آشکارسازی و تعیین مکان متون فارسی-عربی در تصاویر ویدئویی

محی‌الدین مرادی، سعید مظفری و علی اصغر اروجی
دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران

چکیده

استخراج اطلاعات متنی از تصاویر ویدئویی، در کاربردهایی نظیر تحلیل معنایی ویدئو، بازیابی اطلاعات متنی و بازیابی اطلاعات مربوط به تصاویر ویدئویی آرشیوشده نقش مهمی دارد. در این مقاله، روشی جهت استخراج متن فارسی- عربی از تصاویر ویدئویی ارائه می‌شود. در ابتدا، جهت کاهش حساسیت به اندازه قلم، بر اساس تصویر ورودی هرمی دارای سه وضوح ایجاد می‌شود. سپس با استفاده از آشکارساز لبه، لبه‌های موجود در تصویر استخراج و با یافتن مکان تلاقی لبه‌ها، تصویرگوشه تصنعی ایجاد می‌شود. جهت حذف گوشه‌های تصنعی نواحی غیرمتنی، تحلیل هیستوگرام انجام می‌شود. ضرایب تبدیل فوریه کسینوسی گسسته بلوک‌های تصویر استخراج و با ترکیب تعدادی از آنها، تصویر شدت بافت حاصل می‌شود. در ادامه جهت جداسازی نواحی متنی از غیرمتنی، با یک طبقه‌بندی کننده، با تلفیق مشخصه‌های حاصل از تصویرگوشه تصنعی و تصویر شدت بافت، برداری متشکل از مشخصه‌ها تشکیل می‌شود. در پایان، با رسم نمودارهای هنجاری‌سازی شدهٔ شدت بافت، بازبینی نهایی انجام شده و تفکیک خطوط متنی انجام می‌شود.

واژگان کلیدی: متون فارسی- عربی، ویدئو، آشکارسازی متن، گوشه تصنعی، تبدیل فوریه کسینوسی گسسته، طبقه بندی SVM.

۱- مقدّمه

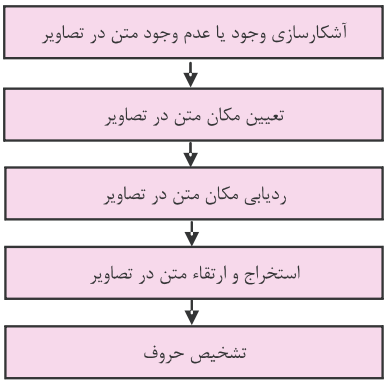
در میان محتویات معنایی موجود در تصاویر ویدئویی، اطلاعات متنی اهمیت فراوانی دارد. آشکارسازی و استخراج متن از تصاویر ویدئویی، یکی از موضوعات چالش‌برانگیز، معروف و شناخته‌شده در تحقیقات بینایی ماشین و تشخیص الگوست به‌طوری که امروزه محققان فراوانی در این زمینه مشغول به فعالیت هستند. با توجه به تحقیقات فراوانی که درخصوص بینایی ماشین و تشخیص الگو انجام شده است، تکامل‌یافته‌های موجود در بازیابی محتویات از داخل تصاویر ویدئویی، موضوعی ضروری است. امروزه بسیاری از تصاویر ویدئویی حاوی اطلاعات متنی شامل متن مرتبط با صدای ویدئو، ترجمه متن، نتایج مسابقات ورزشی یا تلویزیونی، وضعیت بورس یا آب و هوا، اخبار فوری و موارد مشابه آنها هستند. یکی از کاربردهای پرطرفدار اطلاعات متنی، اضافه‌کردن اطلاعات متنی به تصویر، جهت اندیس‌گذاری و جستجوی خودکار متن براساس اندیس‌های موجود در ویدئوست. این موضوع نیاز به استخراج متن از تصاویر ویدئویی را به‌طور چشم‌گیری افزایش داده است. در این تحقیق سعی شده ره‌یافت مناسبی برای استخراج متن فارسی ارائه شود.

متن می‌تواند با سبک‌ها، اندازه‌ها و جهت‌های گوناگون در تصویر قرارگیرد. از این رو جداسازی متن از پس‌زمینه می‌تواند امری چالش‌برانگیز باشد. از جمله خصوصیات مورد استفاده جهت آشکارسازی متن، می‌توان به نحوهٔ قرارگیری متن در تصویر، میزان بسامد نویسه‌ها در نواحی متنی، همبستگی مکانی میان نویسه‌ها و توزیع رنگ در نواحی متنی اشاره کرد. علاوه‌بر مشکلات موجود در استخرج متن از تصاویر ویدئویی، دستگاه‌های نوری شناخت نویسه نیز محدویت‌هایی دارند [Jung, et al., 2004].

۲- بررسی مراحل مختلف استخراج متن از تصاویر ویدئویی

به‌طور کلی سامانه‌های استخراج متن طبقات مختلفی دارند که عنوان هر یک از این طبقات همراه با نحوه ارتباط میان آنها با یکدیگر در شکل (۱) نمایش داده شده است [Jung, et al., 2004].

در مرحلهٔ آشکارسازی، وجود یا عدم وجود متن در تصویر از طریق روش‌های متعددی مشخص می‌شود. در مرحله تعیین مکان متن، مکان متن در داخل تصویر مشخص شده و در اطراف آن کادری کشیده می‌شود. در بسیاری از موارد، مرحلهٔ آشکارسازی و تعیین مکان به‌صورت هم‌زمان با یکدیگر صورت می‌گیرند. در مرحلهٔ استخراج و ارتقای متن، اجزای متن از بخش پس‌زمینهٔ تصویر جداشده و تصویر، قطعه‌قطعه می‌شود. نواحی متن به‌طور معمول وضوح پایینی دارند، لذا در معرض نوفه قرار داشته و لازم است ارتقا داده شوند. در این تحقیق، هدف معرفی روشی درخصوص مراحل آشکارسازی و تعیین مکان متن فارسی در تصاویر ویدئویی است و لذا موضوعات شامل مراحل ردیابی مکان متن، استخراج و ارتقای متن و تشخیص حروف در این نوشتار مد نظر نخواهند بود.



(شکل ۱)- مراحل مورد نیاز جهت استخراج متن از تصاویر ویدئویی

۳- مروری بر تحقیقات انجام‌شده جهت استخراج متن از تصاویر ویدئویی

الگوریتم‌هایی که جهت استخراج متن از تصاویر ویدئویی ارائه شده‌اند، همگی نواقصی دارند که این مطلب ناشی از وجود تغییرات گستردهٔ متن و تصاویر ویدئویی است. از جملهٔ این موارد می‌توان به پایین‌بودن تفاوت شدت روشنایی در تصاویر، پیچیدگی زیاد پس‌زمینه در برخی از تصاویر، تغییرات در اندازه، رنگ و جهت قرارگیری متن در تصاویر اشاره کرد [Jung, et al., 2004,Liang, et al., 2005, Zhang, et al., 2008].

در گذشته، تلاش‌های متعددی جهت استخراج متن از تصاویر ویدئویی انجام گرفته است. برخی از خصوصیات متنی استفاده‌شده جهت استخراج نواحی متنی از تصاویر ویدئویی عبارتند از: رنگ و سطح خاکستری نواحی متنی و پس‌زمینه، اختلاف در شدّت روشنایی و تباین نواحی متنی و پس‌زمینه، تفاوت ساختاری موجود در بافت تصویر، جهت قرارگیری متن در تصویر و ابعاد هندسی ناحیه متنی.

در مرجع [Xiaoqian, et al., 2012] روش مبتنی بر استخراج لبه، جهت افزایش کارایی و دقت سامانه پیشنهادی از تصحیح محتوای فریم‌های متوالی، تعین موقعیت زمانی-مکانی و نیز جداسازی زیرنویس‌های موجود در تصویر بهره می‌گیرد. در [Zhao, et al., 2011] یک سامانه خودکار آشکارسازی متن براساس نقاط گوشه که یکی از ویژگی‌های مهم نواحی متنی است، استفاده شده است. بلوک‌های متنی می‌توانند از طریق تغییر رنگ میان نواحی متنی و پس‌زمینهٔ مجاور آن تشخیص داده شوند. نواحی دارای متن می‌توانند به‌صورت تقریبی از طریق محاسبه چگالی پیکسل‌های در حال تغییر و پیکسل‌های پایدار مجاور مورد شناسایی قرار گیرند [Wonjun and Changick., 2009]. در مرجع [Shivakumara, et al., 2011] متن مصنوعی اضافه‌شده به تصویر و متن ذاتی موجود در تصویر قابل تشخیص است. در استخراج متن ذاتی موجود در تصویر به‌دلیل آنکه اغلب سامانه‌های موجود بر متن افقی تأکید دارند تمرکز سامانه بر جهت متن استوار است. در حوزهٔ فرکانس، لاپلاسین به تشخیص نواحی متنی کمک کرده و تحلیل اسکلت جهت قسمت بندی اجزا به هم متصل شده پیچیده به بخش‌های ساده‌تر مورد استفاده قرار می‌گیرند. در مرجع [Li, et al., 2011] روش استخراج متن بر اساس نقاط کلیدی متن ارائه شده است. نقاط کلیدی متن به‌عنوان نقاطی تعریف می‌شوند که هم‌زمان بافت قوی در جهات عمودی، افقی و قطری دارند. نقاط کلیدی متن از طریق سه باند فرعی فرکانس بالای تبدیل موجک استخراج گشته و افزایش دقت و کارایی را در پی خواهند داشت.

به‌طور کلی می‌توان گفت در تحقیقات انجام شده جهت آشکارسازی متن در تصاویر از دو گروه الگوریتم آشکارسازی با عناوین آشکارسازی بر اساس ناحیه و

آشکارسازی بر اساس بافت تصویر استفاده شده است [Jung, et al., 2004, Lienhart and Wernicke., 2002 , Kim, et al., 2003].

۳- ۱- آشکارسازی متن بر اساس ناحیه

در روش آشکارسازی بر اساس ناحیه از مشخّصه‌های مربوط به نواحی متنی استفاده می‌شود. از جملهٔ این مشخّصات می‌توان به سطح خاکستری نواحی متن و پس‌زمینه اشاره کرد. در این روش می‌توان از طریق تفاوت مشخّصه‌های نواحی متنی و پس‌زمینه، متن را استخراج کرد. در این روش در ابتدا لبه‌ها و اجزای به‌هم متصل شده استخراج می‌شوند. در مرحلهٔ بعد، این ساختارها با یکدیگر گروه بزرگ‌تری را تشکیل داده و در نتیجه، بلوکی از یک ناحیهٔ متن را مشخص می‌کنند. این روش خود به دو صورت قابل پیاده‌سازی است [Jung, et al., 2004,Liang, et al., 2005, Zhang, et al., 2008].

الف- استخراج متن بر اساس لبه‌ها

این روش بر اساس وجود تباین بالا میان نواحی متن و پس‌زمینه استوار است و از این طریق، امکان جداسازی متن از پس‌زمینه فراهم می‌شود. در این روش، ابتدا لبه‌های موجود در تصویر از طریق یک آشکارساز لبه، همانند Canny، استخراج می‌شوند و نواحی غیر متنی بر اساس دسته‌ای قوانین، از نواحی متنی جدا خواهند شد. در ادامه، با استفاده از اُپراتورهای ریخت‌شناسی[1] و یا فیلترهای هموارکننده، لبه‌های تأییدشده با یکدیگر ادغام می‌شوند. کارایی در این روش به مقدار قابل توجهی، به آشکارساز لبه بستگی دارد. در برخی مواقع لبه‌های نواحی متن و پس‌زمینه به‌راحتی قابل تشخیص نبوده و درنتیجه الگوریتم‌های مبتنی بر لبه، فقط در کاربردهای مشخّصی قابل استفاده هستند [Xiaoqian, et al., 2012, Zhao, et al., 2011].

