
بسم الله الرحمن الرحيم

دانشکده : برق و رباتیک
گروه : الکترونیک و مخابرات

استخراج متن از تصاویر صحنه های طبیعی

دانشجو : مریم سبزواری

استاد راهنما :
دکتر علیرضا احمدی فرد

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

ماه و سال انتشار : بهمن ۱۳۹۲

تقدیم به پدر بزرگوار و مادر مهربانم

آن دو فرشته ای که از خواسته هایشان گذشتند، سختی ها را به جان خریدند و خود را سپر بلای مشکلات و ناملایمات کردند تا من به جایگاهی که اکنون در آن ایستاده ام برسم .

تقدیر و تشکر

سپاس بی کران پروردگار یکتا را که هستی مان بخشید و به طریق علم و دانش رهنمونمان شد و به همنشینی رهروان علم و دانش مفتخرمان نمود و خوشه چینی از علم و معرفت را روزیمان ساخت.

از پدر و مادر عزیزم که همواره بر کوتاهی و درشتی من، قلم عفو کشیده و کریمانه از کنار غفلت هایم گذشته اند و در تمام عرصه های زندگی یار و یاوری بی چشم داشت برای من بوده اند؛ از استاد با کمالات و شایسته جناب آقای دکتر علیرضا احمدی فرد که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این پایان نامه را بر عهده گرفتند؛ کمال تشکر و قدردانی را دارم.

چکیده

استخراج متن موجود در تصاویر صحنه‌های طبیعی، امروزه مورد توجه محققین زیادی قرارگرفته است. محتوای صحنه می‌تواند به دو طبقه مهم تقسیم‌بندی شود. محتوای ادراکی و محتوای معنایی. محتوای ادراکی شامل خصوصیات رنگ، شکل و بافت صحنه می‌باشد. در مقابل محتوای معنایی شامل متن، چهره، رفتارها و حرکات انسان است. در میان اطلاعات مختلف موجود در صحنه، اطلاعات متنی از اهمیت ویژه‌ای برخوردارند، چرا که به آسانی توسط انسان قابل فهم بوده و امکان توصیف محتوای یک صحنه را فراهم می‌کنند.

در این پایان نامه، روشی برای استخراج متن از صحنه‌ها با پس‌زمینه پیچیده، بدون در نظر گرفتن زبان نوشتاری ارایه شده است. الگوریتم پیشنهادی از چهار مرحله اصلی تشکیل می‌شود. در مرحله اول به کمک ویژگی تغییر گرادیان در لبه‌های صحنه اقدام به استخراج نواحی کاندید متن می‌نماییم. در مرحله بعد از میان کاندیدها با توجه به این واقعیت که اجزاء تشکیل دهنده یک سطر متن در صحنه دارای راستا و ارتفاع تقریبا یکسانی هستند به گروه بندی نواحی استخراج شده می‌پردازیم. در مرحله سوم از ویژگی‌های هیستوگرام اندازه گرادیان و زاویه گرادیان در نواحی استخراج شده، استفاده نموده تا نواحی غیر متنی را فیلتر نماییم. برای این منظور از یک طبقه بند ماشین بردار پشتیبان که توسط ویژگی‌های هیستوگرام اندازه گرادیان و زاویه گرادیان نواحی متنی و غیر متنی آموزش دیده است استفاده می‌کنیم. در ادامه با قرار دادن معیار فاصله بر مبنای عرض نواحی متنی یافت شده و استفاده از افکنش افقی نتیجه بهبود داده می‌شود. نتایج ارزیابی روش پیشنهادی بر روی صحنه‌ها ،دارای متون فارسی و انگلیسی با قلم‌های مختلف با پس‌زمینه‌های ساده و پیچیده متون می‌باشند که بر اساس تشخیص و ارزیابی نتایج حاصل از سه مجموعه داده ICDAR 2003/2005 Dataset ، Microsoft Street View Text Detection Dataset و مجموعه داده فارسی, روش تشخیص متن

پیشنهاد شده می‌تواند برای متن با قلم‌ها، اندازه، رنگ و جهت‌گیری‌های مختلف کارآمد باشد. این نتیجه در مقایسه با روش‌های موجود بسیار امیدوار کننده است.

کلمات کلیدی:استخراج متن، هیستوگرام گرادیان، گروه بندی، ماشین بردار پشتیبان

لیست مقالات استخراج شده از پایان‌نامه

1- Extracting Text from Natural Scene Images by Features Based on Gradient , International Research Journal of Applied and Basic Sciences

۲- استخراج متن از تصاویر با پس‌زمینه پیچیده به کمک ویژگی‌های آماری، Iranian Journal of Science and Technology Transactions of Electrical Engineering

فهرست

عنوان **صفحه**

فهرست جدول‌ها

عنوان **صفحه**

فهرست شکل‌ها

عنوان **صفحه**

فصل اول

مقدمه

۱- مقدمه

امروزه استخراج متن از صحنه به خاطر اینکه توصیف قدرتمندتری از محتوای صحنه را ارائه می-کند، مورد بررسی بیشتر قرار گرفته است. اما تشخیص خودکار متن در صحنه در شرایط طبیعی کار دشواری است. این مشکلات از محدودیت‌های زیر ناشی می‌شود.

۱) حروف متن می‌توانند در نوع قلم، اندازه، ترتیب قرار گرفتن در یک خط، رنگ وحتی بافت تغییر کنند.

۲) متن معمولا" در یک زمینه پیچیده‌ی رنگی دارای بافت قرار دارد.

۳) نوشته ها همیشه به صورت افقی ویا عمودی نیستند و می‌توانند در جهت‌های مختلف نوشته شوند.

تحلیل اطلاعات متنی موجود در صحنه‌ها، امکان انجام فعالیت‌هایی نظیر شناسایی پلاک خودروها، تشخیص و ترجمه علائم، جستجوی محتوای صحنه‌ها، واقعه نگاری در دنباله‌های ویدیویی و شاخص‌گذاری مبتنی بر متن صحنه‌ها را به صورت خودکار فراهم می‌کند. همچنین با توجه به تنوع قلم، سبک، اندازه، جهت و رنگ متون و پیچیدگی زمینه در صحنه‌های مناظر طبیعی، استخراج متن در این دسته از صحنه‌ها، یکی از مسائل چالش برانگیز در پردازش تصویر است.

۱-۱ طرح مساله

امروزه با گسترش دوربین‌های دیجیتال و تلفن‌های همراه، صحنه‌ها از اهمیت ویژه ای برخوردارند. و برای کاربردهای زیادی از جمله حاشیه نویسی خودکار، شاخص‌گذاری و بازیابی و پایگاه داده‌های چندرسانه‌ای مبتنی بر محتوا و ویژگی‌های توصیفی مرتبط با صحنه مورد نیاز هستند. ویژگی‌هایی مانند بافت،رنگ،شکل و طرح می‌توانند در سطح پردازش پایین استخراج شوند. ویژگی‌های سطح

پایین به آسانی قابل استخراج هستند ولی نمی‌توانند ایده روشنی درباره آنچه در صحنه است به ما بدهند.

متون حاوی اطلاعات شفافی هستند و می‌توانند به عنوان یک ابزار قوی در کاربردهای مختلف به‌کار روند. بنابراین ادراک متن موجود در صحنه از اهمیت زیادی برخوردار است. ولی از آنجا که داده‌های متنی می‌توانند در یک صحنه طبیعی در قلم‌ها، سبک‌ها، اندازه‌ها، جهت‌ها و رنگ‌ها همچنین در یک زمینه پیچیده مخفی شوند،مساله استخراج متن به یک چالش تبدیل شده است.

در اکثر کارهای انجام شده سیستم ادراک متن از چهار گام اصلی تشکیل شده است. این گام‌ها عبارت‌اند از: تشخیص، مکان‌یابی، استخراج و شناسایی متن. از بین این گام‌ها دو بخش اول در کارایی کلی سیستم، نقش حیاتی دارند.

تصاویر مناظر طبیعی، تصاویری از صحنه‌های دنیا واقعی هستند. هیچ دانش اولیه‌ای در مورد وجود یا عدم وجود متن، نوع محیط، نور پردازی و پارامترهای تصویر برداری در دست نمی‌باشد.

۱-۲ کاربردهای استخراج متن از تصاویر مناظر طبیعی

با پیشرفت تجهیزات تصویربرداری مانند دوربین دیجیتال و دوربین‌های تعبیه شده در تلفن‌های همراه کاربردهای سیستم استخراج متن از صحنه‌ها روز به روز بیشتر می‌شود. سیستم ادراک متن می‌تواند به عنوان بخشی از یک سیستم بزرگ، برای ارتباط بهتر محیط و انسان، مورد استفاده قرار گیرد که این ارتباط را یک واسط کامپیوتری برقرار می‌کند. به عبارت دیگر، این سیستم می‌تواند بخشی از یک سیستم بزرگتر مانند یک گوشی همراه باشد یا در یک اتومبیل تعبیه شده باشد و با تصویر برداری از محیط و تشخیص وجود متن در صحنه‌ها و تعیین دقیق محل متن و سپس جدا کردن متن از زمینه و شناسایی آن، درک متن موجود در صحنه‌ها را ممکن می سازد و در صورت لزوم

متن شناسایی شده را به زبان مقصد ترجمه کرده وآن را روی یک مانیتور نشان داده و یا به صورت یک پیغام صوتی پخش کند.

کاربردهای دیگر سیستم عبارتند از :

پردازش و خواندن خودکار اسناد

تشخیص پلاک خودرو

نظارت و جستجوی خودکار

رباتیک(کمک به ربات‌ها در درک بهتر محیط)

آنالیز اسناد حاصل از دوربین

تکنیک‌های پیش‌پردازش برای OCR

۱-۳ چالش‌های فهم متن از تصاویر مناظر طبیعی

تشخیص متون موجود در تصاویر صحنه طبیعی، به دلیل تغییرات زیاد در شرایط تصویربرداری و صحنه‌های مورد نظر با چالش‌های زیادی روبرو است و از بین متون مختلف، متن صحنه طبیعی سخت‌ترین نوع متن برای استخراج محسوب می‌شود. بنابراین چالش‌های تشخیص متن را باید در دو عامل شرایط تصویر برداری و صحنه‌های مورد نظر جستجو کرد.

مشکلات ناشی از تصویربرداری عبارتند از:

زاویه دید: در تصویربرداری از یک صحنه، زاویه دوربین همیشه مستقیم نیست و متون می‌توانند زوایای مختلفی نسبت به دوربین داشته باشند.

تاری: در طول تصویربرداری حرکت دوربین یا غیرفوکوس بودن آن سبب تاری صحنه می‌شود، همچنین تاری ایجاد شده توسط بزرگ‌نمایی نادرست ممکن است کیفیت صحنه را خراب کند.

نورپردازی: در تصاویر واقعی، منابع نوری غیریکنواخت و بازتاب نور اشیا ممکن است سبب تغییر شدت روشنایی رنگ شده و کارایی سیستم تشخیص متن را مختل کند.

درجه تفکیک: محدوده تغییرات درجه تفکیک از وب‌کم‌ها تا دوربین‌های پیشرفته وسیع است و سیستم شناسایی باید بتواند صحنه‌ها با درجه تفکیک پایین را آنالیز نماید.

و صحنه‌های مورد نظر با مشکلات زیر روبرو هستند:

تغییرات متن: متون درتصاویر صحنه‌های طبیعی، می‌توانند از نظر قلم، اندازه، جهت، ترازبندی و مکان تغییرات زیادی داشته باشند.

پیچیدگی زمینه: زمینه تصاویر صحنه‌های طبیعی لزوما یکنواخت نیستند بلکه دارای بافت می-باشند. هرچه زمینه پیچیده‌تر باشد استخراج متن از صحنه سخت‌تر می‌شود.

اشیا غیرمسطح: کج بودن متن موجود روی اشیا غیرمسطح یکی دیگر از مشکلات تشخیص متون در صحنه‌های طبیعی است.

طرح‌بندی ناشناخته: وجود نداشتن اطلاعات اولیه‌ای از ساختار متن، تشخیص آن را با مشکل مواجه خواهد کرد.

اشیاء در فاصله: متغییر بودن فاصله بین متن و دوربین و همچنین اندازه قلم، باعث وسیع شدن محدوده تغییر در اندازه نویسه‌ها می‌شود.

ویژگی‌های خاص زبان فارسی: ویژگی‌هایی مانند کشیدگی زیاد در صحنه‌ها در برخی از حروف، کار تشخیص متن را با مشکل مواجه می‌سازد.

هدف اصلی یک سیستم تشخیص متن این است که متون موجود در صحنه‌ها را با طرح‌ها، رنگ-ها، قلم‌ها و اندازه‌های مختلف و همچنین در شرایط تصویر برداری، نورپردازی مختلف و با پیچیدگی زمینه را تشخیص دهد.

