [۸] گرچه از نرخ فراخوانی بالاتری نسبت به‌روش پیشنهادی برخوردار است ولی بسیار کندتر عمل می‌کند و به‌دلیل تشخیص‌های اشتباهِ بیشتر دارای نرخ دقت و شاخص F کمتری است. در تصاویر با زیرنویس انگلیسی روش پیشنهادی دارای نرخ فراخوانی و MDR یکسانی با روش Laplacian است ولی سریعتر عمل می‌کند و نرخ دقت و شاخص F بالاتری نسبت به آن دارد. در این‌گونه تصاویر نیز روش DCT سریعتر ولی دارای نرخ MDR بالا و نرخ دقت، فراخوانی و شاخص F کمتری است. شکل ۲۲ نتایج روش‌ها در تصاویر با زیرنویس فارسی در مجموعه‌های ZAREI-1 و ZAREI-2 را نشان می‌دهد که روش پیشنهادی توانسته بهتر از سه روش دیگر به خوبی موقعیت‌یابی کند.

۵- نتیجه‌گیری

مشکلات ارایه روش عمومی موثر 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، رزلوشن تصویر، پیچیدگی زبان متن، و تنوع زبان برای برخورداری از سرعت و دقت مناسب دارد. روشی برای موقعیت‌یابی متن ارایه دادیم که نسبت به مشکلات فوق مقاوم و در شرایط سطح متن غیرمسطح و دارای نورپردازی غیریکنواخت به‌خوبی عمل می‌کند. نتایج بسیار مناسب از آزمایشات تجربی برای تصاویر طبیعی متنوع حاصل شده است. نتایج قابل توجه بر روی متون تصاویر دوربین و ویدیو دارای زبان‌های مختلف دارد و به‌خوبی و با سرعت بالا برای کاربردهای محیط واقعی مانند لوگوها و تابلوها عمل می‌کند.از مشکلات این روش فقدان تشخیص همزمان متون با زوایای مختلف در تصاویر و عدم مقیاس‌پذیری آن در راستای داده‌های تصویری عظیم است. اولین هدف ما پشتیبانی از تنوع داده‌ها در ابعاد مختلف برای رفع بسیاری از چالش‌ها تصویر بوده است ولی روش ما برای کاربرد داده‌های عظیم تصویری هنوز مقیاس‌پذیر نیست و فقط بخش تنوع داده‌ها را پوشش می‌دهد. در کارهای آینده ضمن افزایش سرعت، بنا داریم که آن را برای داده‌های بزرگ و کاربردهای خاص مانند متون پیچیده با جهت‌های مختلف ارایه دهیم.

مراجع

[1] Z. Yingying, C. Yao, and X. Bai, "Scene text detection and recognition: Recent advances and future trends," Frontiers of Computer Science, 2015.

[2] L. Rainer, and A. Wernicke, "Localizing and segmenting text in images and videos," Circuits and Systems for Video Technology, IEEE Transactions, 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 Image Processing, International Conference,vol. 1, pp. I-117, 2002.

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

[6] A. Marios, B. Gatos, and 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," In Document Analysis and Recognition (ICDAR), IEEE International Conference, 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, IEEE Transactions, 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, IEEE Transactions,vol. 20, no. 3, pp. 800-813, 2011.

[10] J. Anil, and B. Yu. "Automatic text location in images and video frames." Pattern recognition, vo. 31, no. 12, pp. 2055-2076, 1998.

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

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

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

[14] H. Weilin, Z. Lin, J. Yang, J. Wang, "Text localization in natural images using stroke feature transform and text covariance descriptors," Computer Vision (ICCV), 2013.

[15] N. Tatiana, O. Barinova, P. Kohli, and V. Lempitsky, "Large-lexicon attribute-consistent text recognition in natural images," InComputer 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, IEEE Transactions, 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

[31] W. Tao, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," In Pattern Recognition (ICPR), 21st International Conference, no. 012, pp. 3304-3308.

[32] J. Max, K. Simonyan, A. Vedaldi, and 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," In 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, 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," Frontiers of Computer Vision (FCV), 21st Korea-Japan Joint Workshop IEEE, 2015.

[37] T. Kobchaisawat, and H. C. Thanarat, "A method for multi-oriented Thai text localization in natural scene images using Convolutional Neural Network," Signal and Image Processing Applications (ICSIPA), EEE International Conference, pp. 220-225, 2015.

[38] D. Karatzas, and et. al., "Icdar 2015 competition on robust reading," 13th International Conference on Document Analysis and Recognition (ICDAR) IEEE, Tunis, Tunisia, pp. 1156–1160, 2015.

[39] Z. Yang, et al.,"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," Journal of Electronic Imaging, vol. 26, no. 1, 2017.

[41] A. A. Ben, and et. al., "MapReduce Based Text Detection in Big Data Natural Scene Videos," Procedia Computer Science,vol. 53, pp. 216-223, 2015.

[42] A. Sana, and et. al., "A Review on Text Detection Techniques," VFAST Transactions on Software Engineering, vol. 8, no. 2, 2015.