ب- استخراج متن بر اساس اجزای به‌هم متصل شده

در این روش متن به‌صورت مجموعه‌ای از اجزای مستقل به‌هم متصل‌شده درنظر گرفته می‌شوند. در این وضعیت هر یک از اجزا دارای شدت روشنایی، توزیع رنگ و منحنی بسته مشخصی خواهند بود. با گروه‌بندی اجزای کوچک به اجزای بزرگ‌تر، همهٔ نواحی تصویر مشخص می‌شوند. این روش‌ها به‌طور عمومی از سه مرحلهٔ زیر تشکیل می‌شوند [Jung, et al., 2004,Liang, et al., 2005, Zhang, et al., 2008, Wonjun and Changick., 2009]:

۱- استخراج اجزای به‌هم متصل‌شده

۲- تحلیل اجزای به‌هم متصل‌شده جهت تشخیص متن یا غیر متن بودن آنها

۳- پردازش نهایی جهت اتصال اجزای گوناگون متن به یکدیگر (کلمات و خطوط)

۳-۲- آشکارسازی متن بر اساس بافت تصویر

در این روش، با فرض وجود بافت ویژه در نواحی متنی، از طریق تحلیل‌های مبتنی بر بافت تصویر، جداسازی نواحی متن از پس‌زمینه امکان‌پذیر است. در الگوریتم‌های آشکارسازی مبتنی بر بافت تصویر، مقادیر پیکسل‌ها بررسی می‌شوند و چنانچه نوفه روی مشخصه‌های بافت پیکسل‌های مجاور اثری نداشته باشد، آنگاه اثر نوفهٔ پس‌زمینه در آشکارسازی نواحی متنی حذف خواهد شد. در طراحی طبقه‌بندی‌کننده‌های بافت تصویر، از آموزش‌های آماری استفاده می‌شود. از این رو سامانه دارای مقاومت بالاتری بوده و امکان تعمیم‌دهی آن به وجود می‌آید. با این وجود عمده‌ترین مسئله در این‌گونه الگوریتم‌ها، پیچیدگی محاسباتی است؛ زیرا تمامی نواحی تصویر تحت جستجو قرارگرفته و تحیل بافت تصویر زمان زیادی نیازخواهد داشت. نگرانی دیگری که در این روش وجود دارد مربوط به دقت انتخاب نواحی متنی است. با توجه به اینکه تحلیل بافت، بر اساس همسایگی پیکسل‌ها عمل می‌کند، امکان دستیابی به دقت موجود در روش‌های مبتنی بر اجزای به‌هم متصل شده وجود نخواهد داشت.

۴- طبقه بندی خصوصیات کلی متن در زبان‌های مختلف

در حالت کلی خصوصیات متن به دو دسته تقسیم می‌شوند. دستهٔ اوّل شامل خصوصیاتی است که مستقل از نوع زبان‌اند. دسته دوم شامل خصوصیاتی است که از زبانی به زبان دیگر متفاوت خواهند بود[Jung, et al., 2004, Cai, et al., 2002].

۴-۱- خصوصیات مستقل از نوع زبان

الف- تباین: نواحی متنی در مقایسه با نواحی غیر متنی نسبت به ناحیهٔ پس‌زمینهٔ خود دارای تباین قابل توجهی دارند. میزان این تباین، با توجه به میزان پیچیدگی ناحیهٔ پس‌زمینه، متغیر است. چنانچه ناحیهٔ پس‌زمینه ناحیه

[1] Morphological Operators

ساده‌ای باشد، در آن‌صورت حتی متن دارای تباین پایین نیز قابل شناسایی است. چنانچه ناحیه پس‌زمینه دارای بافت پیچیده‌ای باشد، فقط نواحی متنی دارای تباین بالا قابل تشخیص‌اند.

ب- رنگ: در اغلب جملات، رنگ نویسه‌های متنی مشابه است و تغییر نمی‌کند. با این وجود در فرآیند فشرده‌سازی با اتلاف، در لبه‌های متن و پس‌زمینه مقداری تداخل رنگ ایجاد می‌شود.

پ- متحرک یا ثابت‌بودن متن در تصویر: متن در تصویر جایگاه ثابت یا متحرکی دارد. اغلب پیام‌های بازرگانی و اخبار فوری و مواردی امثال آن به‌صورت زیرنویس متحرک در پایین تصویر نوشته می‌شوند. این موضوع سبب افزایش اهمیت فرآیند ردیابی متن خواهد شد. جهت حرکت متن با توجه به نوع زبان به‌کار رفته می‌تواند تغییر کند. به‌عنوان مثال در زبان فارسی به‌طور عمومی متن از طرف چپ به طرف راست و یا از پایین به طرف بالا حرکت می‌کند.

۴-۲- خصوصیات وابسته به زبان

الف- چگالی تکانه‌های موجود در متن: تکانه‌های ایجاد شده توسط یک قلم، حرکات یا علامت‌هایی هستند که در هنگام نوشتن توسط قلم ایجاد می‌شوند. تکانه درواقع چهارچوب یک نوشته را نمایش می‌دهد. در زبان انگلیسی و فارسی چگالی تکانه‌های موجود در نواحی مختلف متنی به‌طور معمول مقدار یکسانی است. در زبان‌هایی همانند کره‌ای و چینی، این چگالی می‌تواند در نواحی مختلف متن تفاوت داشته باشد. برای مثال اگر چه در زبان چینی برخی حروف فضای یکسانی را اشغال می‌کنند، ولی تعداد تکانه‌های موجود در آنها با یکدیگر تفاوت دارد.

ب- کوچک‌ترین اندازه قلم به‌کاررفته در متن: در تصویر جهت نمایش تمامی تکانه‌های موجود در یک متن لازم است با توجه به چگالی تکانه‌های موجود در حروف، کوچک‌ترین اندازه قلم قابل استفاده تعیین شود. به‌عنوان نمونه نویسه‌های چینی به‌دلیل چگالی بالای تکانه‌های موجود در آنها، در مقایسه با نویسه‌های انگلیسی و فارسی، فضای بیشتری را اشغال کرده و به همین دلیل این نویسه‌ها باید با اندازه قلم بزرگ‌تری نمایش داده شوند. چنانچه بخواهیم روش‌های استخراج متن مبتنی بر تکانهٔ زبان‌های چینی، کره‌ای و انگلیسی را به زبان فارسی اعمال کنیم در آن‌صورت به‌دلیل این تفاوت نتایج مطلوبی حاصل نمی‌شود.

پ- کوچک‌ترین نسبت ابعاد کلمات: زبان چینی زبانی نویسه محور است. به‌عبارت دیگر در این زبان یک نویسه به‌تنهایی می‌تواند یک مفهوم مستقل داشته باشد. زبان‌های فارسی و انگلیسی بر اساس کلمات می‌باشند و برای ایجاد یک کلمه لازم است چندین نویسه به یکدیگر متصل شوند. از این رو کوچک‌ترین نسبت ابعاد کلمات با معنی مستقل، در این زبان‌ها با یکدیگر متفاوت است.

ت- توزیع آماری تکانه‌ها در جهات مختلف: تکانه‌ها در حروف انگلیسی عمدتاً در جهت عمودی قرار دارند و به همین دلیل بسیاری از الگوریتم‌های استخراج متن بر اساس تکانه‌های عمودی استوارند. در زبان‌های فارسی و چینی این موضوع متفاوت است و به‌همین دلیل الگوریتم‌هایی که بر این اساس برای زبان انگلیسی تهیه شده‌اند، برای زبان‌های فارسی و چینی قابل اعمال نیستند.

ث- جهت قرارگیری متن در تصویر: متن در تصاویر باید به‌گونه‌ای قرارگیرد تا در کنار محتویات دیگر موجود در ویدئو به آسانی برای بیننده قابل درک باشد. از این رو در تصاویر ویدئویی به‌طور معمول متن فارسی به‌صورت افقی نوشته شده و لذا کافی است برای زبان‌هایی همانند فارسی و انگلیسی فقط متن افقی مورد توجه قرارگیرد.

۵- اهمیت زبان فارسی و ویژگی‌های آن

نویسه‌های زبان فارسی در چندین زبان دیگر از جمله عربی، اردو و پشتو مورد استفاده قرارمی‌گیرند. درواقع حدود نیم‌میلیارد از مردم کره زمین با این حروف سر و کار دارند. زبان عربی، زبان به‌کار رفته در قرآن بوده و اغلب مسلمانان، که جمعیتی در حدود یک‌چهارم جمعیت کل کرهٔ زمین را تشکیل می‌دهند، توانایی خواندن زبان عربی را دارند. با این وجود هنوز آشکارسازی متن فارسی از تصاویر ویدئویی مورد توجه جدی قرار نگرفته است. برخی از خصوصیات متن فارسی عبارتند از:

الف- نویسه‌های خط فارسی روی یک خط فرضی معروف به خط مبنا به‌صورت شکسته و متصل به یکدیگر نوشته می‌شوند. به‌دلیل اتصال حروف به یکدیگر، بلوکی از نویسه‌ها می‌تواند تشکیل شود. در زبان فارسی حروف کوچک و بزرگ همانند آنچه برای زبان انگلیسی در نظر گرفته می‌شود، وجود ندارد.

ب- نویسه‌های فارسی با توجه به مکان قرارگیری آنها در کلمه می‌توانند بیش از یک شکل داشته باشند. با توجه به

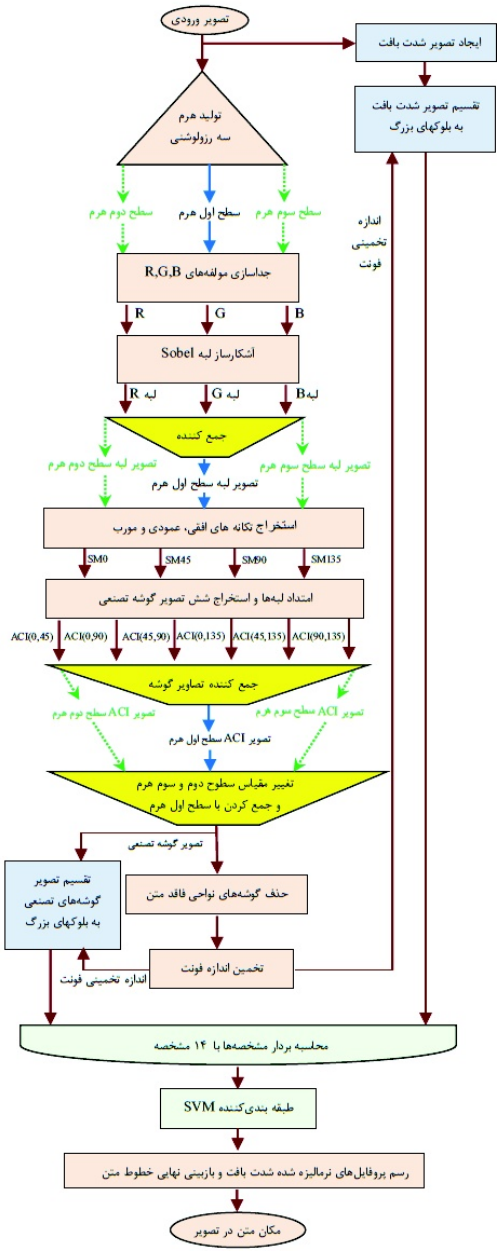
مفهوم بلوک نویسه‌ها، یک نویسه می‌تواند در ابتدا، وسط، انتها و یا در حالت تنها قرار بگیرد.