۱-۴ ساختار پایان‌نامه

پایان‌نامه حاضر در شش فصل طبقه بندی شده است. چنان که گذشت، فصل اول مقدمه‌ای برای آشنایی با موضوع پایان‌نامه، دلایل انجام آن، کاربردها و چالش‌های پیش رو است. در فصل دوم پیشینه‌ای از کارهای انجام شده بیان می‌شود و با دسته‌بندی کارهای گذشته در ۶ گروه به بیان ویژگی‌ها و نقاط ضعف و قوت هر گروه پرداخته می‌شود و در فصل سوم توضیح مختصری در مورد مفاهیم به کار رفته در پژوهش گرد آوری شده است. در فصل چهارم روش پیشنهادی به تفصیل آورده شده است. الگوریتم پیشنهادی از چهار مرحله اصلی تشکیل می‌شود. در مرحله اول به کمک ویژگی تغییر گرادیان در لبه‌های صحنه اقدام به استخراج نواحی کاندید متن می‌نماییم. از میان کاندیدها، با توجه به این واقعیت‌هایی که بیان خواهیم کرد به گروه بندی نواحی استخراج شده می‌پردازیم. در ادامه با استفاده از ویژگی‌های هیستوگرام اندازه گرادیان و زاویه گرادیان در نواحی استخراج شده نواحی غیر متنی را فیلتر می‌نماییم. در آخر با قرار دادن معیار فاصله بر مبنای عرض نواحی متنی یافت شده و استفاده از افکنش افقی نتیجه بهبود داده می‌شود. در فصل پنجم نتایج تجربی بر روی سه مجموعه داده ICDAR 2003/2005 Dataset ، Microsoft Street View Text Detection Dataset و پایگاه داده به زبان فارسی را نمایش داده و به ارزیابی سیستم با توجه به معیارهای موجود پرداخته‌ایم و در فصل ششم نتیجه‌گیری و راه کارهای آینده آمده است.

فصل دوم

مروری بر روش‌های شناسایی متن در صحنه‌ها

۲- پیشینه پژوهش

در این بخش به طور مختصر به مرور و بررسی کارهای انجام شده در این زمینه خواهیم پرداخت. به طور کلی برای دسته‌بندی روش‌های موجود در تشخیص و مکان‌یابی، می‌توان پنج رویکرد مختلف ارائه کرد. این رویکردها ۱) روش‌های مبتنی بر ناحیه ۲)روش‌های مبتنی بر مولفه‌های متصل ۳) روش‌های مبتنی بر آموزش ماشین ۴)روش‌های مبتنی بر رنگ و ۵)روش‌های مبتنی بر بافت می‌باشند. در نهایت به معرفی روش‌هایی که از ترکیب رویکردهای فوق برای کارآیی بهتر سیستم خود استفاده کرده‌اند، می‌پردازیم.

۲-۱ روش‌های مبتنی بر ناحیه

در روش‌های مبتنی بر ناحیه، به طور کلی یک بردار ویژگی که از هر ناحیه استخراج می‌شود، برای تخمین احتمال متن در ناحیه مورد استفاده قرار می‌گیرد[۱]. سپس نواحی متنی همسایه، برای ایجاد بلوک‌های متن باهم ادغام می‌شوند. از آنجا که نواحی متن خصوصیات مجزایی نسبت به غیر متن دارند، این روش‌ها قادر به تشخیص و مکان‌یابی دقیق متون حتی در صحنه‌های نویزی هستند.

در مورد روش‌های مبتنی بر ناحیه، سرعت به نسبت پایین است و کارایی به جهت ترازبندی متن وابسته است[۱]. بیشتر روش‌های مبتنی بر ناحیه، بر اساس این مشاهده استوارند که نواحی متنی خصوصیات مجزایی مانند توزیع گرادیان و ...، نسبت به نواحی غیر متن دارند.

در روش‌های مبتنی بر ناحیه، ابتدا با استفاده از فیلترهای اکتشافی ، نواحی متن در صحنه مشخص می‌شود.سپس این نواحی، با استفاده از روش K-means یا آستانه‌یابی به نواحی متن و پس زمینه تقسیم می‌شود[۲].روش‌های مبتنی بر ناحیه، به الگوریتم امکان پردازش صحنه‌های پیچیده را می‌دهد اما مشکلاتی در هر دو زمینه تشخیص و قطعه‌بندی دارد.[۱]

یک روش اولیه که توسط و'و و همکارانش[۳] پیشنهاد شده، از مجموعه فیلترهای گوسین مرتبه دوم روی سه مقیاس مختلف، برای استخراج خصوصیات بافتی از نواحی محلی صحنه استفاده می‌کند. با پاسخ‌های فیلتر مربوطه، همه پیکسل‌های صحنه به یکی از سه کلاس متن، غیر متن و زمینه پیچیده تقسیم می‌شوند. سپس دسته‌بند K-means و عملگرهای ریخت‌شناسی[۱] برای گروه بندی پیکسل‌های متن به نواحی متنی مورد استفاده قرار می‌گیرند. این الگوریتم برای صحنه‌های اسکن شده طراحی شده، که معمولا نسبت سیگنال به نویز [۲]بهتری نسبت به صحنه‌های طبیعی و فریم‌های ویدیویی دارند. مشکل اصلی این روش، ناحیه‌بندی[۳] بافت و حساس بودن کیفیت ناحیه‌بندی به نویز است.

گلاوالا و همکارانش[۴] از تبدیلات تصاویر مانند تبدیل کسینوسی گسسته و موجک برای استخراج ویژگی استفاده می‌کنند. با آستانه گرفتن از پاسخ این فیلترها، نواحی غیر متن حذف شده و نواحی متن باقی مانده، بر اساس روابط مکانی‌شان گروه بندی می‌شود. یک هدف مهم تبدیل موجک، تجزیه یک سیگنال به زیر باندها در اندازه‌ها و فرکانس‌ها مختلف است. در مورد صحنه‌ها تبدیل موجک برای تشخیص لبه‌ها در جهت‌های مختلف به‌کار می‌رود. این تبدیل با استفاده از بانک فیلتری شامل فیلتر-های بالاگذر و پایین‌گذر پیاده‌سازی می‌شوند. کاربرد آن‌ها برای یک صحنه ، شامل یک فرایند فیلترینگ در جهت افقی و سپس در جهت عمودی است. برای تخمین بردار ویژگی یک پنجره لغزان با اندازه w*h پیکسل بدون هم‌پوشانی روی صحنه تبدیل یافته، حرکت می‌کند. w عرض و h طول پنجره است. برای هر موقعیت پنجره و برای هر زیر باند انحراف معیار هیستوگرام مربوط(H) محاسبه می‌شود.

[1] Morphology
[2] Pick Signal to Noise Rate
[3] Texture segmentation

مقدار انحراف معیار بلوک در هر زیر باند بردار ویژگی هر بلوک را تشکیل می‌دهد[۴]. انتخاب این ویژگی بر اساس این مشاهده است که برای نواحی غیرمتن ضرایب موجک در زیر باندهای LH, HL و HH از یک توزیع لاپلاسین پیروی می‌کنند. در حالی که در متون، این ضرایب حول مقدار گسسته کوچکی متمرکز می‌شوند. در این روش برای دسته‌بندی داده‌ها به سه دسته متن، غیرمتن و زمینه پیچیده از الگوریتم K-means استفاده می‌شود. برای اصلاح مکان متن، متون نامزد توسط نگاشت افکنش[1] مورد ارزیابی بیشتر قرار می‌گیرند. نتایج آزمایش نشان می‌دهد که این ویژگی انتخاب شده برای تمایز بین متن و زمینه مفید است.

شیواکوما در[۵]، روش جدیدی را برای تشخیص متن ارائه می‌دهد که روی دسته‌بندی خودکار صحنه‌ها با تباین بالا و پایین تاکید دارد. دلیل اصلی نرخ پایین تشخیص متن در بسیاری از روش‌ها این است که، فرض می‌کنند متن تباین بالایی نسبت به زمینه دارد، و به همین دلیل یک آستانه ثابت را برای جدا کردن متن از زمینه در نظر می‌گیرند. در صحنه‌ها با تباین پایین، هرچه این آستانه افزایش یابد، نواحی اشتباه بیشتری را تحت عنوان متن خواهیم داشت و در صحنه‌ها با تباین بالا، وجود حد آستانه پایین، نواحی اشتباه بیشتری را ایجاد خواهد کرد. بنابراین حداقل دو آستانه، یکی برای صحنه‌ها با تباین بالا و دیگری برای صحنه‌ها برای تباین پایین لازم است، تا کارایی تشخیص متن بهبود یابد. برای این منظور باید صحنه‌ها با تباین بالا و پایین را دسته‌بندی کرد.

در این مقاله[۵]، با استفاده از عملگر سوبل[2] لبه چهارجهته را، که جهت و قدرت لبه‌ها را در چهارجهت نشان می‌دهد، به دست آورده و متوسط آن‌ها را محاسبه می‌کند. برای توصیف صحنه از ویژگی‌های میانگین، انحراف معیار، انرژی وآنتروپی استفاده می‌شود. سپس از الگوریتم K-means برای دسته‌بندی ویژگی‌ها به دو دسته متن و زمینه استفاده می‌شود و روی نواحی نامزد عملیات مورفولوژی

[1] Projection Profile Analysis
[2] Sobel operation

بازکردن و گسترش را اعمال کرده تا مولفه‌های متصل ایجاد شود. در اینجا از یک فیلتر ریاضی برای تاری بلوک و یک فیلتر میانه برای حذف نویز استفاده شده است. پس از دسته‌بندی صحنه‌ها ، دو حد آستانه برای صحنه‌ها با تباین بالا و پایین با استفاده از گرادیان میانگین نقشه لبه و بر اساس این واقعیت که، در صحنه‌ها با تباین بالا بیشتر مقادیر گرادیان بزرگتر یا مساوی یک است و در مورد صحنه‌ها با تباین پایین برعکس است، به صورت خودکار محاسبه می‌شود.

در این مقاله برای حداقل کردن نقاط اشتباه، علاوه بر ارتفاع و پهنای نواحی نامزد متن، ویژگی-های جدیدی مانند درجه صاف بودن و شکستگی معرفی شده است. اگر مرکز یک مولفه در همان مولفه قرار گیرد،آن مولفه صاف و در غیر این صورت آن را شکسته می‌نامند[۵]. تعداد فاصله‌ها (N_S) در بلوک تشخیص داده می‌شود و مجموع پیکسل‌ها در بلوک متن تشخیص داده شده، تحت شرایط زیر حذف خواهند شد:

$$((H<6) \cup (W<5) \cup (A<24) \cup (H>70)) \quad (۱-۲)$$

$$((A_{HW}>0.3) \cap (N_S=0) \cap (N_{CT}=0)) \quad (۲-۲)$$

که H ارتفاع مولفه و W پهنا آن می‌باشد. N_{CT} تعداد مولفه‌های صاف و A_{HW} چگالی مولفه است.

پارک و همکارانش در [۶]، فرض می‌کنند که در متون روی خط مرکزی صحنه‌ها و به صورت افقی قرار گرفته اند. بنابراین برای تشخیص متن از یک روش لبه‌یابی ساده استفاده می‌کنند. بدین صورت که با با عملگر کنی[1]، لبه‌های موجود در صحنه‌های خاکستری را پیدا کرده و با تحلیل نمودار افقی و سپس عمودی افکنش، مکان متون موجود در صحنه‌ها را پیدا می‌کنند. فرض در نظر گرفته شده مبنی بر این که متون روی خط مرکزی صحنه‌ها و به صورت افقی قرار گرفته اند ، بسیار محدود کننده و روش ارائه شده، بسیار ساده است.

[1] Canny

۲-۲ روش‌های مبتنی بر مولفه متصل

روش‌های مبتنی بر نواحی متصل [1] نواحی متنی را با استفاده از تشخیص لبه و خوشه‌بندی رنگ در اطراف لبه‌ها پیدا می‌کنند. سپس اجزای غیر متن با استفاده از دسته‌ای از قوانین فیلتر می‌شوند. از آن‌جا که تعداد مولفه‌های پیدا شده، به نسبت کم است این روش‌ها هزینه محاسباتی کمی دارند و مولفه های متن می‌توانند به طور مستقیم برای شناسایی مورد استفاده قرار گیرند. اگرچه روش‌های موجود از نظر مکان‌یابی کارآیی خوبی دارند، ولی مسائل حل نشده‌ای نیز وجود دارد.به عنوان مثال روش‌های مبتنی بر مولفه متصل بدون داشتن اطلاعات اولیه‌ای از مکان و اندازه متن، نمی‌توانند با دقت مولفه‌های متن را پیدا کنند. علاوه بر این، به دلیل وجود اجزای غیر متن زیاد، طراحی سریع و مطمئن یک تحلیل‌گر مولفه متصل سخت است.[۷]

برخلاف روش‌های مبتنی بر ناحیه، روش‌های مبتنی بر مولفه متصل بر این اساس عمل می‌کنند که متون می‌توانند به عنوان مجموعه‌ای از اجزای متصل در نظر گرفته شوند که هرکدام ویژگی‌های هندسی مجزایی دارند و اجزای همسایه روابط مکانی و هندسی نزدیکی با هم دارند. این روش‌ها معمولا از سه گام تشکیل می شوند.