[43] S. Palaiahnakote, 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, IEEE Transactions, vol. 22, no. 8, pp. 1227-1235, 2012.

[44] L. Neumann, and J. Matas, "Real-time lexicon-free scene text localization and recognition," IEEE Transactions Machine Intelligence, IEEE Transactions, vol. 36, no. 5, pp. 970-983, 2014.

[18] W. John, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 2, pp. 210-227, 2009.

[19] E. Michael, and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," Image Processing, IEEE Transactions, 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. Lim Tan, "A laplacian approach to multi-oriented text detection in video," Pattern Analysis and Machine Intelligence, IEEE Transactions, 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." ICDAR, 2003.

[23] L. Simon, "ICDAR 2005 text locating competition results," Document Analysis and Recognition, Proceedings, IEEE Eighth International Conference, pp. 80-84, 2005.

[24] C. Huizhong, 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," In Image Processing (ICIP), 2011 18th IEEE International Conference on, pp. 2609-2612, 2011.

[25] L. Su, and K. E. Barner, "Weighted DCT coefficient based text detection," In Acoustics, Speech and Signal ProcessingICASSP 2008, IEEE International Conference, 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," In Image Processing Proceedings, International Conference, 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.

Recognition, CVPR 2004, Proceedings of the 2004 IEEE Computer Society Conference, vol. 2, pp. II-366, 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.

امین‌اله مه‌آبادی تحصیلات خود را در رشته مهندسی برق سخت‌افزار و معماری کامپیوتر به انجام رسانده و اکنون استادیار گروه مهندسی کامپیوتر و فناوری اطلاعات دانشگاه شاهد است. تحقیقات مورد علاقه نامبرده پردازش تصویر، حمل و نقل هوشمند و شبیه‌سازی هوشمند است.

آدرس پست‌الکترونیکی ایشان عبارت است از:

mahabadi@shahed.ac.ir

عیرضا زارعی لیسانس خود را در رشته مهندسی کامپیوتر نرم‌افزار از دانشگاه پیام‌نور شیراز و فوق‌لیسانس خود را در رشته معماری کامپیوتر از دانشگاه شاهد اخذ کرده است. تحقیقات مورد علاقه نامبرده پردازش تصویر و بینایی ماشین است.

آدرس پست‌الکترونیکی ایشان عبارت است از:

a.zarei1600@gmail.com

اطلاعات بررسی مقاله:

تاریخ ارسال: ۱۳۹۶/۰۱/۳۱

تاریخ اصلاح: ۱۳۹۶/۰۳/۰۴

تاریخ قبول شدن: ۱۳۹۶/۰۴/۲۲

نویسنده مرتبط: دکتر امین‌اله مه‌آبادی، دانشکده فنی و مهندسی، دانشگاه شاهد، تهران، ایران.

[1]Text Detection (TD)
[2]Text Localization (TL)
[3]Optical Character Recognition (OCR)
[4]Image Framing (IF)
[5]Big Data Images (BDI)
[6]Texture-Based Methods (TBM)
[7]Connected Component-Based Methods (CCM)
[8]Hybrid Methods (HM)
[9]Statistical
[10]Structural and Spectral (SS)
[11]Image Serach Space (ISS)
[12]Stroke Width (SW)
[13]Stroke Width Transform (SWT)
[14]Maximally Stable External Regions (MSER)
[15]Straightness
[16]False Positive (FP)
[17]Gradient Vector Flow (GVF)
[18]Edge Block (EB)
[19]Eccentricity
[20]Strength
[21]Truly Detected Block (TDB)
[22]Falsely Detected Block (FDB)
[23]Text Block with Missing Data (TBMD)
[24]Average Processing Time (APT)

on Pattern Analysis and Machine Intelligence, vol. 38, no. 9, pp. 1872-85, 2016.

[45] Z. Jingchao, L. Xu, B. Xiao, R. Dai, and S. Si. "A robust system for text extraction in video," In Machine Vision, 2007. ICMV 2007. IEEE International Conference on, 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, IEEE Transactions,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," In Document Analysis and Recognition, Proceedings. Eighth International Conference, 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," In Computer Vision (ICCV), IEEE International Conference, pp. 1241-1248, 2013.

[49] W. Huang, Q. Yu, and X. Tang, "Robust scene text detection with convolution neural network induced mser trees," In Computer Vision–ECCV, pp. 497-511, 2014.

[50] Z. Yu, K. Karu, and A. K. Jain, "Locating text in complex color images," In Document Analysis and Recognition, Proceedings of the Third International Conference, vol. 1, pp.146-149, 1995.

[51] L. Huiping, D. Doermann, and O. Kia, "Automatic text detection and tracking in digital video," Image Processing, IEEE Transactions, 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, IEEE Transactions, 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, IEEE Transactions,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, IEEE Transactions,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," Springer Berlin Heidelberg, 2010.

[57] C. Xiangrong, and A. L. Yuille, "Detecting and reading text in natural scenes," In Computer Vision and Pattern