پ- برخی از قلم‌ها در زبان فارسی رایج‌تر بوده و در کتب، مقالات و نامه‌های رسمی از آنها استفاده می‌شود. این قلم‌ها عبارتند از Lotus ، Mitra ، Nazanin ، Traffiic ، Yaghut، Zar ، Homa، Titr [Khosravi and Kabir., 2010].

ت- مستقل از نوع و سبک قلم، برای یک اندازه قلم مشخص، ارتفاع همه قلم‌ها به‌طور تقریبی یکسان است. برای مثال چنانچه قلم‌های متداول ده‌گانه فارسی با اندازهٔ بیست را درنظر بگیریم، در یک تصویر پال ارتفاع همه آنها به‌طور تقریبی برابر ۱۷ تا ۲۳ پیکسل خواهد بود. از این رو سعی می‌کنیم تصویر را بر اساس ارتفاع قلم هنجارسازی کنیم. این امر ناشی از این واقعیت است که هر اندازه قلمی را می‌توان با یک ارتفاع معادل‌سازی کرد. در اغلب متون فارسی موجود در تصاویر ویدئویی، اندازه قلم به‌طور معمول میان ۱۵ تا ۲۵ متغیر است. از این رو به‌عنوان یک فرض منطقی اولیه، می‌توان اندازه قلم موجود در تصویر را برابر ۲۰ درنظر گرفت.

۶- طرح پیشنهادی

در شکل (۲) روندنمای طرح پیشنهادی نشان داده شده است. جهت کاهش حساسیت طرح استخراج متن به تغییر اندازه قلم، در ابتدا هرم سه وضوحی تصویر ورودی ایجاد می‌شود (بخش ۶-۱). سپس جهت استخراج لبه‌ها با استفاده از آشکارساز لبه، چهار طرح تکانه[1] (SM) افقی، عمودی و مورب تولید می‌شوند (بخش ۶-۲). به‌منظور استخراج گوشه‌های نواحی متنی، با استخراج گوشه‌های تصنعی سطوح مختلف هرم، تصویر گوشهٔ تصنعی[2] ACI، شده (بخش ۶-۳) و با رسم هیستوگرام افقی هنجارسازی‌شده و اعمال محدودیت‌های ابتکاری، اندازه واقعی قلم به‌کمک تصویر گوشه تصنعی تخمین زده می‌شود (بخش ۶-۴). در ادامه به‌منظور شناسایی بافت نواحی مختلف تصویر، از روی تصویر ورودی تصویر شدت بافت ایجاد می‌شود (بخش ۶-۵). سپس بردار مشخصه‌ها محاسبه شده و جهت جداسازی نواحی متن، از طبقه‌بندی‌کنندهٔ SVM[3] استفاده می‌شود (بخش ۶-۶). در نهایت با رسم نمودارهای افقی و عمودی، مکان دقیق متن در تصویر ویدئویی مشخص خواهدشد (بخش ۶-۷). در این روش، در هر ثانیه فقط دو فریم مورد بررسی قرارمی‌گیرند. این انتخاب بر اساس حداقل زمانی که مغز جهت رؤیت و درک کامل یک تصویر پیچیده لازم دارد صورت گرفته است (۲ تا ۳ ثانیه) [Leon and Gasulll., 2005].

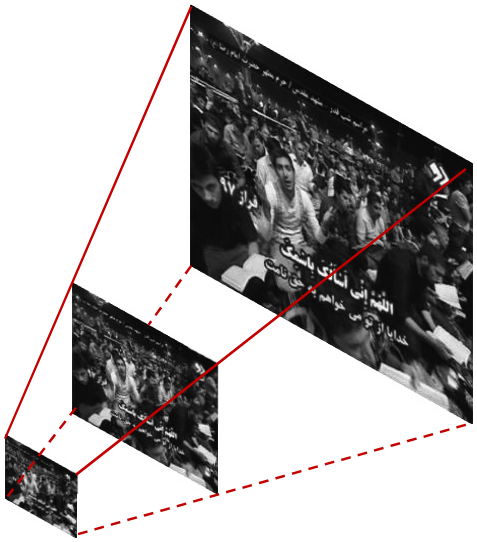


(شکل ۲)- طرح پیشنهادی جهت آشکارسازی و تعیین مکان متن فارسی- عربی از تصاویر ویدئویی

[1] Stroke Map
[2] Artificial Corner Image
[3] Support Vector Machine

۶-۱ ایجاد هرم سه وضوحی جهت کاهش حساسیت آلگوریتم به تغییرات اندازه قلم

یکی از مشکلات موجود در موضوع استخراج متن از تصاویر ویدئویی، تغییر اندازه قلم متن موجود در تصاویر است. چنانچه بتوان در ابتدا اندازه قلم را مشخص کرد آنگاه آشکارسازی متن با دقت بیشتری انجام می‌شود. متأسفانه تاکنون هیچ روش مؤثّری در این خصوص ارائه نشده است. جهت کاهش حساسیت آلگوریتم پیشنهادی به تغییر اندازه قلم و رفع خطای احتمالی ناشی از اندازه قلم پیش‌فرض بیست تصویر ورودی از طریق یک فیلتر گوسی، به‌صورت متوالی فیلتر می‌شود تا یک هرم سه وضوحی مرتبط با آن ایجاد شود شکل (۳). در ساخت این هرم، وضوح در هر مرحله به نصف حالت پیشین خود کاهش داده می‌شود. همان‌گونه که در ادامه شرح داده خواهدشد، تمامی پردازش‌های لازم جهت استخراج گوشه‌های تصنعی، روی هر سه سطح هرم انجام شده و با افزودن گوشه‌های تصنعی سطوح مختلف به یکدیگر، یک تصویر گوشه حاصل خواهد شد.



(شکل ۳)- هرم سه وضوحی تصویر ورودی

۶-۲ آشکارسازی لبه‌ها

یکی از مشخّصه‌های مهم نواحی متنی در تصویر لبه است و می‌تواند جهت استخراج متن از تصاویر ویدئویی مورد استفاده قرارگیرد. متون فارسی به‌طور معمول تکانه‌هایی در جهت افقی، عمودی و مورب دارند. از این رو چنانچه ناحیه‌ای از تصویر دارای چگالی بالایی از لبه در این جهات باشد (به‌ویژه افقی و عمودی) آنگاه به احتمال زیاد این ناحیه دارای متن خواهد بود. درصورت استخراج لبه از روی تصویر خاکستری، در برخی از تصاویر دارای متون رنگی، با خطای عدم تشخیص لبه‌های متنی از پس‌زمینه مواجه می‌شویم. لذا جهت غلبه بر مشکلات ناشی از رنگی‌بودن تصویر، تصویر به سه مؤلّفهٔ رنگ‌های سبز، آبی و قرمز تفکیک شده و سه اپراتور آشکارساز لبه Sobel به تصاویر مؤلّفه‌های رنگی اعمال می‌شوند. سپس با ادغام لبه‌های موجود در سه مؤلّفهٔ رنگ، یک تصویر لبه ایجاد می‌گردد شکل (۴- ب).

(الف)

(ب)

(شکل ۴)- الف: تصویر ورودی ب: تصویر لبه

با توجه به ویژگی‌هایی اپراتور آشکارساز لبه Sobel، در در روش پیشنهادی از این آشکار ساز لبه استفاده می‌شود. این اپراتور دارای مشخصه ایزوتروپیک بوده و درنتیجه پاسخ یک‌نواختی را در تمامی جهات ایجاد می‌کند. لبه‌های آشکار شده توسط این اپراتور دارای چگالی بالایی بوده و سبب

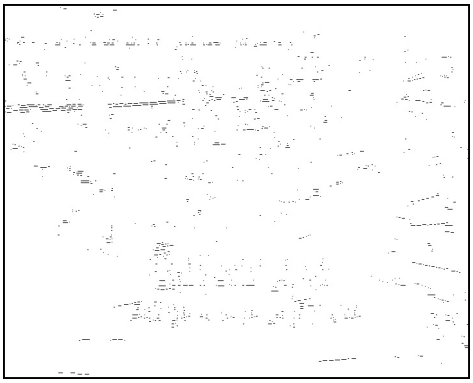
افزایش لبه در نواحی متنی خواهد شد. با توجه به عدم آشکارسازی تمامی لبه‌ها توسط این اپراتور، نواحی دارای تباین پایین مربوط به پس‌زمینه، به‌راحتی توسط فرآیند آستانه‌گذاری حذف شده و تشخیص لبه‌های مربوط به نواحی متنی آسان‌تر انجام می‌شود. با توجه به ویژگی‌های بیان شده درخصوص متن فارسی و دلایل انتخاب اندازه قلم پیش فرض ۲۰، مجموعه‌ای متشکل از هشتاد تصویر ویدئویی شامل قلم‌های متنوع با اندازهٔ قلم حدود بیست‌پیکسل از تصاویر موجود در پایگاه داده، که در بخش (۷) به آن پرداخته خواهد شد، تهیه شد. اگرچه می‌توان جهت افزایش دقت و حصول به نتایج مطلوب‌تر تعداد تصاویر ویدئویی را بیش‌تر انتخاب کرد امّا به‌دلیل وجود اغلب پارامترهای متغیر مربوط به متن در تصاویر انتخاب شده و نیز افزایش سرعت پیاده‌سازی در مرحلهٔ آموزش طبقه‌بندی کننده SVM، انتخاب هشتاد تصویر کافی و مناسب خواهد بود. جهت حذف لبه‌های مربوط به نواحی غیر متنی و تفکیک لبه‌های نواحی متنی در جهت‌های افقی، عمودی، مورب ۴۵ و ۱۳۵ درجه، به‌ترتیب لبه‌های بزرگ‌تر از ۱۵، ۱۸، ۸، ۱۰ پیکسل و کوچک‌تر از سه‌پیکسل از تصاویر لبه حذف می‌شوند. این مقادیر طی یک فرایند تجربی مشخص شده‌اند که در زیر به آن خواهیم پرداخت. در ابتدا با تخصیص مقادیر پیش فرض مختلف به لبه‌های بزرگ و کوچک، این مقادیر را روی هشتاد تصویر ویدئویی جمع آوری شده با اندازهٔ قلم بیست‌پیکسل اعمال می‌نماییم. سپس با استفاده از مشاهده نتایج، بهترین مقادیری که می‌توانند لبه‌های مربوط به متن را حفظ کرده و لبه‌های بزرگ و کوچک مربوط به نواحی غیر متن را حذف نمایند، مشخص می‌کنیم. نتایج به‌گونه‌ای است که به‌ترتیب برای جهت‌های افقی، عمودی، مورب ۴۵ و ۱۳۵ درجه چنانچه لبه‌ای بزرگ‌تر از ۱۵، ۱۸، ۸، ۱۰ پیکسل و کوچک‌تر از سه‌پیکسل باشد آنگاه این لبه مربوط به متن نخواهد بود. درنهایت با انجام این پردازش چهار طرح تکانه (SM) حاصل خواهد شد که در هر یک فقط یکی از لبه‌های افقی، عمودی، مورب ۴۵ درجه و مورب ۱۳۵ درجه حضور دارند شکل (۵).