۱) استخراج مولفه‌های متصل برای جداکردن متن از صحنه.
۲) تحلیل مولفه‌های متصل برای فیلتر کردن اجزای غیر متن با استفاده از کلاسه‌بندها.
۳) پیش پردازش برای گروه کردن مولفه‌های متنی به بلوک‌های متن.

از طرف دیگر روش‌های مبتنی بر مولفه متصل بدون داشتن اطلاعات اولیه از مکان و اندازه متن، نمی‌توانند با دقت مولفه‌های متن را جدا کنند. علاوه بر این، به دلیل وجود اجزای غیر متن زیاد، طراحی سریع و مطمئن یک تحلیل‌گر مولفه متصل سخت است. برای غلبه بر مشکلات فوق، می‌توان

[1] Connected Component

از یک روش ترکیبی برای تشخیص و مکان‌یابی مطمئن متون درتصاویر صحنه‌های طبیعی استفاده کرد و از مزایای هر دو روش بهره گرفت.

در مرجع[۸] بعد از باینری کردن صحنه‌ها، اجزای متصل از هم تفکیک می‌شود و در مرحله بعدی به استخراج ویژگی از نواحی متصل پرداخته می‌شود. از جمله ویژگی‌ها می‌توان به ویژگی‌های هندسی از قبیل مساحت، عرض، ارتفاع، منطقه محدب[1]، محیط و نسبت طول به عرض ناحیه اشاره کرد. ویژگی بعدی نظم ناحیه متصل از لحاظ فرورفتگی‌ها و تعداد حفره‌های احتمالی در ناحیه است. بعد از شناسایی نواحی کاندید و یک مرحله آموزش توسط forest learning، نواحی برای ایجاد کلمات باهم گروه‌بندی می‌شوند.نتایج نشان می‌دهد که استفاده از نواحی متصل در مواقعی که قسمتی از صحنه در روشنایی و قسمتی تیره‌تر است نتایج بهتری نسبت به استفاده از روش‌های متکی بر رنگ دارد.

۲-۳ روش‌های مبتنی بر یادگیری ماشین

روش‌های مبتنی بر یادگیری ماشین [2]از ۳ گام تشکیل شده اند .

۱) تشخیص متون نامزد: به منظور ایجاد تعادل بین کارایی و زمان پردازش ابتدا نامزدهای متن را شناسایی می‌کنند. هدف این گام، نرخ فراخوانی بالاو هزینه محاسباتی پایین است. بنابراین نرخ دقت کمی مورد انتظار است. برای این منظور باید ویژگی‌های ساده‌ای برای تشخیص متون نامزد مورد استفاده قرار می‌گیرد. در مراحل بعدی این متون مورد بازبینی قرار گرفته و نواحی متن و غیرمتن متمایز خواهد شد.

[1] Convex hull
[2] Machine Learning

۲) اصلاح متون نامزد: از آن‌جا که الگوریتم تشخیص متون نامزد نسبتا تجربی است و ممکن است نقاط اشتباه زیادی را تولید کند. برای حذف این نواحی اشتباه یک فرایند اصلاح برای استخراج ویژگی از نواحی نامزد دسته‌بند مبتنی بر یادگیری ماشین، به کار می‌رود.

۳) اصلاح نهایی: در بعضی از روش‌ها، نواحی متن حاصل از گام دوم با استفاده از بعضی قوانین غیر مستدل یا روش‌های اصلاح خطوط مورد بررسی قرار می‌گیرند.

در ادامه روش‌هایی که از ابزار یادگیری ماشین استفاده کرده‌اند مرور شده است.

در مرجع [۹]، ازadaboost که یک کلاسه‌بند قوی را به کمک ترکیب کلاس‌بندهای ضعیف طراحی می‌کند استفاده شده است. شش نوع استراتژی برای تشخیص متن در نظر گرفته شده که شامل مقدار انرژی محلی فیلتر گابور در چهار جهت، واریانس، اندازه‌گیری بافت آماری هیستوگرام صحنه، اندازه‌گیری واریانس ضریب موجک، تشخیص لبه و محاسبه فاصله لبه و تجزیه و تحلیل اجزای متصل می‌باشد.تعداد ۶۹ ویژگی استخراج شده توسط adaboost آموزش داده می‌شود. بعد از آموزش داده‌ها برای بهبود نتایج به‌دست آمده از adaboost یک عملیات پس‌پردازش و مورفولوژی بر روی صحنه‌ها انجام می‌گیرد. نتایج نشان می‌دهد که این الگوریتم در مواردی که حتی شدت روشنایی کم باشد پاسخ مناسبی می‌دهد.

پارک و همکارانش در [۱۰] به دنبال شناسایی اجزای متن توسط یک بردار ماشین پشتیبان هستند. در ابتدا به خوشه بندی متن بر اساس رنگ پرداخته و آن را به گروه‌هایی با توجه به توزیع سه رنگ قرمز، آبی و سبز تقسیم می‌کنند. در مرحله بعدی هر گروه به حوزه موجک برای استخراج ویژگی برده می‌شود و از ضرایب موجک برای طبقه‌بندی اجزا با کلاسه‌بند svm استفاده می‌شود. الگوریتم پیشنهادی نسبت به تغییرات نور مقاوم بوده و به خاطر استفاده از ضرایب موجک به ویژگی‌های ظاهری متن نظیر اندازه، جهت وشکل وابسته نیست.

۲-۴ روش‌های مبتنی بر رنگ

روش‌های مبتنی بر رنگ با در نظر گرفتن این موضوع که نویسه‌های متنی دارای رنگ یکسانی هستند به خوشه بندی صحنه بر اساس رنگ و نمودار شدت می‌پردازد.

فو و همکاران[۱۱] یک روش تشخیص متن را بر اساس محدودیت‌های متعدد در پس‌زمینه‌های پیچیده معرفی می‌کند. تقسیم‌بندی اولیه توسط خوشه بندی k-means بر اساس بردار رنگ Ycbr اجرا شده است. تعداد خوشه‌ها سه و یا چهار، بسته به تعداد قله‌های هیستوگرام یک صحنه تعیین می‌شود. بعد از مشخص شدن نواحی متصل سه محدودیت برای از بین بردن نواحی اضافه باقی‌مانده استفاده می‌شود.

۱) محدودیت رنگ که تمام نواحی متصل هم رنگ را شامل می‌شود.

۲) محدودیت لبه که باید لبه نواحی متصل قوی باشد.

۳) محدودیت عرض که ارتفاع و عرض بیشتر از یک آستانه حذف می‌شود.و در نهایت با ترکیب سه شرط ناحیه متنی محاسبه می‌شود.

همانطور که مشخص است اگر متن دارای رنگ‌های متفاوت و یا مشابه پس زمینه باشد، تکیه بر این روش با مشکل مواجه می‌شود. این روش به تنهایی پاسخ‌گوی مناسبی برای تفکیک متن و پس-زمینه نمی‌باشد.

۲-۵ روش‌های مبتنی بر بافت[1]

علاوه بر خواص ناحیه‌ای، خواص بافتی اطلاعات مفیدی برای استخراج متن از پس‌زمینه ارائه می-دهند. در این روش‌ها از ویژگی‌های آماری، تبدیل فرکانس، فیلتر گابور و روش‌های مبتنی بر یادگیری ماشین برای توصیف و تشخیص بافت متن از پس‌زمینه استفاده می‌شود.

کلین و همکاران، [۱۲] و [۱۳] یک روش خوشه‌بندی متریک را به کار برده‌اند. آن‌ها بر این نکته رسیدند که فضای RGB نسبت به فضاهای دیگر از تنوع بهتری در صحنه‌های طبیعی برخوردار است. در خوشه‌بندی از فاصله اقلیدسی و شباهت کسینوسی، در فضای رنگ RGB و فضای مکمل رنگ، استفاده شده است. در این روش سه خوشه کلی پس‌زمینه، پیش‌زمینه و نویز تشکیل می‌شود و سپس برای پیدا کردن پیش‌زمینه متنی به نظم خوشه‌های استخراج شده توجه می‌شود. در مرحله بعدی از یک فیلتر گابور برای ترکیب اطلاعات مکانی و اطلاعات فرکانسی برای شناسایی نویسه‌ها استفاده می-شود و برای استخراج ویژگی‌های فرکانسی از روش‌های مبتنی بر متن در حوزه فرکانس نظیر تبدیل فوریه و تبدیل کسینوسی گسسته و تبدیل موجک کمک گرفته می‌شود.

در[۱۴] شیواکومارا و همکارانش، برای مکان‌یابی خودکار متون الحاقی در صحنه‌های فشرده JPEG و فریم‌های ویدیویی، روش جدیدی ارائه کرده اند،که در آن نواحی متن با استفاده از خصوصیات بافتی از زمینه‌شان جدا می‌شوند. برخلاف روش‌های پیشین، که قبل از استخراج متن، ویدیو را کاملا از حالت فشرده خارج می‌کردند، در این روش نواحی نامزد متن را مستقیما در حوزه فشرده تبدیل کسینوس گسسته و با استفاده از اطلاعات تغییر شدت موجود در این حوزه، استخراج می‌کنند. ضرایب تبدیل کسینوسی گسسته در صحنه‌ها، به عنوان شاخص بافت برای تشخیص نواحی متن مورد استفاده قرار می‌گیرند. این ضرایب در صحنه‌های فشرده، ویژگی‌های محلی مربوط به جهت و تناوب را

[1] texture

استخراج می‌کنند. هر بلوک در صحنه‌های فشرده شده، براساس تغییرات محلی عمودی و افقی شدت، به یکی از کلاس‌های متن با غیر متن دسته‌بندی می‌شوند. علاوه بر این، فرایند پیش‌پردازشی شامل عملگرهای ریخت‌شناسی و تحلیل مولفه‌های متصل برای اصلاح متون تشخیص داده شده، انجام می-گیرد.از آنجا که الگوریتم و فرایند اصلاح و پردازش بر روی صحنه‌ها فشرده انجام می‌گیرد، این الگوریتم بسیار سریع است.

نتایج آزمایش‌ها نشان می‌دهد اگرچه ضرایب تبدیل کسینوسی در تشخیص متون موفق است ولی نواحی غیر متن زیادی را ، به صورت اشتباه، متن تشخیص می‌دهد. بنابراین می‌توان نتیجه گرفت که روش‌های مبتنی بر بافت، بر اساس آماره‌های محلی شدت، به تنهایی برای متمایز کردن نواحی متن و غیرمتن کافی نیستند و همچنین اگر نویسه‌ها خیلی بزرگ باشند یا به صورت پراکنده پخش شده باشند به طوری که بافت متمایز کننده‌ای ایجاد نکند این روش نتایج خوبی ارائه نمی‌دهد.

در روش مرجع[۱۵] از میان موجک‌های موجود، موجک هار به علت داشتن سرعت پردازش بالا انتخاب می‌شود. این موجک بر روی اجرای رنگ (R,G,B) اعمال می‌شود تا لبه‌های متن شناسایی شوند. لبه‌های قوی که اکثرا لبه‌های متن را در خود دارند با اعمال فیلتر سوبل بر روی زیر باندهای جزئیات آشکارسازی می‌شوند. با باینری کردن زیرباندهای جزئیات و اعمال عملگر منطقی AND، نواحی نامزد متن شناسایی می‌شوند.سپس نواحی غیرمتن با مجموعه‌ای از قواعد شهودی حذف خواهند شد. در نهایت با برچسب‌زنی و آستانه‌یابی، متن از تصویر استخراج می‌شود.

۲-۶ روش‌های ترکیبی

روش‌های موجود از ترکیب ویژگی‌های مختلف برای استخراج متن بهره می‌گیرند.

دینه و همکاران[۱۶] مکان متن را با استفاده از ویژگی‌های لبه و رنگ مشخص می‌کنند. لبه‌ها توسط آشکارساز لبه سوبل مشخص شده و با توجه به محدودیت‌های هندسی عمل فیلتر کردن انجام می‌شود. نواحی کاندید متن توسط لبه‌هایی که به یکدیگر متصل هستند ایجاد می‌شوند. مناطق متنی ایجاد شده توسط ویژگی رنگ نیز بر اساس محدودیت‌های هندسی ناحیه رنگی بر اساس ابعاد و ارتفاع بوده و دو نتیجه حاصله برای ایجاد نتیجه نهایی ترکیب می‌شوند و تشخیص نهایی را امکان‌پذیر می‌سازند.

یی و همکاران[۱۷] ویژگی عرض قلم و رنگ را برگزیده و با استفاده از شدت رنگ و نسبت ابعاد، به پالایش نواحی می‌پردازند. برای پارتیشن بندی بر اساس رنگ ، از خوشه‌بند k-means استفاده شده است. با این روش تصویر به چند لایه رنگی تفکیک می‌گردد. در مرحله بعدی برای حذف نواحی غیر متنی از دو روش گروه بندی استفاده می‌شود. در روش اول بر مبنای محدودیت‌های هندسی مانند شباهت در ارتفاع بلوک‌ها، چیدمان بلوک‌ها، رنگ و فاصله دو ناحیه مجاور نواحی غیرمتنی حذف می‌شوند. گروه‌بندی خط متن توسط تبدیل هاف [1] بر اساس اختلاف زاویه و اختلاف طول پاره‌خط متصل بین دو ناحیه مجاور صورت می‌گیرد.

در [۱۸] روش‌های مبتنی بر رنگ، لبه و بافت تصویر را ترکیب می شوند. نگاشت لبه‌ها در جهت‌های عمودی، افقی، مورب چپ و راست ایجاد می‌شود. سپس توسط روش‌های بدون سرپرستی، بلوک متن از زمینه تشخیص داده می شود. به این که، در ابتدا با کمک لبه یاب، لبه های تصویر را پیدا می‌شود. با این کار تاثیر زمینه کم می‌شود و به طور موثری می توان کاندیدای اولیه برای متن را انتخاب کرد. ثانیا ویژگی‌های بافتی متن را برای هر پیکسل از تصاویر لبه حساب شده و سپس از الگوریتم k-means برای دسته‌بندی ویژگی‌ها به دو دسته متن و زمینه استفاده می‌شود. در پایان، کاندیداهایی

[1] Hough trasformation

که به عنوان متن تشخیص داده شده بودند دستخوش آنالیز و قوانین تجربی قرار می‌گیرند و محل-هایی که متن هستند شناسایی شده و برای استخراج و پالایش آماده می‌شوند. این روش می تواند متن را با سایز ،رنگ و قلم‌های مختلف تشخیص دهد.

در[۱۹] با استفاده از الگوریتم EM اقدام به استخراج پارامترهای مخلوط گوسی [1]در فضای رنگ لبه‌های صحنه شده است. بدین صورت که ۸ همسایگی مجاور هر پیکسل را مورد بررسی قرار داده و بیشترین و کمترین شدت رنگ را در سه بعد تحت عنوان یک بردار ویژگی ۶ بعدی در نظر گرفته می‌شود و همچنین چون مکان پیکسل‌ها حایز اهمیت است و با توجه به این فرض که نوشته‌ها در راستای افقی هستند از مولفه x پیکسل‌ها صرف‌نظر شده و مولفه y با سه بار تکرار به ۶ ویژگی قبلی اضافه و یک بردار ویژگی ۹ بعدی ایجاد می‌کند. در مرحله بعدی صحنه با استفاده از این ویژگی-ها به ۵ زیر تصویر تقسیم می‌شود. و سپس به استخراج ناحیه متنی با توجه به لبه‌ها از هر کدام از زیر تصویرها پرداخته می‌شود. در ادامه برای حذف نواحی غیر متنی از کلاسه بند ماشین بردار پشتیبان [2] با تمرکز بر ویژگی‌های گرادیان و عرض قلم و نسبت مساحت پیش‌زمینه به پس‌زمینه، استفاده می-شود.

در این فصل روش‌هایی که تا به حال برای تشخیص ارائه شده بررسی شد. به طور کلی کارهای انجام شده با تمرکز بر روی لبه و اطلاعات مربوط به خواص بافت‌های متنی و اطلاعات فراوانی می-باشد. اطلاعات لبه ممکن است با استفاده از عملگر سوبل، لبه یاب کنی یا روش‌های دیگر استخراج شود. اطلاعات فراوانی با استفاده از تبدیل فوریه یا تبدیل موجک و روش‌های دیگر به دست می‌آید. معمولا رنگ، نسبت ابعاد و ویژگی‌های متن نیز مورد بهره‌برداری قرار می‌گیرد. محققان از تکنیک‌ها و تقسیم بندی‌های ساده و از شبکه‌های عصبی پیچیده برای طبقه‌بندی متن در صحنه‌ها استفاده می-

[1] Gaussian mixture model(GMM)
[2] Support vector mashin

کنند. کارهای مرور شده نشان می‌دهد که، اگر چه موفقیت‌هایی در استخراج متن حاصل شده ولی مسایل حل نشده زیادی وجود دارد. در مورد روش‌های مبتنی برناحیه، سرعت به نسبت پایین است و کارایی به جهت ترازبندی متن وابسته است. از طرف دیگر روش‌های مبتنی بر مولفه متصل بدون داشتن اطلاعات اولیه‌ای از مکان و اندازه متن، نمی‌توانند با دقت مولفه‌های متن را جدا کنند. علاوه بر این، به دلیل وجود اجزای غیر متن زیاد، طراحی سریع و مطمئن یک تحلیل‌گر مولفه متصل سخت است. برای غلبه بر مشکلات فوق، می‌توان از یک روش ترکیبی برای تشخیص و مکان‌یابی متون در تصاویر صحنه استفاده کرد و از مزایای هر دو روش بهره گرفت. ولی در عین حال، تعداد ویژگی‌ها نباید خیلی زیاد باشد. در مورد ابزار یادگیری ماشین برای دسته‌بندی نواحی به دو دسته‌ی متن و غیرمتن، ماشین بردار پشتیبان به دلیل آموزش سریع، آسان‌تر، نیاز به نمونه‌های آموزشی کمتر و توانایی تعمیم بیشتر، نسبت به دیگر ابزار مانند شبکه های عصبی، درخت‌های تصمیم و الگوریتم Adaboost، از کارایی بیشتری در تشخیص متون برخوردار است.[۷]

فصل سوم

تئوری

۳ - تئوری

در این فصل به معرفی مفاهیم استفاده شده در پژوهش می‌پردازیم.

۳-۱ ماشین بردار پشتیبان

ماشین بردار پشتیبان از روش‌های یادگیری با نظارت[1] است که برای طبقه‌بندی[2] استفاده می‌شود. ماشین بردار پشتیبان در سال 1992 توسط vapnik و همکارانش بر پایه تئوری یادگیری آماری[3] معرفی شد. در اصل ماشین بردار پشتیبان یک الگوریتم برای ماکزیمم کردن یک تابع ریاضی با توجه به مجموعه داده‌های موجود می‌باشد [۲۰]. در طبقه‌بندی به کمک ماشین بردار پشتیبان هدف جداسازی مجموعه داده‌های دو کلاس است. اگر هر داده را به صورت یک بردار p بعدی در نظر بگیریم داده‌های دو کلاس را می‌توان با یک ابر صفحه در فضای p بعدی جدا کرد. در این صورت عمل جداسازی خطی نامیده می‌شود. ابر صفحه‌های زیادی وجود دارند که می‌توانند داده‌ها را جدا کنند. مفهوم آموزشی که اشیا بتوانند به عنوان نقاط در یک فضای با ابعاد بالا دسته‌بندی شوند و پیدا کردن خطی که آن‌ها را جدا کند منحصر به فرد نیست. آنچه ماشین بردار پشتیبان را از سایر جداکننده‌ها متمایز می‌کند چگونگی انتخاب ابرصفحه است. در ماشین بردار پشتیبان ماکزیمم کردن حاشیه بین دو کلاس مدنظر است. بنابراین ابر صفحه‌ای را انتخاب می‌کند که فاصله‌ی آن از نزدیک‌ترین داده‌ها در هر دو طرف جدا کننده‌ی خطی، ماکزیمم باشد. اگر چنین ابر صفحه‌ای وجود داشته باشد، ابر صفحه ماکزیمم حاشیه[4] شناخته می‌شود[۲۰].

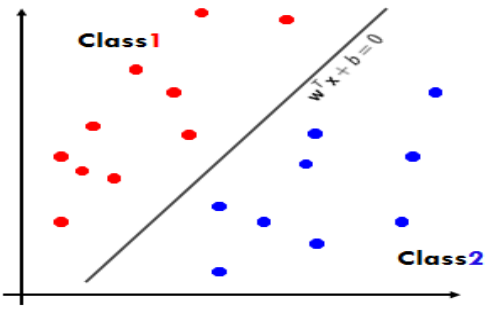
[1] Supervised learning
[2] classification
[3] Statistical learning theory
[4] Maximum margin hyper plan

تابع تصمیم‌گیری برای جدا کردن داده‌ها با یک زیرمجموعه‌ای از مثال‌های آموزشی که بردار پشتیبان نزدیک‌ترین داده آموزشی به ابر صفحه جدا کننده نامیده می‌شوند، تعیین می‌شوند. در واقع ابر صفحه بهینه در ماشین بردار پشتیبان، جداکننده‌ای بین بردارهای پشتیبان است. در صورت استفاده مناسب از *svm* این الگوریتم قدرت تعمیم خوبی خواهد داشت[۲۰].

۳-۱-۱ شبکه ماشین بردار پشتیبان برای سیستم‌های خطی جداپذیر

فرض کنید نمونه های آموزش بصورت $\{x_1, x_2, \dots, x_n\}$ باشند در این حالت y بصورت معادله(۳-۱)تعریف می شود.

$$y_i = \begin{cases} 1 & if\ x_i\ in\ class\ 1 \\ -1 & if\ x_i\ in\ class\ 2 \end{cases} \qquad (۱-۳)$$



شکل ۳-۱ : صفحه جداساز بهینه با حداکثر مقدار حاشیه

خط جداسازی که تمام داده ها را جدا می کند بصورت معادله (۳-۲) می باشد. این خط در شکل (۳-۱) نمایش داده شده است.

$$w^T x + b = 0 \qquad (۲-۳)$$

تابعی بصورت معادله (۳-۱) تعریف می شود.

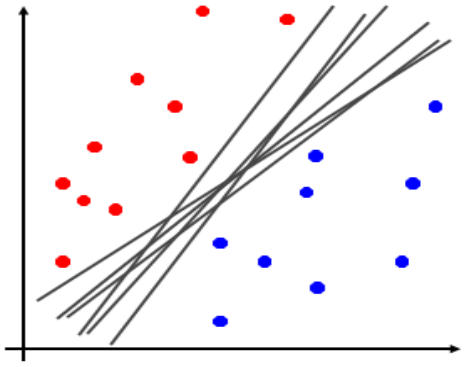
$$f(x) = sgn(w^T x + b) \qquad \begin{matrix}(w^T x_i) + b > 0 \;\; if \; y_i = 1 \\ (w^T x_i) + b < 0 \;\; if \; y_i = -1\end{matrix} \qquad (۳-۳)$$

تعداد زیادی انتخاب برای w و b وجود دارد که هر کدام یک خط جداساز مانند شکل (۳-۲) می‌دهد. اما تنها یک انتخاب بهینه برای خط جداساز وجود دارد[۲۰].

بر اساس شرایط بیان شده، زمانی مجموعه‌ای از نقاط به صورت بهینه با یک صفحه جداسازی می-شوند که :

۱- بدون اشتباه در کلاس مربوط به خود قرار گرفته باشند.

۲- فاصله بین نزدیکترین نقاط هر کلاس داده تا صفحه جدا کننده بیشینه باشد.



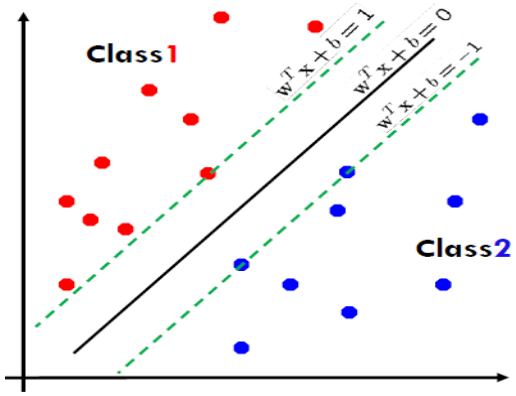
شکل ۳-۲ : خطوط جداساز مختلف برای مقادیر مختلف $\boldsymbol{w}$ و $\boldsymbol{b}$

بر این اساس، پارامترهای w و b باید به گونه‌ای محاسبه گردند که دو شرط ذکر شده برقرار باشد.