۶-۳ آشنایی با مفهوم گوشه تصنعی و استفاده از آن در تشخیص نواحی متنی

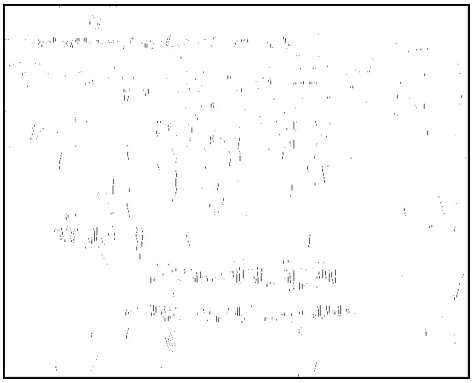
با توجه به پایین‌بودن چگالی تکانه‌ها در زبان فارسی، چنانچه جهت استخراج متن فارسی الگوریتم‌های ارائه شده درخصوص استخراج متون چینی و یا انگلیسی به‌کار روند، امکان آشکارسازی دقیق متن فارسی وجود نخواهد داشت. به‌عنوان اولین تلاش جهت رفع مشکل و افزایش چگالی تکانه‌ها در نواحی متنی، می‌توان تکانه‌ها را امتداد داد.

از این رو با توجه به اندازه قلم پایه، هر یک از تکانه‌های ۰ ، ۹۰ ، ۴۵ و ۱۳۵درجه از طریق یک اپراتورگشایش با المان ساختاری دارای طول بیست‌پیکسل، در جهت مربوطه امتداد داده می‌شوند شکل (۶). این پردازش مشکل کم‌بودن چگالی تکانه‌ها در نواحی متنی را مرتفع می‌نماید، ولی سبب افزایش لبه‌های اضافی در نواحی غیر متنی خواهد شد. برای غلبه بر این مشکل مفهوم گوشهٔ تصنعی معرفی می‌شود. محل برخورد دو تکانه امتداد داده شده را یک گوشهٔ تصنعی می‌نامیم. صفت تصنعی به این گوشه‌ها ناشی از این واقعیت است که این گوشه‌ها، از طریق تکانه‌های امتداد داده‌شده حاصل می‌شوند و نه از طریق محل تلاقی تکانه‌های واقعی موجود در تصویر. انتخاب المان ساختاری با طول بیست‌پیکسل جهت اپراتورگشایش بر این اساس است که با توجه به اینکه اندازهٔ قلم پایه برابر بیست در نظر گرفته شده است لذا با استفاده از اپراتورگشایش با المان ساختاری دارای طول بیست‌پیکسل، می‌توان تکانه‌های مختلف داخل یک کلمه و نیز تکانه‌های کلمات مختلف داخل یک سطر از متن فارسی را به مقدار کافی امتداد داده تا از این طریق علاوه‌بر افزایش چگالی تکانه، امتداد تکانه‌ها با یکدیگر برخورد کرده و گوشه‌های مصنوعی ایجاد شوند. چنانچه اثر امتداددهی تکانه‌ها در روند رشد تعداد گوشه‌های تصنعی مورد بررسی قرارگیرند، نتایج حاکی از رشد چشم‌گیر گوشه‌های تصنعی در نواحی متن، در مقایسه با نواحی پس‌زمینه است. در توجیه این پدیده کافی است به تغییرات گرادیان لبه‌ها در یک تصویر توجه شود. یکی از ویژگی‌های نواحی متنی، تغییرات شبه پریودیک گرادیان لبه افقی است. از این رو با افزایش طول لبه‌های موجود در تصویر، در نواحی متنی نسبت به نواحی غیر متنی تغییرات شبه‌گرادیان لبه در امتداد افقی، رشد بیشتری خواهد داشت. به‌طور کلی یک ناحیه متنی دارای ترکیب شبه منظمی از لبه‌های افقی، عمودی و مورب است. لذا با افزایش طول تکانه‌ها، نرخ رشد گوشه‌های تصنعی نواحی متنی در اغلب موارد بیشتر خواهد بود.

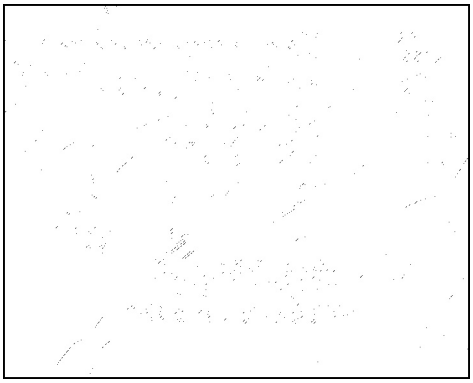
پس از استخراج لبه‌ها در چهار جهت اصلی، از محل تلاقی هر دو لبه متفاوت، مطابق رابطهٔ زیر یک تصویر گوشهٔ تصنعی(ACI) ایجاد می‌شود.



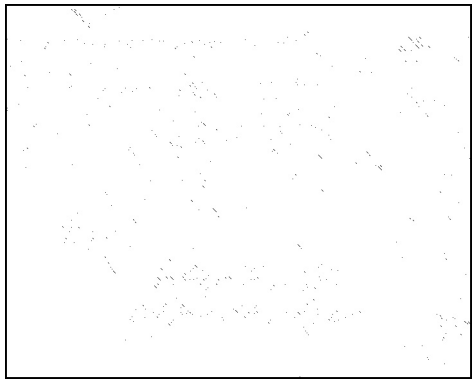
(الف)



(ب)

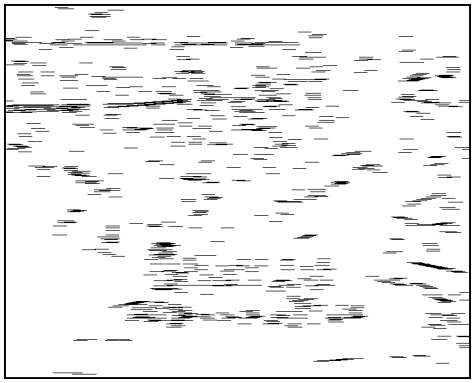


(پ)

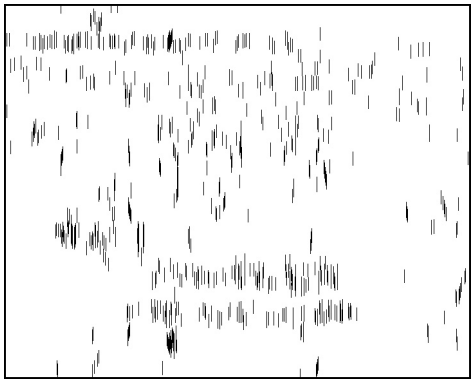


(ت)

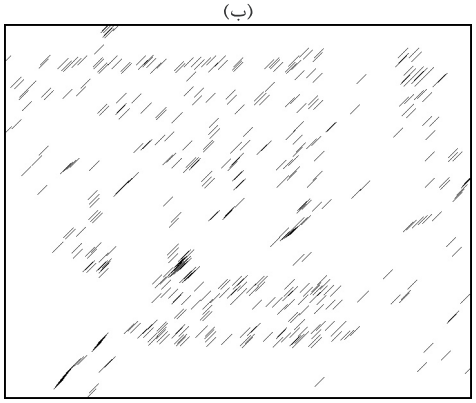
(شکل ۵)- تصاویر طرح تکانه ۰، ۹۰، ۴۵ و۱۳۵ درجه



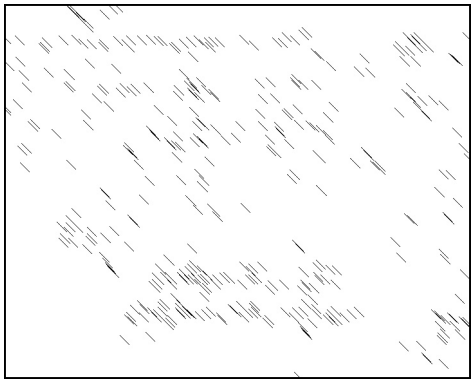
(الف)



(ب)



(پ)

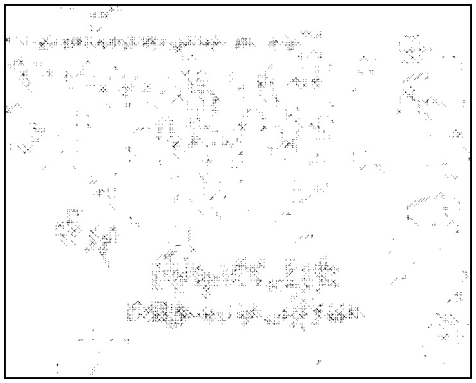


(ت)

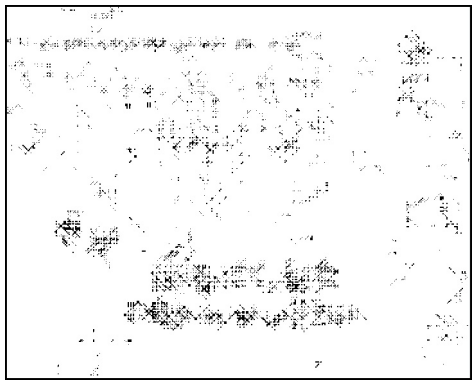
(شکل ۶) - تصاویر لبه ۰، ۹۰، ۴۵ و۱۳۵ درجه امتداد داده شده

$$ACI\,x,y = AND\left(SMx, SMy\right) \qquad x,y = 0,45,90,135 \quad x \neq y \tag{۱}$$

با افزودن شش‌گوشه تصنعی به یکدیگر، یک تصویر گوشهٔ تصنعی نهایی حاصل خواهد شد (شکل۷- الف).



(الف)



(ب)

(شکل ۷)- الف- تصویرگوشه تصنعی سطح اول هرم
ب- تصویر گوشه تصنعی حاصل از سه تصویر موجود در هرم سه وضوحی

پس از استخراج گوشه‌های تصنعی موجود در سطوح دوم و سوم هرم، تصاویر گوشه سطوح دوم و سوم هرم، جهت رسیدن به اندازه‌ای برابر با اندازه سطح اول هرم، تغییر مقیاس داده می‌شوند. گوشه‌های تصنعی تصاویر گوشه تغییر مقیاس یافته، به تصویر گوشه مربوط به اولین سطح هرم اضافه شده و تصویر گوشهٔ حاصل‌شده دارای این ویژگی است که تا حدّ زیادی می‌تواند سبب کاهش وابستگی روش پیشنهادی به تغییرات در اندازهٔ قلم شود شکل (۷- ب).