جهت حل این مسئله و برای کنترل جداپذیری داده‌ها نیز معادله (۳-۴) برای حاشیه بیان می‌گردد

$$w^T x + b \begin{cases} \geq 1 & for \; y_i = 1 \\ \leq -1 & for \; y_i = -1 \end{cases} \qquad (۴-۳)$$

در شکل (۳-۲) معادلات در نظر گرفته شده برای حاشیه‌ها و صفحه جداساز بهینه مشاهده می‌شود.



شکل ۳-۳: صفحه جداساز و حاشیه‌ها

جهت معرفی صفحه جداسازی که از بیشترین حاشیه ممکن برخوردار باشد، سعی می‌شود تا فاصله بین دو حاشیه در نظر گرفته شده، بیشینه گردد. برای محاسبه فاصله این دو حاشیه و بیشینه نمودن آن از معادله (۳-۵) استفاده می‌شود.

$$M = \frac{|(w^T x+b-1)-(w^T x+b+1)|}{\|w\|} = \frac{2}{\|w\|} \qquad (۵-۳)$$

در این رابطه $\|w\|$ نرم تابع نامیده می‌شود. بر اساس خروجی محاسبه شده از معادله (۳-۱)، اگر $\frac{2}{\|w\|}$ بیشینه گردد، حاشیه مورد نظر ماکزیمم خواهد شد. اما برای سادگی کار می‌توان مقدار به‌دست آمده را معکوس نموده و آن را کمینه نمود که در این حالت به صورت $\frac{1}{2}w^T w$ نوشته خواهد شد. بر اساس شرایط بیان شده در حالت کلی جهت بیشینه نمودن فاصله حاشیه‌ها و یافتن بهینه‌ترین ابر صفحه جداساز از معادله (۳-۶)استفاده می‌شود.

$$min_{w,b}\ \frac{1}{2}w^T w$$

$$subject\ to \quad y_i(w^T x + b - 1) \geq 1 \qquad (۶-۳)$$

در اینجا هدف کمینه سازی تابع $f(x)$ با توجه به محدودیت $g(x) \geq 0$ می باشد، در نتیجه تابع لاگرانژ $L(x,\alpha) = f(x) - \alpha g(x)$ با در نظر گرفتن $\alpha \geq 0$ کمینه خواهد شد. با جاگذاری $f(x)$ و $g(x)$ از معادله (۳-۶) به معادله (۳-۷) می رسیم.

$$L_p(w,b,\alpha) = \frac{1}{2}w^T w - \sum_{i=1}^{N}\alpha_i \left\{ y_i \left[w^T x_i + b \right] - 1 \right\} \qquad (۳-۷)$$

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w_0 = \sum_{i=1}^{N}\alpha_i x_i y_i$$

اگر از رابطه معادله (۳-۷) نسبت به w و b مشتق جزئی گرفته شده و مساوی صفر قرار داده شود، مقدار بهینه W به دست خواهد آمد که این کار در معادله (۳-۸) انجام شده است.

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i y_i = 0 \qquad (۳-۸)$$

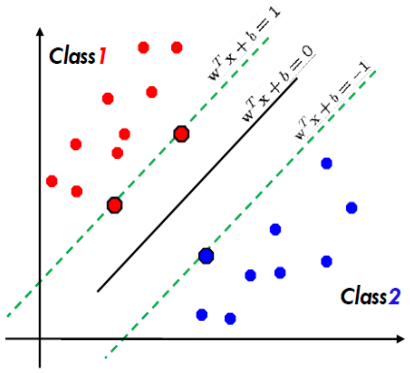
حال اگر مقدار w به دست آمده از معادله (۳-۸) در معادله (۳-۷) قرار داده شود، معادله اساسی ماشین برداری به صورت رابطه معادله (۳-۹) معرفی می‌شود. بنابراین، هدف در ماشین برداری حل معادله معادله (۳-۹) با توجه به محدودیت‌های مشخص شده است. در ماشین برداری سیستم های خطی جداپذیر، مقدار ضریب لاگرانژ باید بزرگتر از صفر باشد[۲۱].

$$\text{Max} \quad L_d(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j x_i{}^T x_j$$
$$\text{S.t} \begin{cases} \alpha_i \geq 0 \\ \sum_{i=1}^{N} \alpha_i y_i = 0 \end{cases} \qquad (۹-۳)$$

مقدار بهینه b نیز از طریق رابطه $b = y_i - w^T x_i$ و میانگین‌گیری از تمامی مقادیر به دست آمده، محاسبه می‌شود. معادله کلی محاسبه مقدار بهینه b را می‌توان به صورت معادله (۳-۱۰) بیان نمود.

$$b_0 = \frac{1}{N} \sum_{S=1}^{N} \left(y_S - w^T x_S \right) \qquad (۱۰-۳)$$

با حل مسئله بهینه‌سازی معادله (۳-۹)و استفاده از معادله (۳-۱۰) می‌توان به بهینه‌ترین صفحه جداساز دست یافت. ابر صفحه جداساز بهینه در شکل (۳-۴) نشان داده شده است.

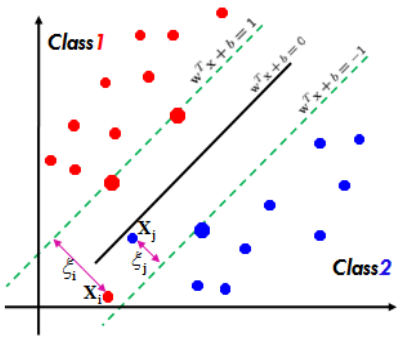


شکل ۳-۴: صفحه جداساز بهینه

نکته قابل توجه در شکل (۳-۴) داده‌هایی است که روی حاشیه‌ها قرار گرفـته‌اند، این داده‌ها همان بردارهای پشتیبان هستند که ماشین برداری برای طبقه‌بندی صحیـح داده‌ها از آن‌ها استفاده می‌کند[۲۱].

۳-۱-۲ شبکه ماشین بردار پشتیبان برای سیستم‌های خطی جداناپذیر

گاهی در سیستم‌های خطی شرایطی ایجاد می‌شود که تعدادی از داده‌ها در کلاس مربوط به خود قرار نمی‌گیرند که این حالت در شکل (۳-۵) نمایش داده شده است. در چنین شرایطی برای داده‌هایی که در طرف دیگر مرز قرار می‌گیرند خطایی به پارامتر $\xi_i \geq 0$ در نظر می‌گیریم که اگر داده به آن اندازه جابجا شود به طرف خود بر می‌گردد.



شکل ۳-۵:سیستم‌های خطی جداناپذیر

در نتیجه معادله جدید بصورت معادله (۳-۱۱) نوشته می شود.

$$\begin{aligned} &min_{w,b} \ \frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i \\ &S.t \quad y_i(w^T x + b) \geq 1 - \xi_i \end{aligned} \qquad (۱۱-۳)$$

در رابطه فوق، C ضریب موازنه جهت بیشینه‌نمودن حاشیه‌ها و کمینه‌سازی خـطای تابع است. در این حالت معادله اساسی بصورت معادله (۳-۱۲) می باشد.

$$L_p(w,b,\xi,\alpha,\beta) = \frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\left\{y_i\left[w^T x_i + b\right] - 1 + \xi_i\right\} - \sum_{i=1}^{N}\beta_i\xi_i \qquad (۱۲-۳)$$

لاگرانژ کلاسیک دوگانه قادر است مسئله اولیه معادله (۳-۱۲) را به مسئله دوگانه آن تبدیل کند. مسئله دو گانه این رابطه، توسط معادله (۳-۱۳) تعریف می‌شود.

$$max\ \ w(\alpha,\beta) = max_{\alpha,\beta}(min_{w,b,\xi} L(w,b,\alpha,\xi,\beta)) \qquad (۱۳-۳)$$

اگر از معادله (۳-۱۳) نسبت به w، b و ξ_i مشتق گرفته شده و مساوی صفر قرار داده شود، مقادیر معادله (۳-۱۴) به دست می‌آیند.

$$\frac{\partial L}{\partial w} = 0 \Longrightarrow w = \sum_{i=1}^{N} \alpha_i y_i x_i \qquad (۱۴-۳)$$

$$\frac{\partial L}{\partial \xi} = 0 \Rightarrow \alpha_i + \beta_i = C$$

با قرار دادن این روابط در معادله (۳-۱۲)، معادله اساسی ماشین برداری در حالت خطی جداناپذیر به دست می‌آید که به صورت معادله (۳-۱۵) خواهد بود.

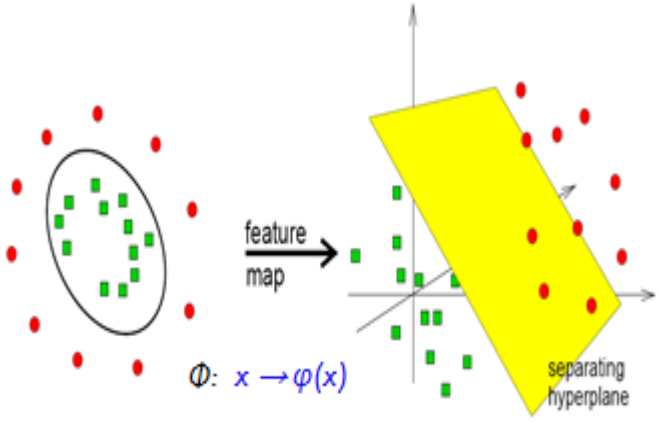
$$\text{Max}\quad L_d(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j x_i^T x_j$$

$$\text{S.t}\quad \begin{cases} 0 \le \alpha_i \le C \\ \sum_{i=1}^{N} \alpha_i y_i = 0 \end{cases} \qquad (۱۵-۳)$$

همان‌گونه که مشاهده می شود، تابع هدف سیستم‌های جداناپذیر خطی مشابه با سیستم‌های جداپذیر خطی است و تنها تفاوت این دو، اصلاح کران‌های ضرایب لاگرانژ می‌باشد. پارامتر C نیز که قابلیت کنترل ظرفیت اضافی طبقه‌بندی کننده را مشخص می‌سازد، در این سیستم‌ها باید انتخاب گردد.[۲۰]

۳-۱-۳ شبکه ماشین بردار پشتیبان غیرخطی

در بسیاری از موارد داده ها بصورت خطی جدا پذیر نیستند که در شکل (۳-۶) نمونه ای نمایش داده شده است.در این حالت از تابعی استفاده می شود که داده های آموزشی را به فضای دیگر منتقل کند که در فضای جدید امکان تفکیک خطی کلاس ها وجود داشته باشد. این تابع کرنل نام دارد.



شکل ۳-۶: نحوه جداسازی داده ها در حالت غیر خطی

در حالتی که داده بصورت خطی جداپذیر باشند از معادله (۳-۱۶) استفاده می شود ولی در حالت غیر خطی این معادله به معادله (۳-۱۷) تبدیل می شود.

$$w = \sum_{i=1}^{N} y_i \alpha_i x_i \qquad (۳-۱۶)$$

$$w = \sum_{i=1}^{N} y_i \alpha_i \varphi(x_i) \qquad (۳-۱۷)$$

در این حالت تابع لاگرانژ بصورت معادله (۳-۱۸) می باشد.

$$L_d(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \emptyset_i{}^T \emptyset_j \qquad (۳-۱۸)$$

تابع کرنل بصورت معادله (۳-۱۹) تعریف می شود.

$$K(x_i, x_j) = \emptyset^T(x_i)\emptyset(x_j) \quad (۱۹-۳)$$

توابع کرنل شناخته‌شده در جدول(۳-۱) آورده شده‌اند.

جدول ۳-۱ : چند نوع تابع کرنل

نوع طبقه بندی	تابع کرنل
تابع خطی از درجه p	$K(x_i, x_j) = ({x_i}^T, x_j)^p$
تابع چند جمله ای از درجه p	$K(x_i, x_j) = ({x_i}^T x_j + 1)^p$
تابع گوسی	$K(x_i, x_j) = e^{\frac{\|x_i - x_j\|^2}{2\sigma^2}}$
تابع پرسپترون چند لایه	$K(x_i, x_j) = \ tanh(\gamma {x_i}^T x_j + \mu)$
تابع دریکله برای مسائل شرایط	$K(x_i, x_j) = \frac{sin((n+1/2)(x_i - x_j))}{2sin((x_i - x_j)/2)}$

در نهایت تابع اساسی بصورت معادله (۳-۲۰) می باشد.

$$\text{Max } L_d(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j K(x_i, x_j)$$

$$\text{Subject to } \begin{cases} 0 \le \alpha_i \le C \\ \sum_{i=1}^{N} \alpha_i y_i = 0 \end{cases} \quad (۲۰ -۳)$$

۳-۱-۴ ماشین بردار پشتیبان چند کلاسه

با گسترش شبکه ماشین بردار پشتیبان برای حالت دو کلاسه که در بالا معرفی شد می توان از آن برای جداسازی چند کلاسه نیز استفاده کرد. دو روش برای جداسازی چند کلاسه با استفاده از ماشین بردار پشتیبان وجود دارد که عبارتند از :

روش یکی در مقابل همه [۱]

روش یکی در مقابل یکی [۲]

در هر دو روش مسئله طبقه‌بندی چند کلاسه به چند مسئله طبقه‌بندی دو کلاسه شکسته می-شود. در روش اول اگر تعداد کلاس‌های موجود برابر با P باشد می بایست P شبکه را آموزش داد و P صفحه جدا کننده طراحی کرد به طوری که در هر بار آموزش شبکه، یک کلاس در مقابل مابقی قرار گیرد. روش دوم نیز همانند روش اول بوده با این تفاوت که کلاس ها دو به دو با هم مقایسه می شوند و مرز بین هر دو کلاس متوسط یک طبقه‌بند دو کلاسه تعیین می‌گردد.[۲۰]

۳-۱-۵ نقاط قوت و ضعف شبکه ماشین بردار پشتیبان

۳ نقطه قوت برای شبکه ماشین بردار پشتیبان برشمرده‌اند:

۱- آموزش شبکه ماشین بردار پشتیبان نسبتاً آسان است.

۲- به کمک کرنل‌های غیر خطی امکان انتقال داده‌ها به فضای بالاتر وجود دارد و به همین دلیل امکان طراحی مرزهای غیرخطی توسط ماشین بردار پشتیبان وجود دارد.

۳- موازنه بین پیچیدگی طبقه‌بند و میزان خطا قابل کنترل است.

[1] One Against All (OAA)
[2] One Against One (OAO)

مشکل ماشین برداری که به عنوان نقطه ضعف این ماشین بیان شده، نیاز این ماشین به انتخاب کرنل مناسب است زیرا در صورت عدم مناسب بودن آن، نتایج ارائه شده رضایت بخش نخواهند بود.[۲۱] همچنین مشکل دیگر دو کلاسه بودن آن است که در حالت چند کلاسه باید چندین مدل آموزش داد.

۳-۲ تخمین پارزن

روش پارزن در گروه روش‌های ناپارامتری قرار می‌گیرد. اساس این روش قانون بیز می‌باشد با این تفاوت که به دنبال تخمین پارامترهای یکPDF خاص برای کل داده‌ها نیستیم، بلکه برای داده‌ها به صورت محلیPDF تخمین زده و بر اساس آن، داده‌ها را در کلاس‌های مختلف دسته‌بندی می‌کنیم. این روش تعبیری برای تخمین محلی تابع چگالی احتمال است.[۲۲] به کمک این روش می‌توان از روی مجموعه داده‌های تولید شده از یک فرایند آماری چگالی احتمال توزیع را تخمین زد.

۳-۲-۱ روش پنجره پارزن

اگر n داده داشته باشیم و k، تعداد داده‌هایی از n داده باشد که در حجم V به مرکز x قرار گرفته-اند، آن‌گاه یک تقریب خوب برای چگالی احتمال x به صورت معادله (۳-۲۱) می‌باشد:

$$\hat{p}(x) = \frac{k}{nV} \qquad (۲۱-۳)$$

در روش پنجره پارزن، حجم V را ثابت فرض می‌کنیم. در این حالت K، تعداد داده‌های قرار گرفته در این حجم را نشان می‌دهد.[۲۲]

برای درک راحت‌تر پنجره پارزن، حجم V را به صورت یک ابر مکعب فرض می‌کنیم، با فرض اینکه داده‌ها در فضای ۳ بعدی قرار دارند و تعریف h، طول لبه‌های یک ابر مکعب، حجم ابر مکعب با رابطه زیر داده می‌شوند:

$$V = h^3 \qquad (۲۲-۳)$$

با تعریف تابع پنجره به صورت:

$$\varphi(u)\begin{cases}1 & \text{if } |u_j| \le \frac{1}{2} \\ 0 & \text{otherwise}\end{cases} \qquad (۲۳-۳)$$

(که در آن u_j مولفه u در جهت محور j ام است) می‌توانیم تعداد نمونه‌های درون این ابر مکعب را با رابطه (۳-۲۴) نشان دهیم:

$$k = \sum_{i=1}^{n} \varphi(\frac{x_i - x}{h})) \qquad (۲۴-۳)$$

در این صورت با قرار دادن رابطه فوق در رابطه (۳-۲۱) داریم.

$$p(x) = \frac{1}{n}(\sum_{i=1}^{n} \frac{1}{V} \varphi(\frac{x_i - x}{h})) \qquad (۲۵-۳)$$

تابع فوق به عنوان یک تابع چگالی احتمال هم باید مثبت باشد و هم انتگرال آن بر روی x، برابر ۱ گردد.نشان داده می‌شود که تابع پنجره، $\varphi(u)$، اگر خود یک تابع چگالی باشد آنگاه p(x) نیز تابع چگالی احتمال می‌گردد. بنابراین φ را با شرایط زیر انتخاب می‌کنیم:

$$\varphi(u) > 0 \qquad (۲۶-۳)$$

$$\int \varphi(u)\, du = 1 \qquad (۲۷-۳)$$

توابع زیادی وجود دارند که شرایط (۳-۲۶) و (۳-۲۷) را برآورده می‌کنند. این آزادی در انتخاب φ، این امکان را به وجود می‌آوردکه تابع پنجره را به گونه‌ای تعریف کنیم که به ازای نقاط نزدیک به

مرکز مقادیر بزرگتری را برگرداند. درحقیقت با این کار داده‌های نزدیک به مرکز پنجره با وزن بیشتری در تابع چگالی احتمال حضور می‌یابند و به نوعی به آن‌ها اهمیت بیشتری در تصمیم‌گیری می‌دهیم.

انتخاب h نقش مهمی در تابع چگالی احتمال p(x) دارد. اگر h خیلی بزرگ انتخاب گردد، p(x) به آرامی تغییر می‌کند و بسیار هموار می‌گردد. همچنین با انتخاب h های کوچک،p(x) به صورت پالس-های تیز به مرکز داده‌ها خواهند بود.

در این روش برای دسته‌بندی داده نامعلوم در یکی از کلاس‌ها، ابتدا حجم V به مرکز این داده جدید را در نظر می‌گیریم. سپس p(x) را برای هر کلاس و با استفاده از داده‌های آموزشی آن کلاس در حجم V، با معادله (۳-۲۱) محاسبه می‌کنیم. داده جدید را در کلاسی دسته‌بندی می‌کنیم که دارای p(x) بزرگتری باشد.[۲۲]

۳-۳ مدل مخلوط توابع گوسی[1]

مخلوط توابع گوسی، تابع چگالی احتمال ۲بردارهای ویژگی را توسط ترکیبی از چند تابع گوسی وزن‌دار مدل می‌کند. در حقیقت GMM مجموعی از K تابع گوسی متفاوت است که هر کدام از این توابع دارای یک وزن خاص هستند. برای هر بردار ویژگی $X \in R^d$ مدل تلفیقی به صورت رابطه (۳-۲۸) تعریف می‌شود [۲۳].

$$f(X|\theta) = \sum_{j=1}^{K} \alpha_j \, N(X|\mu_j, \Sigma_j) \qquad (۳-۲۸)$$

$$N(X|\mu_j, \Sigma_j) = \frac{1}{2\pi^{d/2}|\Sigma_j|^{1/2}} exp\left\{-\frac{1}{2}(X-\mu_j)^T \Sigma_j^{-1}(X-\mu_j)\right\} \qquad (۳-۲۹)$$

[1] Goassian mixture model
[2] PDF

پارامتر θ مجموعه پارامترهای لازم برای تعریف GMMاست و به صورت $\theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^{K}$ نوشته می‌شود بطوریکه:

$\alpha_j > 0$ وزن تابع گوسی jام می‌باشد بطوریکه $\sum_{j=1}^{K} \alpha_j = 1$ است.

$\mu_j \in R^d$ میانگین تابع گوسی jام است.

$\sum j$ ماتریس کواریانس تابع گوسی j ام است که دارای ابعاد d×d می‌باشد(که d ابعاد داده‌های مدل شده می‌باشد).

برای مجموعه‌ای از بردارهای ویژگی X1، X2، ... و Xn به ابعاد d تخمین ماکزیمم احتمال[1] از پارامترهای مجهول θ برای تابع احتمال L(.) برابر است با:

(۳۰-۳)

$$\theta_{ML} = \arg\max_{\theta} L(\theta | X_1, X_2, ..., X_n) = \arg max_{\theta} \sum_{i=1}^{n} \log f(X_i | \theta)$$

الگوریتم‌های زیادی برای تخمین پارامترهای مدل GMM یعنی θ_{ML} وجود دارد. یکی از متداول-ترین الگوریتم‌ها در این زمینه، الگوریتم EM[2] است[۲۳]. EM یک الگوریتم بازگشتی است که از یک مقدار اولیه برای θ شروع کرده و در هر بار تکرار این بردار را به‌روز رسانی می‌کند تا همگرا شود. بروز رسانی پارامترها در تکرار t ام در دو مرحله پیش‌بینی[3] (E) و ماکزیمم سازی[4] (M) صورت می‌پذیرد که به صورت زیر می‌باشند[۲۴].

[1] Maximum likelihood
[2] Expectation Maximization
[3] Expectation
[4] Maximization

۳-۳-۱ مرحله پیش‌بینی

در این مرحله احتمال‌های پسین Wij محاسبه می‌شود. این احتمال بیان می‌دارد بردار ویژگی Xi با چه احتمالی جزئی از تابع گوسی j ام در تکرار t ام است.

$$W_{ij} = \frac{\alpha_j^t f(X_i|\mu_j^t, \Sigma_j^t)}{\sum_{l=1}^{K} \alpha_l^t f(X_i|\mu_l^t, \Sigma_l^t)} \quad (۳۱-۳)$$

۳-۳-۲مرحله ماکزیمم‌سازی

در این مرحله پارامترهای تکرار t+1 ام مطابق روابط زیر بدست می‌آیند.

$$\alpha_j^{t+1} = \frac{1}{n} \sum_{i=1}^{n} W_{ij} \quad (۳۲-۳)$$

$$\mu_j^{t+1} = \frac{\sum_{i=1}^{n} W_{ij} X_i}{\sum_{i=1}^{n} W_{ij}} \quad (۳۳-۳)$$

$$\Sigma_j^{t+1} = \frac{\sum_{i=1}^{n} W_{ij} \left(X_i - \mu_j^{t+1}\right)\left(X_i - \mu_j^{t+1}\right)^T}{\sum_{i=1}^{n} W_{ij}} \quad (۳۴-۳)$$

ماتریس‌های کواریانس توابع می‌توانند قطری فرض شوند. در این صورت ویژگی‌ها مستقل از هم می‌باشند[۲۳]. اگر ماتریس کواریانس‌ها را قطری درنظر بگیریم، پیچیدگی محاسباتی برای تخمین پارامترها در *روش* GMM کاهش یافته و سرعت این روش بالاتر می‌رود.

در تکرار اول روش بهینه سازی EM نیاز به یک نقطه شروع دارد. یکی از روش‌های بدست آوردن نقطه شروع استفاده از خوشه‌بندی به روش K-means است. یعنی داده‌ها به روش K-means دسته‌بندی شده و میانگین داده‌ها در هر دسته و برچسب داده‌ها استخراج می‌شود. توسط داده‌هایی که در هر دسته قرار می‌گیرند می‌توان ماتریس کواریانس ویژگی‌ها در هر دسته را بدست آورد. البته می‌توان مراکز خوشه‌ها را در شروع به روش تصادفی نیز انتخاب کرد. در این روش K بردار ویژگی به

عنوان مراکز دسته‌ها درنظر گرفته می‌شوند. وزن‌های تمامی توابع بطور مساوی $\frac{1}{K}$ قرار داده می‌شوند. ماتریس کواریانس دسته‌ها برای همه یکسان و قطری درنظر گرفته می‌شود. عنصر j ام روی قطر اصلی ماتریس کواریانس برابر با واریانس ویژگی j ام برای تمامی داده‌های آن دسته است.

۳-۴ فیلتر گابور

تبدیل گابور، در دسته تبدیل‌های موجک با پنجره مدوله شده قرار می‌گیرد. با استفاده از تبدیل موجک دو بعدی گابور می‌توان ویژگی‌های جهت‌دار صحنه را در مقیاس مختلف استخراج کرد. فیلتر های گابور، فیلتر های میان‌گذری هستند که انتخاب جهت و فرکانس را با هم دارند و دارای درجه تفکیک ایده ال در هر دو فضای مکانی و فرکانسی می باشند[۲۵].