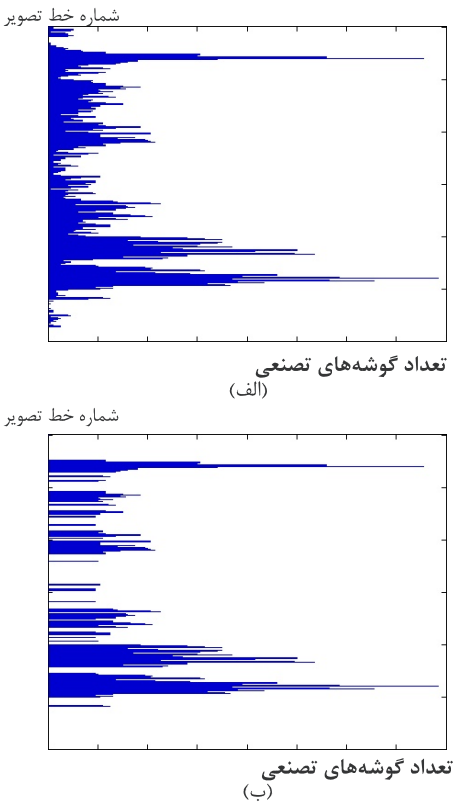
۶-۴ تخمین اندازه واقعی قلم

با محاسبهٔ تعداد گوشه‌های تصنعی موجود در هر خط از تصویر گوشهٔ تصنعی، هیستوگرام افقی مربوط به تصویرگوشهٔ تصنعی از طریق رابطه زیر قابل ترسیم خواهد بود (شکل ۸- الف).

$$H(i) = \sum_{j} ACI(i,j) \tag{۲}$$

پس از رسم هیستوگرام افقی، جهت حذف گوشه‌های نواحی غیر متنی و تخمین اندازهٔ قلم، از میانگین ستون‌های هیستوگرام،m ، استفاده می‌شود.

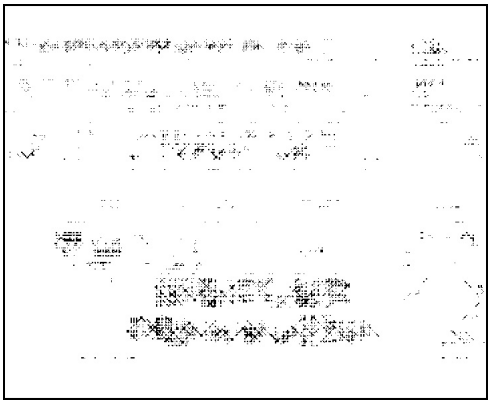
$$m = \frac{\sum_{i} H(i,j)}{Num(H(i))} \tag{۳}$$



(شکل ۸)- الف- هیستوگرام تصویرگوشه تصنعی
ب- حذف گوشه‌های تصنعی نظیر نواحی غیر متنی

با حذف ستون‌هایی از هیستوگرام که از m کمتر هستند، بسیاری از ستون‌های مربوط به نواحی غیرمتنی حذف خواهند شد شکل (۸- ب). سپس گوشه‌های نظیر این ستون‌ها از تصویر گوشه تصنعی حذف می‌شوند شکل (۹). در توجیه استفاده از مقدار میانگین جهت حذف گوشه‌های

نواحی غیر متنی، کافیست به این نکته توجه شود که به‌طور معمول چگالی گوشه‌ها در نواحی متنی نسبت به میانگین آن در کل تصویر گوشه بیشتر است. میانگین عرض ستون‌های باقی‌مانده در هیستوگرام را محاسبه کرده و مقدار به‌دست آمده، تخمینی از مقدار واقعی قلم موجود در متن خواهد بود. چنانچه این تخمین با مقدار قلم پایه بیش از پنجاه‌درصد اختلاف داشته باشد، آنگاه اندازه قلم پایه با مقدار به‌دست‌آمده از مرحلهٔ تخمین، جایگزین می‌شود.



(شکل ۹)- تصویر گوشه تصنعی حاصل از حذف گوشه‌های نظیر ستونهای باریک در هیستوگرام افقی

چنانچه اندازهٔ قلم از مقدار نصف ارتفاع تصویر بیشتر شود، آنگاه در محاسبه اندازهٔ قلم خطا رخ داده است و یا متنی در تصویر وجود ندارد. در این حالت، جهت استخراج متن احتمالی در تصویر، با فرض اندازه قلمی برابر قلم پایه، ادامهٔ پردازش‌ها جهت آشکارسازی متن انجام خواهند شد.

۶-۵ ایجاد تصویر شدّت بافت

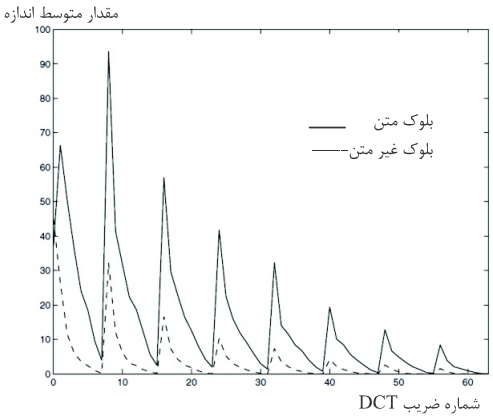
تصویر شدّت بافت، تصویری است که بافت نواحی مختلف تصویر را مشخص می‌کند. یکی از مزایای استفاده از بافت تصویر، تشخیص نواحی متنی با قلم کوچک است. روش استفاده از تبدیل فوریه کسینوسی گسسته(DCT)[1] با ابعاد ۸×۸، یکی از روش‌هایی است که در ساخت تصویر شدت بافت مورد استفاده قرار می‌گیرد. ضرایب DCT مربوط به یک بلوک ۸×۸ تصویر f(x,y) با رابطهٔ زیر بیان می‌شوند[Lim, et al., 2000, Zhong, et al., 1999].

$$AC_{uv} = \frac{1}{8} C_u C_v \times \sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y)\cos\frac{(2x+1)u\pi}{16}\cos\frac{(2y+1)v\pi}{16}$$

$$C_u, C_v = \begin{cases} 1 & u,v \neq 0 \\ \frac{1}{\sqrt{2}} & u,v = 0 \end{cases} \qquad (۴)$$

در رابطهٔ بالا u و v مختصات عمودی و افقی را نمایش داده و مقادیر آنها میان ۰ تا ۷ تغییر می‌کند. انتخاب مجموعه ضرایبی که بتوانند میان نواحی متنی و غیرمتنی تصاویر ویدئویی تمایز قائل کرده و نواحی دارای متن را آشکارسازی نمایند، می‌تواند از طریق یک فرایند تجربی انجام گیرد. با توجه به وجود ۶۴ ضریب برای هر بلوک، تعداد کل ترکیبات موجود برابر $\sum_{i=1}^{64}\binom{64}{i} \cong 1.8\times10^{19}$ می‌باشند.

درنتیجه، با توجه به تعدد ترکیبات موجود انتخاب مناسب ضرایب از طریق یک فرایند تجربی کار آسانی نخواهد بود. یکی از روش‌های تسهیل‌کننده جهت انجام آسان‌تر فرآیند تجربی فوق، استفاده از روشی است که جزئیات کامل آن در [Crandall, et al., 2003] ذکر شده است. به‌طور خلاصه این روش مبتنی بر محاسبهٔ مقدار متوسط اندازه هر یک از ۶۴ ضریب، در نواحی متنی و غیر متنی است. مقدار متوسط اندازه ضرایب در نواحی متنی و غیر متنی برای تعدادی از تصاویر موجود در پایگاه داده در شکل (۱۰) نشان داده شده است.



(شکل ۱۰)- مقدار متوسط اندازه ضرایب برای نواحی متنی و غیر متنی

همان‌گونه که در این شکل نشان داده شده است، برخی از ضرایب اختلاف بیشتری را برای نواحی متنی و غیر متنی نشان می‌دهند. با رسم این نمودار برای هشتاد تصویر

[1] Descrite Cosine Transform

ویدئویی شامل قلم‌های متنوع با اندازهٔ قلم حدود بیست پیکسل از تصاویر موجود در پایگاه داده ، ضرایب AC را که در آنها میان مقدار متوسط اندازه نواحی متنی و غیر متنی بیشترین تفاوت وجود دارد انتخاب می کنیم. به عبارت دیگر، پس از اعمال تبدیل فوریه کسینوسی گسسته روی بلوکهای ۸×۸ تصویر خاکستری ورودی، جهت مؤثربودن ضرایب DCT در تشخیص نواحی متن فارسی، از میان ۶۴ ضریب ایجاد شده برای هر بلوک، ضرایبی که به بهترین صورت، نشان‌گر تغییر فرکانس مکانی مورب، افقی و عمودی نواحی دارای متن، انتخاب می‌شوند. نتایج تجربی حاکی از وجود نوزده ضریب با ویژگی فوق خواهد بود. با فرض T(i,j) به عنوان شدّت بافت (i,j) امین بلوک۸×۸، با کمک نوزده ضریب انتخاب شده در مرحله قبل، T(i,j) از طریق رابطهٔ زیر محاسبه می‌شود. در این رابطه اندیس‌ها به شمارهٔ ضریب انتخاب شده اشاره دارند.

$$T(i,j)=k_{00}|AC_{00}|+k_{01}|AC_{01}|+k_{02}|AC_{02}|+k_{03}|AC_{03}| +k_{04}|AC_{04}|+k_{07}|AC_{07}|+k_{10}|AC_{10}|+k_{11}|AC_{11}|+k_{12}|AC_{12}| +k_{13}|AC_{13}|+k_{17}|AC_{17}|+k_{20}|AC_{20}|+k_{21}|AC_{21}|+k_{22}|AC_{22}| +k_{27}|AC_{27}|+k_{30}|AC_{30}|+k_{31}|AC_{31}|+k_{37}|AC_{37}| \quad (۵)$$

در این رابطه $k_{i,j}$ ضرایب وزن‌دهی مربعی هستند که محاسبه آنها از طریق رابطهٔ زیر صورت می‌گیرد. با استفاده از ضرایب $k_{i,j}$ فرکانس‌های مکانی بالاتر که به‌طور معمول به نواحی دارای متن مربوط هستند مورد تاکید خواهند بود [Kenneth and Barnerl., 2008].

$$k_{i,j}=\begin{cases}0 & u=v=0\\ (\frac{u+v}{2})^2 & u+v\geq 1\end{cases} \quad (۶)$$

پس از محاسبهٔ شدت بافت همهٔ بلوک‌های ۸×۸ تصویر ورودی، با کنار هم قرارگیری این مقادیر، تصویر شدّت بافت حاصل می‌شود. جهت مشاهدهٔ بهتر این تصویر با هشت برابر بزرگ‌نمایی در شکل (۱۱) نشان داده شده است.

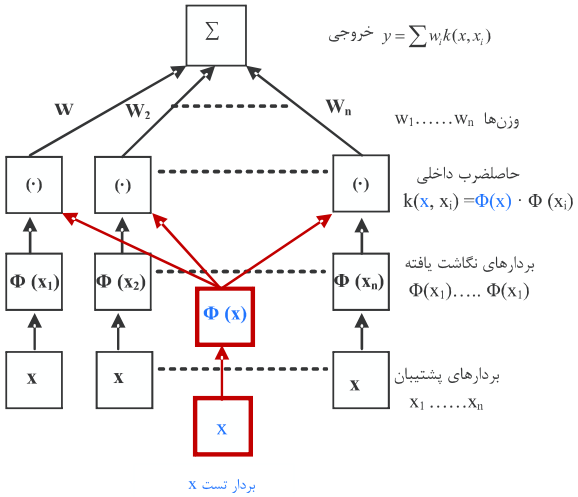
(شکل ۱۱)- تصویر شدّت بافت

۶-۶ ایجاد بردار مشخّصه و آموزش SVM

روش SVM یک روش پیاده‌سازی تقریبی از روش کمینه‌سازی خطرپذیری ساختاری[1] است. دلایل استفاده از SVM آموزش‌دهی آسان، عدم نیاز به تعداد نمونه‌های زیاد و امکان تعمیم‌دهی آسان آن است.