فیلتر گابور با معادلات (۳-۳۵) و (۳-۳۶) و (۳-۳۷) مشخص می شود.

$$G(x,y,f,\theta) = \exp\left\{\frac{-1}{2}\left[\frac{\acute{x}^2}{\delta_x{}^2} + \frac{\acute{y}^2}{\delta_y{}^2}\right]\right\}\cos(2\pi f\acute{x}) \qquad (۳۵-۳)$$

$$\acute{x} = xsin\theta + ycos\theta \qquad (۳۶-۳)$$

$$\acute{y} = xcos\theta - ysin\theta \qquad (۳۷-۳)$$

پارامتر f فرکانس سینوسی فیلتر در راستای جهت θ از محور عمود و پارامترهای δ_x و δ_y متناظراً ثابت های فضای پنجره گوسی در راستای x و y هستند.

تبدیل موجک گابور به عنوان ابزاری برای استخراج ویژگی در زمینه‌هایی همچون طبقه‌بندی، بازیابی صحنه ، تشخیص قلم و تحلیل بافت و ... مورد استفاده قرار می‌گیرد[۲۵].

فصل چهارم

روش پیشنهادی

۴- روش پیشنهادی

سیستم تشخیص متن پیشنهاد شده در این پایان‌نامه، بر مبنای ارتباط میان اجزای متنی کار می‌کند. این سیستم با در نظر گرفتن برخی ویژگی‌های متنی به حذف اجزای غیر متنی می‌پردازید.

الگوریتم پیشنهادی از چهار مرحله اصلی تشکیل می‌شود. در مرحله اول به کمک ویژگی تغییر گرادیان در لبه‌های صحنه ، اقدام به استخراج نواحی کاندید متن می‌نماییم(بخش ۴-۱). در مرحله بعد از میان کاندیدها، با توجه به این واقعیت که اجزاء تشکیل دهنده یک سطر متن در صحنه دارای راستا و ارتفاع تقریبا یکسانی هستند به گروه بندی نواحی استخراج شده می‌پردازیم(بخش ۴-۲). در مرحله سوم از ویژگی‌های هیستوگرام اندازه گرادیان و زاویه گرادیان در نواحی استخراج شده استفاده نموده تا نواحی غیر متنی را فیلتر نماییم(بخش ۴-۳). در آخر با قرار دادن معیار فاصله بر مبنای عرض نواحی متنی یافت شده و استفاده از افکنش افقی نتیجه بهبود داده می‌شود(بخش ۴-۴).

از مزیت‌های روش معرفی شده:

۱) نسبت به تغییرات رنگ و اندازه و قلم متون پایدار است .

۲) سیستم برای دو زبان نوشتاری انگلیسی و فارسی پاسخ‌گو خواهد بود.

۴-۱ استخراج نواحی کاندید متن

ما به دنبال اجزایی از صحنه هستیم که امکان حضور متن در آن‌ها بیشتر باشد و ویژگی‌هایی از جمله لبه‌های صحنه ، رنگ‌های یکسان در صحنه ، بخش‌های متصل در صحنه توجه ما را به خود جلب کرد.

اولین گام تفکیک نواحی صحنه با توجه به رنگ موجود در آنها است. در این زمینه در [۱۹] به گروه‌بندی پیکسل‌های لبه با توجه به رنگ همسایه‌های مجاور پرداخته است. این روش با استفاده از

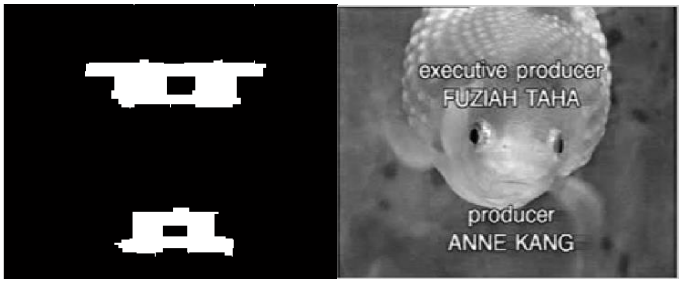
الگوریتم EM به استخراج پارامتر های مخلوط گوسی در فضای رنگ لبه‌های صحنه پرداخته‌است. بدین صورت که ۸ همسایگی مجاور هر پیکسل را مورد بررسی قرار داده است و بیشترین و کمترین شدت رنگ را در سه بعد تحت عنوان یک بردار ویژگی ۶ بعدی در نظر گرفته است و همچنین چون مکان پیکسل‌ها حایز اهمیت است و با توجه به این فرض که نوشته‌ها در راستای افقی هستند از مولفه x پیکسل‌ها صرف‌نظر کرده و مولفه y با سه بار تکرار به ۶ ویژگی قبلی اضافه و یک بردار ویژگی ۹ بعدی را ایجاد کرده است. صحنه با استفاده از این ویژگی‌ها به ۵ لایه تقسیم شده است (تعداد گوسی ها در الگوریتم gmm ۵ است) و سپس به استخراج ناحیه متنی با توجه به لبه‌ها از هر کدام از این لایه‌ها پرداخته است. نتایج تجربی نشان می‌دهد که این روش در استخراج لبه‌های متنی و در ادامه استخراج ناحیه متن مناسب نیست. به این خاطر که نواحی متن از هم جدا شده و در ۵ لایه تقسیم می‌شوند و همچنین در هر بار انجام الگوریتم به خاطر تصادفی بودن فرایند خوشه‌بندی به کمک GMM، نتایج متفاوتی در ۵ لایه برای عکسی یکسان مشاهده می‌شود. این مشکل در صحنه‌های که دارای تنوع رنگ خیلی زیاد (۱۰ رنگ به بالا) و یا صحنه‌ها با تنوع رنگ پایین(کمتر از ۳ رنگ) بیشتر نمایان می‌شود. شکل (۴-۱) این مشکل را نشان می‌دهد.

شکل ۴-۱: نواحی متنی تفکیک شده با استفاده از روش تفکیک نواحی صحنه با توجه به رنگ موجود در آن‌ها

در [۲۶] برای استخراج نواحی متنی از تبدیل موجک سطح یک استفاده شده و در مرحله بعد لبه‌های هر زیر باند با استفاده از یک آستانه فیلتر می‌شود. این آستانه با توجه به ماکزیمم مقدار هیستوگرام و اندازه هر زیر باند تعیین می‌شود. با استفاده از این فرض که در همسایگی ۸*۸ پیکسل‌ها بیشتر از یک چهارم پیکسل‌ها باید لبه باشند، لبه‌های منفرد در هر زیر باند حذف شده و اگر یک پیکسل در هر سه زیر باند مقدار یک را داشته باشد به عنوان پیکسل متنی انتخاب می‌شود. در ادامه با چند عملیات ریخت شناسی بازکردن و بستن [۱] در راستای افقی و عمودی ناحیه متنی به دست می‌آید. نتایج تجربی نشان دهنده آن بود که این روش در صحنه‌ها با پس‌زمینه ساده یا پیچیدگی کم پاسخ مناسبی می‌دهد ولی در صحنه‌ها با پس زمینه پیچیده با لبه‌های بسیار پاسخ مناسبی را شاهد نخواهیم بود. این روش نیز روش کارآمدی به حساب نیامد.

تشخیص ناحیه متنی در شکل (۴-۲) که به نسبت صحنه با پس زمینه ساده‌ای است به درستی صورت گرفته است اما در شکل (۴-۳) با پس زمینه شلوغ تشخیص درست نیست.

[1] closing



شکل۴ -۲ :تشخیص ناحیه متنی صحنه با پس‌زمینه ساده با استفاده از روش تبدیل موجک

شکل ۴-۳: تشخیص ناحیه متنی صحنه با پس‌زمینه غیره ساده با استفاده از روش تبدیل موجک

در ادامه برای یافتن روشی کارآمد با استفاده از لبه یاب Canny لبه‌های صحنه را مشخص می-کنیم. روش‌های مبتنی بر لبه بر این فرض استوارند که صحنه دارای تباین بالا بین بخش‌های متنی و غیرمتنی است. این ایده بر این واقعیت بنا شده است که لبه‌ها یک مشخصه بسیار قوی صرف نظر از رنگ، چگالی و جهت برای تشخیص متن می‌باشند.

آشکارساز لبه Canny به عنوان یک آشکار ساز لبه موفق عمل می‌کند. آشکارساز لبه Canny از فیلترینگ گوسی و الگوی مشتق‌گیری استفاده می‌کند، هر چند در مسئله پیش‌رو به دلیل گستره وسیعی از نواسانات در شدت، تباین و روشنایی در صحنه به ازای یک آستانه مشخص ممکن است برخی لبه‌های ضعیف را از قلم بیندازد و مانع آشکارسازی نویسه‌های متن به هم چسبیده شود.

لبه‌های صحنه استخراج شده در هر سطر مورد پردازش قرارمی‌گیرد. لبه‌هایی که فاصله‌ی بین دو لبه در آن‌ها از یک آستانه کمتر باشد و جهت گرادیان روشنایی در آن‌ها خلاف هم باشد به عنوان دو لبه متنی شناخته شده و ناحیه ما بین آن‌ها به عنوان ناحیه متنی مارک می‌شود. از جمله معایب این روش در مواردی که حروف نزدیک به هم باشند ناحیه بین حروف نیز به عنوان متن انتخاب می‌شود. و

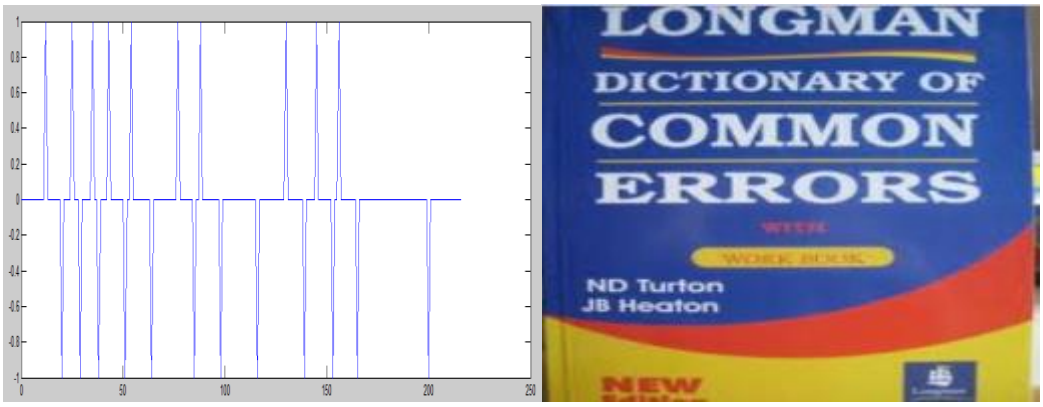
همچنین در مورد متون فارسی لبه‌های دو طرف حروفی مانند "ب" ، "پ"،"ت" و امثال آن در قسمت انتهایی زیاد است و این نواحی تحت عنوان متن شناسایی نمی‌شوند. برای بهبود این روش ما روش زیر را ارائه کردیم.

در ابتدا گرادیان صحنه را استخراج می‌نماییم. در میان لبه‌های موجود در صحنه به دنبال لبه‌های متنی هستیم. با توجه به این واقعیت که در لبه‌های متنی صحنه تغییر در روند مقادیر گرادیان(صعودی و نزولی) را شاهد هستیم و با قرار دادن یک مقدار آستانه (در اینجا ۲۰ در نظر گرفته شده که به صورت تجربی به دست آمده است) لبه‌های صحنه مشخص می‌شود. برای لبه‌های صعودی مقدار مثبت یک و برای لبه‌های نزولی مقدار منفی یک را اختصاص می‌دهیم. مقدار مثبت یک مربوط به زمانی است که در یک سطر مقادیر گرادیان سیر صعودی داشته و در لبه مورد نظر به بعد مقدار گرادیان کاهش می‌یابد. به بیان دیگر شاهد یک قله در مکان لبه هستیم و مقدار منفی یک مربوط به زمانی است که در یک سطر مقادیر گرادیان سیر نزولی داشته و در لبه مورد نظر به بعد مقدار گرادیان افزایش می‌یابد و دره راخواهیم داشت. مقادیر گرادیان در یک سطر صحنه و اختصاص مقادیر مثبت یک و منفی یک در جدول ۴-۱مشاهده می‌شود.