ساختار SVM به‌عنوان آشکارساز متن، در شکل (۱۲) نشان داده شده است. همان‌گونه که در شکل مشخص شده است، این ساختار از سه لایه تشکیل می‌شود. ورودی شبکه از طریق تعدادی از پیکسل‌های تصویر، که با یکدیگر یک بلوک بزرگ را تشکیل می‌دهند، تأمین می‌شود. لایهٔ پنهان، یک نگاشت غیرخطی Φ از فضای ورودی به فضای مشخصه انجام داده و حاصل ضرب نقطه‌ای میان ورودی خود و بردارهای نگاشت یافته را محاسبه می‌کند. در عمل، دو مرحلهٔ بالا با معرفی تابع هسته k می‌تواند در یک گام انجام شوند. این محاسبه به‌صورت حاصل ضرب نقطه‌ای دو الگوی نگاشت داده شده با رابطهٔ $k(x,x_i)=\Phi(x).\Phi(x_i)$ تعریف می‌شود. تابع هسته به‌کار رفته در این مقاله، تابع مبنا شعاعی[2] زیر است که در آن σ بیان‌گر انحراف از استاندارد است.

$$k(x,x_i)=e^{\frac{-|x-x_i|}{2\sigma^2}} \quad (۷)$$



(شکل ۱۲)- ساختار SVM به‌عنوان آشکارساز متن

[1] Structural Risk Minimization
[2] Radial Basis Function

علامت خروجی، y ، توسط وزن‌دهی به خروجی لایهٔ پنهان مشخص می‌شود و وضعیت طبقه را نمایش خواهدداد. جهت آموزش، طبقه متن با ۱+ و طبقه غیر متن با ۱- نمایش داده می‌شود[Shin, et al., 2000].

جهت تشخیص نواحی متنی از نواحی غیر متنی با توجه به اندازهٔ قلم به‌دست آمده در مرحلهٔ تخمین قلم، تصویر ورودی به نواحی کوچک‌تری به‌نام بلوک بزرگ تقسیم می‌شود شکل (۱۳). جهت کاهش خطای احتمالی در تخمین قلم، چنانچه اندازه قلم پایه از چهل کمتر باشد، ابعاد بلوک‌های بزرگ ۸۰×۴۰ و چنانچه اندازه قلم پایه از ۱۲۰ بیش‌تر باشد، ابعاد بلوک‌های بزرگ ۲۴۰×۱۲۰ درنظر گرفته می‌شوند. چنانچه اندازهٔ قلم پایه در محدوده ۴۰ تا ۱۲۰ باشد، آنگاه هر بلوک بزرگ، در جهت عمودی دارای اندازه‌ای برابر با قلم پیکسل و در جهت افقی، دارای اندازه‌ای برابر با دو قلم پیکسل است. جهت استخراج چهارده مشخصه، روابط ۸ تا ۱۴ بر تصویر گوشه و تصویر شدّت بافت اعمال می‌شوند. به‌عبارت دیگر، در این روابط، F یک‌بار با تصویر گوشه و بار دیگر، با تصویر شدّت بافت جایگزین می‌شود. در این روابط، h و w ارتفاع و عرض بلوک بزرگ و $M_2, M_1, \mu, I, H, Et, E$ به‌ترتیب بیان‌گر انرژی، آنتروپی، همسانی، اینرسی، میانگین، گشتاور مرتبهٔ اول و گشتاور مرتبهٔ دوم یک بلوک بزرگ هستند[Shivakumara, et al., 2009].

(شکل ۱۳)- تقسیم بندی تصویر به بلوک‌های بزرگ

$$E = \sum_{i,j\in Bl} F(i,j)^2 \quad (۸)$$

$$Et = \sum_{i,j\in Bl} F(i,j)\log F(i,j) \quad (۹)$$

$$H = \sum_{i,j\in Bl} \frac{1}{1+(i-j)^2} F(i,j) \quad (۱۰)$$

$$I = \sum_{i,j\in Bl} (i-j)^2 F(i,j) \quad (۱۱)$$

$$\mu = \frac{1}{w\times h}\sum_{i,j\in Bl} F(i,j) \quad (۱۲)$$

$$M_1 = \frac{1}{w\times h}\sum_{i,j\in Bl} (F(i,j)-\mu)^2 \quad (۱۳)$$

$$M_2 = \frac{1}{w\times h}\sum_{i,j\in Bl} (F(i,j)-\mu)^3 \quad (۱۴)$$

پس از محاسبهٔ مشخصه‌های فوق، درجمع برای هر بلوک بزرگ، چهارده مشخصه به‌دست می‌آید. مقادیر مربوط به تمامی بردارهای مشخصه میان ۰ و ۱ هنجارسازی شده و با ورود این مشخصه‌ها به یک طبقه‌بندی کننده SVM، متن یا پس‌زمینه بودن هر یک از بلوک‌های بزرگ تعیین می‌شوند. جهت آزمایش روش پیشنهادی، از ۲۸۷۱ قاب تصویر موجود در پایگاه داده استفاده شده است. از این تعداد چهارصد قاب انتخاب شده و پس از تعیین چشمی بلوک‌های بزرگ متنی و غیر متنی، در مرحله آموزش از آنها استفاده شده‌است.

۶-۷ رسم نمودارهای هنجارسازی‌شدهٔ شدت بافت و بازبینی نهایی خطوط متن

با فرض آنکه MTI(R) نمایش‌گر شدت بافت متوسط[1] ناحیه بلوک متنی R باشد شکل (۱۴)، آنگاه مقدار آن به‌صورت زیر محاسبه می‌شود[Qian and Liu ., 2006].

$$MTI(R) = \frac{1}{N(R)}\sum_{i\in R}\sum_{j\in R} T(i,j) \quad (۱۵)$$

در رابطهٔ بالا N(R) تعداد بلوک‌های موجود در ناحیه - R است. چنانچه شدت بافت متوسط به قدر کافی بزرگ باشد، آنگاه ناحیه R نامزد شده، یک ناحیهٔ متنی است. وضعیت ناحیه R توسط خصوصیات بلوک متن و تباین میان ناحیه متن و پس‌زمینه به‌صورت زیر تعیین می‌شود.

$$MTI(R) > \max\{MTI_{LR}, MTI_{UD}, MTI_A\} \quad (۱۶)$$

$$MTI_A = \frac{1}{M\times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} T(i,j) \quad (۱۷)$$

در رابطهٔ بالا M و N تعداد بلوک‌های افقی و عمودی موجود در یک تصویر و MTI_{UD} ,MTI_{LR} ,MTI_A به‌ترتیب میانگین شدّت بافت درکل تصویر، ناحیه واقع در همسایگی بالا - پایین و ناحیه واقع در همسایگی چپ - راست ناحیه

[1] Mean Texture Intensity

R هستند. این موضوع در شکل (۱۴) نشان داده شده است. ناحیهٔ همسایگی R در جهت افقی در هر طرف دارای مقدار W بلوک و در جهت عمودی در هر طرف دارای مقدار H بلوک است. محاسبهٔ مقادیر MTI_{LR} ، MTI_{UD} از طریق رابطهٔ مربوط به محاسبه MTI_A بوده و فقط ناحیهٔ محاسبه متفاوت است.

(شکل۱۴) – کادر احاطه‌کننده یک بلوک متن، همراه با بلوک نواحی همسایگی بالا، پایین، چپ و راست.

به‌طور معمول کادر متن دقیق مربوط به هر خط متن، از طریق نمودارهای افقی و عمودی در حوزهٔ پیکسل با استفاده از چگالی لبه و یا اطلاعات شدت روشنایی پیکسل حاصل می‌شود. در این بخش سعی داریم روش تعیین مکان متن در ویدئو را با استفاده از نمودارهای هنجارسازی شده شدت بافت بلوکی به‌دست آوریم. مثالی از چگونگی استخراج یکی از نواحی بلوک متن در یک تصویر ویدئویی همراه با کادرهای احاطه‌کننده مربوط به آن و نمودارهای هنجارساز افقی و عمودی مربوطه، در شکل (۱۵) نمایش داده شده است. فرض کنیم Hprof(i) نمایش‌گر پروفایل افقی هنجارسازی‌شده شدت بافت خط i ام از ناحیه بلوک متن بازبینی شده باشد شکل (۱۵- ب). پروفایل افقی هنجارسازی‌شده، Hprof(i)، به‌صورت زیر بیان می‌شود.

$$Hprof(i) = \sum_{i \in R} \frac{T(i,j)}{HN(i)} \qquad (۱۸)$$

در این رابطه، HN(i) تعداد بلوک‌های موجود درiأمین ردیف است. به‌طور مشابه، پروفایل عمودی هنجارسازی‌شده شدت بافت (شکل ۱۵- پ) ، Vprof(j)، به‌صورت زیر به‌دست می‌آید که در آن VN(j) تعداد بلوک‌های موجود درjأمین ستون است.

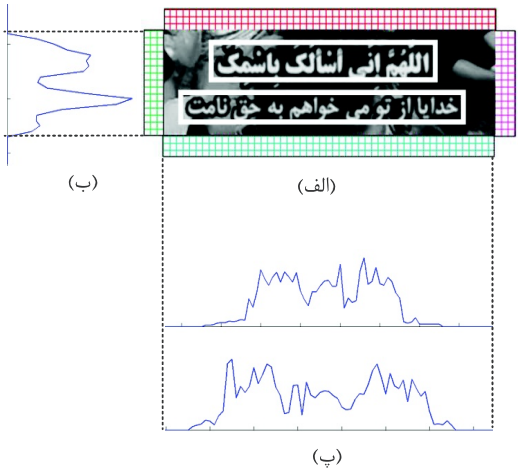
$$Vprof(j) = \sum_{j \in R} \frac{T(i,j)}{VN(j)} \qquad (۱۹)$$

در پروفایل نمایش افقی هنجارسازی‌شده شدّت بافت، از طریق دره‌های میان دو قلهٔ مجاور یکدیگر، به‌طور دقیق می‌توان سطرها را از هم تفکیک و جداسازی کرد. در کادر پیرامون خطوط متنی، برخی از بلوک‌های پس‌زمینه ممکن است در میان خطوط متنی باقی بمانند. جهت حذف نواحی غیر متنی موجود میان خطوط و رسم کادر احاطه‌کننده دقیقتر پیرامون هر یک از خطوط، از دو رابطهٔ تجربی زیر استفاده می‌شود.

$$Hprof(i) > \max\{MTI_{UD}, MTI_A, mean(Hprof(i))\} \qquad (۲۰)$$

$$Vprof(j) > \max\{MTI_{LR}, MTI_A, mean(Vprof(j))\} \qquad (۲۱)$$

چنانچه این روابط برقرار شوند، مفهوم آن متن‌بودن بلوک *(i,j)* أم است. با استفاده از شکل (۱۵) مشاهده می‌شود که اغلب بلوک‌های پس‌زمینه، بطور مؤثّری حذف شده است و کادرهای احاطه‌کننده دقیقی در اطراف متن حاصل می‌شود.



(شکل ۱۵) – الف– ناحیه نامزد متن همراه با خطوط متنی تفکیک‌شده ب– پروفایل افقی هنجارسازی‌شده شدّت بافت پ– پروفایل‌های عمودی هنجارسازی‌شده شدت بافت هر یک از خطوط متنی موجود

۷– مجموعه داده مورد استفاده

در بررسی نتایج حاصل از الگوریتم‌های پردازش تصویر، انتخاب پایگاه داده اهمیت فراوانی دارد. چنانچه انتخاب تصاویر مورد استفاده در پایگاه داده به‌درستی صورت نگیرد، در آن‌صورت نتایج به‌دست آمده چندان قابل اعتماد نخواهند بود. با وجود آنکه موضوع استخراج متن از تصویر ویدئویی موضوع جدیدی نیست؛ ولی جهت استخراج متن فارسی از

تصاویر ویدئویی، پایگاه داده معتبری وجود ندارد. از این رو در این تحقیق، پایگاه داده مناسب، جهت استخراج متن فارسی از تصاویر ویدئویی ارائه می‌شود. با توجه به طیف وسیع فرمت‌های تصاویر ویدئویی و مشخصات تصاویر تشکیل دهندهٔ ویدئو، پایگاه داده ایجاد شده باید تمامی حالات ممکن را پوشش دهد. با توجه به استفاده ایران از سیستم رنگی پال، تمامی تصاویر ویدئویی مورد استفاده در این پایگاه داده به‌صورت پال بوده و دارای ابعاد ۷۲۰ ×۵۷۶ پیکسل هستند. جهت ایجاد این پایگاه داده، از شصت نماهنگ ویدئویی استفاده شده است. پارامترها و نکاتی که جهت ایجاد این پایگاه داده در نظر گرفته شده است در جدول زیر مشخص شده است. برای دست یابی به این متغیرها از تصاویر خبری، تصاویر مسابقات ورزشی، فیلم‌های سینمایی، تصاویر پویانمایی، تصاویر مستند، تصاویر مربوط به آگهی‌های بازرگانی و تصاویری که بر اثر بایگانی طولانی دارای کیفیت پایین باشند استفاده شده است. شایان ذکر است با توجه به اینکه متن موجود در یک تصویر می‌تواند چندین ویژگی داشته باشد؛ لذا مجموع درصدهای موجود در جدول زیر از صددرصد بیش‌تر است.

(جدول ۱)- ویژگی‌های موجود در پایگاه داده و درصد آنها

ویژگی	نوع تصویر مورد نیاز	درصد
روشنایی	تصاویر با روشنایی متفاوت	۱۵
کنتراست	تصاویر دارای تباین کم و زیاد	۱۲
رنگ بودن	تصاویر رنگی	۸۶
فاقد رنگ بودن	تصاویر سیاه و سفید	۱۴
دوران	تصاویر با متون دوران یافته	۱۳
اندازه و سبک قلم به‌کاررفته	تصاویر دارای متن با اندازه و سبک قلم مختلف	۴۷
پیچیدگی پس‌زمینه	تصاویر با پس‌زمینه‌های ساده و پیچیده	۳۲
مقاومت در برابر نوفه	تصاویر دارای نوفه	۴۳
تغییر اندازه و سبک قلم در داخل یک تصویر	تصاویر ویدئویی که در یک صحنه به‌صورت هم‌زمان چندین اندازه و سبک قلم مختلف وجود دارند	۳۸
ثابت و متحرک بودن متن	تصاویر دارای متن ثابت و متحرک، به‌گونه‌ای که متن متحرک در جهات اصلی در تصویر حرکت نماید	۳۱
جهت حرکت متن	استفاده از تصاویری که متن موجود به‌صورت هم‌زمان در دو جهت چپ و راست دارای حرکت باشند	۲۳

۸- پیاده سازی و آزمون

جهت آزمون مقاومت و کارایی روش‌های ارائه شده در آشکارسازی متن از تصاویر ویدئویی، از معیاری نرخ دقت[1] *(P.R)* و نرخ فراخوانی[2] *(R.R)* استفاده می‌شود[1].

نرخ دقت به‌صورت زیر تعریف می‌شود.

$$P.R = \frac{C}{C + FP} \qquad (۲۲)$$

خطای مثبت[3] (FP): چنانچه نواحی غیرمتنی به‌صورت نواحی متنی آشکار شوند، خطایی که در این حالت رخ می‌دهد خطای مثبت است.

در این رابطه تعداد نواحی متنی که به‌درستی به‌عنوان ناحیه متن آشکارشده‌اند با C نمایش داده شده‌است.

نرخ فراخوانی به‌صورت زیر تعریف می‌شود:

$$R.R = \frac{C}{C + FN} \qquad (۲۳)$$

خطای منفی[4] (FN) : چنانچه نواحی متنی به‌صورت نواحی غیر متنی آشکار شوند، خطایی که در این حالت رخ می‌دهد خطای منفی است.

به‌طور معمول میان نرخ فراخوانی و نرخ دقت مصالحه‌ای برقرار است. لذا جهت اندازه‌گیری دقیق‌تر، به‌طور معمول نمرهٔ نرخ فراخوانی و نرخ دقت از طریق فرمول زیر محاسبه می‌شود.

$$Score = \frac{2 \times P.R \times R.R}{P.R + R.R} \qquad (۲۴)$$

در نمره، اثر هر دو پارامتر در نظر گرفته می‌شود و از این رو معیار درست‌تری جهت بررسی الگوریتم‌های استخراج متن از ویدئو است. نتایج حاصل از بررسی تصاویر ویدئویی موجود در پایگاه داده، نشان‌گر نرخ فراخوانی۸۹/۲۳ و نرخ دقت ۸۳/۱۲ هستند. درنتیجه، نمرهٔ حاصل از روش پیشنهادی برابر با ۸۶/۱۰ است. جهت مقایسهٔ روش پیشنهادی با روش‌های دیگر، روش ارائه‌شده در این مقاله را با دو روش ارائه‌شده در مراجع [Moradi and Mozaffaril., 2010,2011]، که توسط مؤلّفان این مقاله درقبل در دو کنفرانس ارائه‌شده است و نیز روش‌های ارائه شده در مراجع [Halima, et al., 2010, Ahmad, et al., 2012] که از جمله نوین‌ترین روش‌های پیشنهادی است مقایسه می‌کنیم. با توجه به این‌که امکان دسترسی به پایگاه دادهٔ این دو روش

[1] Precision Rate
[2] Recall rate
[3] False Positive
[4] False Negative

وجود نداشته و لازم است نتایج براساس پایگاه داده یکسانی گزارش شوند، لذا این دو روش موجود در [Halima, et al., 2010, Ahmad, et al., 2012] پیاده‌سازی شده و نتایج حاصل از پنج روش مذکور با به‌کارگیری پایگاه داده در جدول(۲) نشان داده شده است. شایان ذکر است در مراجع [Halima, et al., 2010, Ahmad, et al., 2012] تأکید بر متون عربی است که آشکارسازی متن عربی به‌دلیل افزایش تکانه‌های ناشی از اعراب، از متن فارسی آسان‌تر خواهد بود.

(جدول ۲)- مقایسه نرخ دقت، نرخ فراخوانی و نمره در روش‌های موجود

	نرخ دقت	نرخ فراخوانی	نمره
Moradi and Mozaffaril., 2010	۷۲/۸	۷۸/۵۴	۷۵/۵۰
Moradi and Mozaffaril., 2011	۸۱/۸۰	۸۷/۵۴	۸۴/۵۷
Halima, et al., 2010	۹۱/۲۳	۹۲/۴۷	۹۱/۴۵
Ahmad, et al., 2012	۹۵	۹۶	۹۵/۴۹
روش ارائه‌شده در این مقاله	۸۳/۱۲	۸۹/۲۳	۸۶/۱۰

تعدادی از تصاویر پایگاه داده و نتایج حاصل از روش پیشنهادی در شکل (۱۶) نشان داده شده‌است. در این شکل تصویر (الف) یک صحنه فیلمبرداری‌شده در خارج استودیو است که در آن پس‌زمینه به‌صورت پیچیده است. در تصویر (ب) یک صحنه پویانمایی نشان داده شده است که در آن به‌دلیل پایین‌بودن تباین و روشنایی، خطای مثبت و منفی هر دو اتفاق افتاده‌اند. تصویر (پ) مثالی از خطوط متنی متحرک همراه با پس‌زمینهٔ پیچیدهٔ متحرک را نمایش می‌دهد. در این تصویر به‌دلیل سرعت بالای پس‌زمینه، خطای مثبت رخ داده است. تصویر (ت) نمونه‌ای از وجود قلم با اندازهٔ متفاوت است که به‌طور دقیق متن در آن آشکارشده است. تصویر (ث) نمونه‌ای از وجود متن فارسی و انگلیسی بوده که به‌دلیل وجود نوفه، خطای منفی رخ داده است. در تصویر (ج) هر دو متن ثابت و متحرک به‌صورت هم‌زمان وجود دارند. تصویر (چ) نمونه‌ای از تصویر با پس‌زمینهٔ پیچیده است. در تصویر (ح) متن دارای چرخش و اندازهٔ قلم متفاوت است. در این تصویر به‌دلیل چرخش متن در تشخیص ناحیهٔ متنی خطای منفی رخ داده است. نرخ دقت، فراخوانی و نمرهٔ تصاویر نمونه، در جدول (۳) نشان داده شده است.

(جدول ۳)- نرخ دقت، نرخ فراخوانی و نمره در تصاویر نمونه

تصویر	نرخ دقت	نرخ فراخوانی	نمره
الف	۱	۰/۸۳	۰/۹۰
ب	۰/۸۵	۰/۶۶	۰/۷۴
پ	۰/۸۰	۱	۰/۸۸
ت	۱	۰/۸۷	۰/۹۳
ث	۱	۰/۴۰	۰/۵۷
ج	۱	۱	۱
چ	۰/۷۵	۱	۰/۸۵
ح	۰	۰/۳۳	۰

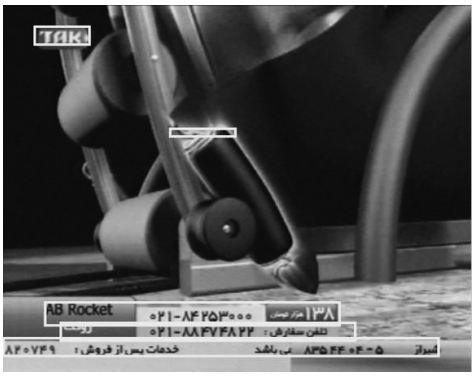
اثر برخی متغیرهای موجود در متن که می‌توانند در نمرهٔ نهایی الگوریتم استخراج متن از تصاویر ویدئویی تأثیر داشته باشند در نمودار شکل (۱۷) نشان داده شده‌است. همان‌گونه که در این نمودار نشان‌داده شده است، کمترین مقدار مربوط به ستون دوران متن است. به‌عبارت دیگر، این روش مقاومت خوبی در مقابل دوران متن ندارد و در صورت دوران متن، توانایی آشکارسازی آن‌را نخواهد داشت. با توجه به این نمودار، از جمله پارامترهای دیگری که پس از دوران، می‌توانند سبب پایین‌بودن نمرهٔ نهایی اختصاص داده شده به روش پیشنهادی شوند، می‌توان به وجود نوفه و پایین‌بودن سطح تباین و روشنایی تصویر اشاره کرد. از جمله نقاط قوّت این الگوریتم، مقاوم‌بودن آن در مقابل تغییر اندازه و سبک قلم، حرکت متن، پیچیدگی پس‌زمینه و سیاه و سفیدبودن تصویر است.

۹- نتیجه‌گیری

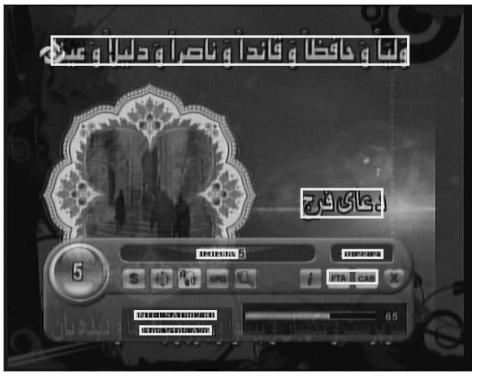
همان‌گونه که نتایج نشان می‌دهند، با تکمیل پایگاه داده و بهینه‌سازی روش‌های موجود در دو مقاله [Moradi and Mozaffaril., 2010,2011]، ارائه‌شده توسط مؤلّفان این مقاله و ارائه برخی از نو آوری هایی که در زیر به‌اختصار به آنها اشاره می‌شود، نتایج بهتری جهت آشکارسازی و استخراج متون فارسی و عربی از تصاویر ویدئویی حاصل می‌شوند.دلایل افزایش نمره در روش پیشنهادی نسبت به دو روش ارائه شده قبل توسط مؤلّفان با توجه به مطالب ارائه‌شدهٔ در بخش ۶ را می توان به‌طور اختصار به‌صورت زیر بیان کرد: افزایش سطوح هرم چند وضوحی از دو به سه، استفاده از مؤلّفه‌های رنگی تصویر ورودی به جای استفاده از تک تصویر خاکستری ورودی، تغییر روش تشخیص اندازهٔ قلم به‌کار رفته در متن و عدم

(الف)

(ب)



(پ)



(ت)

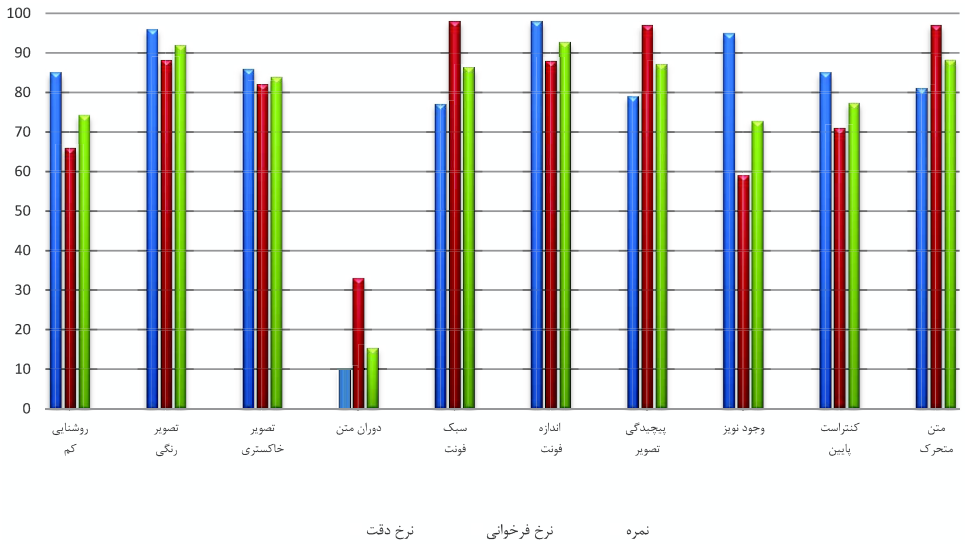
(ث)

(ج)

(چ)

(ح)

(شکل ۱۶)- نمونه‌هایی از متن آشکار شده در تصاویر ویدئویی



(شکل ۱۷)- اثر برخی از متغیرهای موجود در متن داخل تصاویر روی نرخ فراخوانی، نرخ دقت و نمره الگوریتم استخراج متن از تصاویر ویدئویی

استفاده از پارامترهای نامناسب در حذف گوشه‌های تصنعی نواحی غیر متنی، عدم استفاده از محدودیت‌های هندسی و اپراتورهای مورفولوژی جهت حذف نواحی غیر متنی و در نتیجه حذف خطای ناشی از روش‌های ابتکاری موجود در مرجع [Moradi and Mozaffari.,2010]، تکمیل و بهینه‌سازی پایگاه داده به‌کار رفته در مراجع [Moradi and Mozaffari.,2010,2011]، بهینه‌سازی بردار استفاده شده در طبقه‌بندی‌کنندهٔ SVM و حذف برخی از مشخصه‌ها جهت افزایش سرعت و دقت در مرحلهٔ طبقه‌بندی نواحی دارای متن از نواحی فاقد متن.

مراجع

Ahmad, A. M. A., Alqutami, A. J., Atoum , A Robust Algorithm for Arabic Video Text Detection, Proceedings of the 2011 2nd International Congress on Computer Applications and Computational Science Advances in Intelligent and Soft Computing (145) (2012), pp 261-266

Cai, M., Song, J., and Lyu, M.R., A New Approach for Video Text Detection, in IEEE Int. Conf. on Image Processing, (2002), pp. I-117 – I-120

Crandall, D., Antani, S., and Kasturi. R., Extraction of special effects caption text events from digital video, in IEEE International Journal on Document Analysis and Recognition (IJDAR) 5 (2003), pp. 138–157

Halima, B.M., Karray, H., Alimi, A. M., A comprehensive method for Arabic video text detection, localization, extraction and recognition, lecture notes in computer science, Advances in Multimedia Information Processing (2010), p.p 648-659

Jung ,K., Kim, K.I., and Jain, A.K., Text information extraction in images and video: A survey, Pattern Recognition 37(5) (2004) 977–997

Kim, K.I., Jung, K., and Hyung, J. , Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(12) (2003), pp. 1631–1639

Khosravi, H., Kabir, E., Farsi font recognition based on Sobel–Roberts features, Pattern Recognition Letter 31(1) (2010) 75–82

Kenneth, S.L., Barner, E., Weighted DCT coefficient based text detection, in IEEE Int. Conf. on International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (2008), pp. 1341–1344

Leon, M., A. Gasull, Text Detection In Images And Video Sequences, in International Conference on Multimedia, Image Processing and Computer Vision (Madrid, Spain, 2005)

Lim, Y.K., Choi, S.H., and Lee, S.W., Text extraction in MPEG compressed video for content-based indexing, in Proceedings of the International Confe-

rence on Pattern Recognition, (Barcelona, Spain, 2000), pp. 409–412

Liang ,J., Doermann, D., and Li, H.P., Camera-based analysis of text and documents: A survey, Document Analysis and Recognition 7(2-3) (2005) 84–104

Li, Z., Liu, G., Quin, X., Guo, D., and Jiang, H., Effective and Efficient video text extraction using key text points, IET Image Processing, 5(8) (2011), pp. 671–683

Lienhart, R., Wernicke, A., Localizing and seg-enting text in images and videos, IEEE Transactions on Circuits and Systems for Video Technology, 12(4) (2002), pp. 256–268

Moradi, M., Mozaffari, S., Orouji A.A., Farsi/Arabic text extraction from video images by corner detection, 6th Iranian Conference on Machine Vision and Image Processing (MVIP), 2010

Moradi, M., Mozaffari, S., Orouji A.A., Farsi/Arabic text extraction from video images , 19th Iranian Conference on Electrical Engineering (ICEE), 2011

Qian, X., Liu, G., Text detection localization and seg-mentation in compressed videos, in IEEE Int. Conf. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2 (2006)

Shivakumara, P., Trung, Q.P., and Chew, L. T., A Laplacian Approach to Multi-Oriented Text Detection in Video, Pattern Analysis and Machine Intelligence, IEEE Transactions, 33(2) (2011), pp. 412 – 419

Shin, C.S., Kim, K.I., Park, M.H. and Kim, H.J., Support vector machine-based text detection in digital video, in IEEE Signal Processing Society Workshop, 2 (2000), pp. 634–641

Shivakumara, P., Phan, T.Q. and Tan, C.L., A robust Wavelet transform based technique for video text detection, in IEEE 10th Int. Conf. on Document Analysis and Recognition (2009), pp. 1285–1289

Wonjun, K., Changick, K., A new approach for overlay text detection and extraction from complex video scene, IEEE Trans. Image Process, 18(2) (2009), pp. 401–411

Xiaoqian, L., Weiqiang ,W., Robustly Extraction Captions in Videos Based on Stroke-Like Edges and Spatio-Temporal Analysis: IEEE Transactions 14(2) (2012), pp. 482 – 489

Zhang ,J., Kasturi, R., Extraction of text objects in video documents: Recent progress, in Eighth International Association for Pattern Recognition (IAPR) Workshop on Document Analysis Systems (Nara, Japan, 2008), pp. 5–17

Zhao, X., Kai-Hsiang, L., Yun, F., Yuxiao, H., Yuncai, L., and T.S. Huang, Text From Corners: A Novel Approach to Detect Text and Caption in Videos. Image Processing, IEEE Transactions, 20(3) (2011), pp. 790–799

Zhong,Y., Zhang, H., and Jain, A.K., Automatic caption localization in compressed video, in IEEE Int. Conf. Image Processing, 2 (1999), pp. 96–100

محی‌الدین مرادی مدرک کارشناسی ارشد خود را در رشتهٔ الکترونیک در سال ۱۳۸۹ از دانشگاه سمنان دریافت کرد و در حال حاضر دانشجوی دکترای الکترونیک در دانشگاه صنعتی خواجه نصیرالدین طوسی می باشد. موضاعات تحقیقاتی ایشان بازشناسی آماری الگو، پردازش و بازشناسی تصاویر دیجیتال و فشرده سازی ویدئو است.

نشانی رایانامهٔ ایشان عبارت است از:

Mohieddin.Moradi@gmail.com

سعید مظفری مدرک کارشناسی، کارشناسی ارشد و دکترای خود را در رشتهٔ الکترونیک به‌ترتیب در سال‌های ۱۳۷۸، ۱۳۸۰ و ۱۳۸۶ از دانشگاه صنعتی امیرکبیر دریافت کرد. ایشان فرصت مطالعاتی خ ود را به‌مدت یک‌سال در دانشگاه صنعتی برانشوایگ در کشور آلمان سپری کرد. وی از سال ۱۳۸۷ در دانشکدهٔ مهندسی برق و کامپیوتر دانشگاه سمنان مشغول به‌فعالیت است. زمینه‌های تحقیقاتی ایشان پردازش تصویر، شناسایی الگو و پردازش متون است.

نشانی رایانامهٔ ایشان عبارت است از:

mozaffari@semnan.ac.ir

علی اصغر اروجی در سال ۱۳۴۵ در نیشابور متولد شد. ایشان مدرک کارشناسی و کارشناسی ارشد خود را در رشتهٔ مهندسی الکترونیک به‌ترتیب در سال‌های ۱۳۶۸، ۱۳۷۱ از دانشگاه علم وصنعت ایران دریافت کرد. وی در سال ۱۳۸۴ موفق به اخذ مدرک دکتری از انستیتو فنی هند (IIT) شد. ایشان از سال ۱۳۷۱ در دانشکدهٔ مهندسی برق و کامپیوتر دانشگاه سمنان مشغول به فعالیت می‌باشد. زمینه‌های تحقیقاتی ایشان طراحی مدارهای مجتمع آنالوگ و مدل‌سازی ترانزیستورهای SOI,MOS است.

نشانی رایانامهٔ ایشان عبارت است از:

aorouji@ieee.org