جدول ۴-۱ : تغییرات گرادیان و مقداردهی به قله و دره

گرادیان	۱۱,۵	۱۰۳	۳۷	۱۰۹	۳,۵	۰,۵	۵
مقادیر اختصاص داده شده	۰	۱	۰	۱	۰	۰	۰

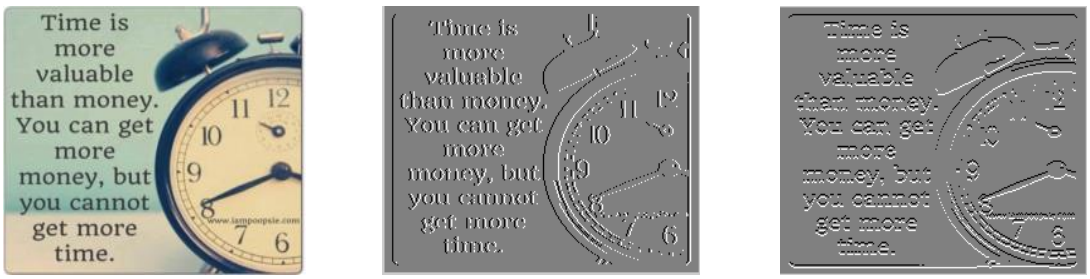
همان‌طور که در جدول۴-۱ مشاهده می‌شود میزان افزایش گرادیان در دو ستون آخر کمتر از میزان آستانه مورد نظر بوده و به همین خاطر به عنوان لبه آشکار نشده است. گراف نشان داده شده در شکل (۴-۴) صعودی و نزولی بودن لبه‌ها در سطر چهارم متن، سطری که کلمه errors در آن قرار دارد، را نمایش می‌دهد.



الف ب

شکل ۴-۴: صعودی و نزولی بودن لبه‌ها الف) تصویر اصلی ب) مقادیر اختصاص داده شده

این فرایند در هر دو راستای x و y بر روی صحنه اعمال می‌شود. نتایج آن در شکل (۴-۵) مشاهده می‌شود.



صحنه اصلی لبه در راستای x لبه در راستای y

شکل ۴-۵ : مقداردهی گرادیان در راستایx و y. از سمت چپ ستون اول صحنه اصلی و ستون دوم لبه در راستای x و ستون سوم لبه در راستای y است.

در مرحله بعدی بر روی توالی۱- و ۱+ های موجود در هر سطر تمرکز می‌کنیم. هدف پیدا کردن ۱- و۱+ هایی است که در فاصله محدودی نسبت به هم باشند که این فاصله محدود توسط یک مقدار آستانه کنترل می‌شود و نشانگر این واقعیت است که دو لبه‌ای می‌توانند متعلق به بخش متنی باشند که فاصله آنها از هم از مقدار مشخصی تجاوز نکند.

در هر سطر صحنه این توالی بررسی شده و هر دو پیکسل لبه که چنین ویژگی داشته باشند وارد فاز بعدی می‌شوند. در فاز بعدی مرکز بین جفت پیکسل‌های مذکور مشخص می‌شود. تا این مرحله مراکز بخش‌های کاندید متن مشخص شده است.

برای استخراج بخش کاندید با این استراتژی عمل می‌شود که در هر سطر به اندازه فاصله دو پیکسل متوالی لبه، از مرکز فاصله گرفته و این ناحیه را با مقادیر پیکسل‌ها در تصویر رنگی پر می-کنیم.

نکته قابل تامل در این مرحله اختلاف رنگ میان پس‌زمینه و پیش‌زمینه است. اگر متن ما دارای رنگی روشن‌تر از پس‌زمینه خود باشد توالی ۱+ به ۱- متن را مشخص نموده و حایز اهمیت خواهد بود و اگر متن ما دارای رنگی تیره تر از پس‌زمینه خود باشد توالی۱- به ۱+ حایز اهمیت خواهد بود. لذا در هر صورت یکی از دو تصویر حاصل مبهم خواهد بود و در تصویر دیگر متن ظاهر می‌شود شکل(۴-۶).

خروجی سیستم پیشنهادی در این مرحله دو صحنه می‌باشد. در صحنه‌های با متون روشن صحنه شماره ۱ و در صحنه‌های با پس‌زمینه روشن صحنه شماره ۲ مناطق کاندید متن را نمایش می‌دهد. شکل(۴-۶) این صحنه‌ها را برای دو صحنه نمونه نشان می‌دهد. ستون اول تصویر اصلی است. ستون دوم توالی ۱+ به ۱- متن را مشخص نموده و ناحیه بین آن را ناحیه متنی در نظر گرفته است. ستون سوم توالی ۱- به ۱+ متن را مشخص نموده و ناحیه بین آن ناحیه متنی در نظر گرفته است.

صحنه شماره ۲ صحنه شماره۱ صحنه اصلی

شکل ۴-۶: نواحی کاندید استخراج‌شده. از سمت چپ ستون اول تصویر اصلی.ستون دوم توالی ۱+ به ۱- متن را مشخص نموده و ناحیه بین آن ناحیه متنی در نظر گرفته است. ستون سوم توالی ۱- به ۱+ متن را مشخص نموده و ناحیه بین آن ناحیه متنی در نظر گرفته است.

در مرحله بعدی اجزای متصل در صحنه‌های به دست آمده را استخراج کرده که بلوک‌های متنی و غیر متنی را برای ما تولید می‌کنند و در این مرحله به دنبال روشی برای جداسازی بلوک‌های متنی از غیر متنی هستیم.

۴-۲ گروه بندی نواحی استخراج شده

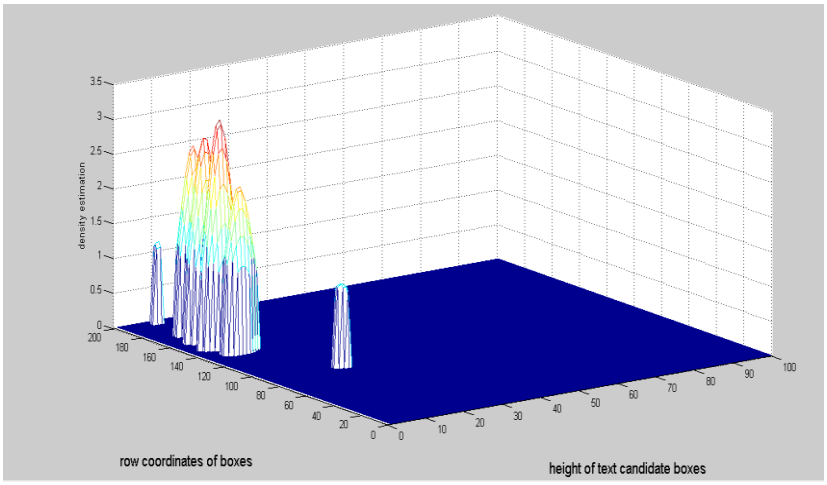
دز این بخش به گروه بندی نواحی استخراج شده پرداخته می‌شود به همین منظور پنجره محیط بر هر ناحیه متصل استخراج می‌شود. حال اقدام به گروه بندی بلوک‌های موجود در صحنه به کمک دو ویژگی می‌کنیم. ویژگی اول مختصات سطری مرکز بلوک‌ها می‌باشد. ویژگی دوم ارتفاع پنجره‌های کاندید متن است که مورد توجه قرار می‌گیرد. همانطور که در شکل (۴-۷) مشخص است نواحی متنی دارای بلوک‌هایی با ارتفاع‌های نزدیک به هم هستند.

شکل ۴-۷: نمایش بلوک‌های متنی با ارتفاع نزدیک به هم در دو صحنه نمونه ی خروجی سیستم پیشنهادی در این مرحله

از هر بلوک کاندید این دو ویژگی استخراج شده و چگالی بلوک‌ها با استفاده از پنجره پارزن در فضای دو ویژگی استخراج شده، تخمین زده می‌شود. قله‌های مرتفع در چگالی مربوط به خوشه‌هایی هستند که از بلوک‌های هم راستا با ارتفاع نزدیک به هم تشکیل شده‌اند. این خوشه‌ها به احتمال زیاد مربوط به خطوط متن در صحنه می‌باشند. با آستانه گذاری بر روی خوشه‌ها، خوشه‌هایی که قله آن‌ها بزرگتر از آستانه تعیین شده باشد به عنوان خوشه متنی کاندید می‌گردد.

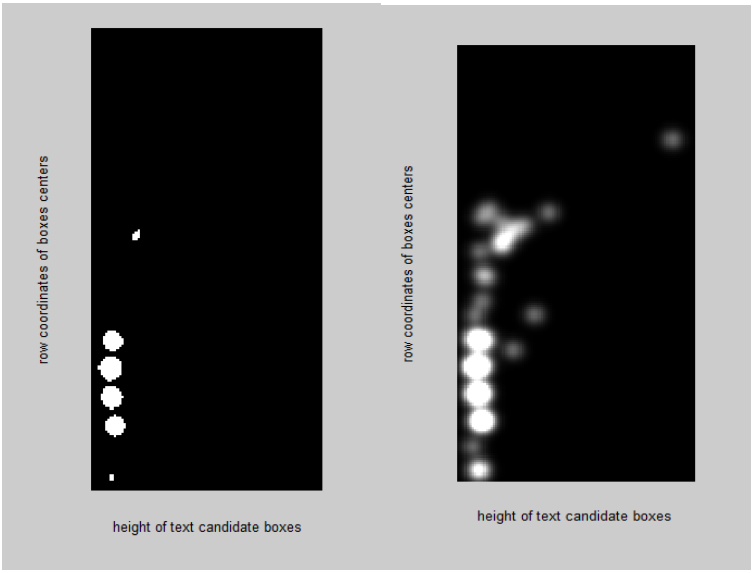
در شکل (۴-۸) خروجی تخمین چگالی بلوک‌ها قابل مشاهده است. این گراف مربوط به صحنه اول از شکل(۴-۷) است که در آن محور x مقادیر ارتفاع بلوک‌ها و محورy مختصات مرکز y بلوک‌ها و محور z مقادیر چگالی تخمین‌زده می‌باشند.

قله‌ای که دارای پیک بزرگتری است نشان دهنده این واقعیت است که فراوانی بلوک‌ها با سطری مشابه و ارتفاع‌های نزدیک به هم در ناحیه‌ای از فضای ویژگی زیاد است. این سطرها به عنوان سطری که مراکز بلوک‌های متنی در آن‌ها قرار دارد در نظر گرفته می‌شوند. در مثال زیر بلوک‌های متنی ما دارای ارتفاعی بین ۲ تا ۲,۷ بوده و از نظر موقعیت y بین ۱۳۵ تا ۲۰۰ می باشند.



شکل ۴-۸ :خروجی تخمین چگالی بلوک‌ها

نتیجه فیلترینگ قله‌های تخمین پارزن در دو بعد در شکل(۴-۹) مشاهده می‌شود که در آن محور x مقادیر ارتفاع بلوک‌ها و محورy مختصاتy مرکز بلوک‌ها می‌باشد.



ب الف

شکل ۴-۹ : قله‌های تخمین پارزن الف) قبل از فیلترینگ ب)بعد از فیلترینگ

با انجام این کار موقعیت خطوط اصلی متن برای ما مشخص می‌شود. و بلوک‌های موجود در راستای این خطوط انتخاب می‌شوند.

در خروجی این بخش برخی از مناطق غیر متنی حذف می‌شود.شکل (۴-۱۰)

الف

ب

شکل ۴-۱۰ : بلوک‌های متنی انتخاب شده الف) قبل و ب) بعد از گروه‌بندی

همانطور که در صحنه خروجی قابل مشاهده است هنوز بلوک‌های غیر متنی در صحنه وجود دارد که در مرحله بعدی تلاش برای حذف آن‌ها صورت می‌گیرد.

۴-۳ استخراج ویژگی

در این بخش به دنبال استخراج ویژگی از بلوک‌های موجود برای حذف بلوک‌های غیر متنی هستیم.

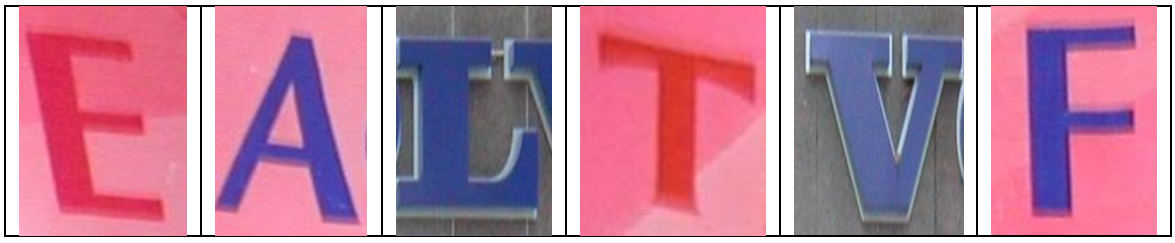
۴-۳-۱ پیشنهاد اول :عرض قلم

در اولین پیشنهاد به سراغ عرض قلم در بلوک‌ها رفته و به محاسبه آن می‌پردازیم. با این استراتژی پیش می‌رویم که عرض قلم در بلوک متنی یکسان است و برای تشخیص از معیارهای پراکندگی استفاده می‌شود. مهمترین معیار پراکندگی واریانس است. واریانس به ما نشان می‌دهد که

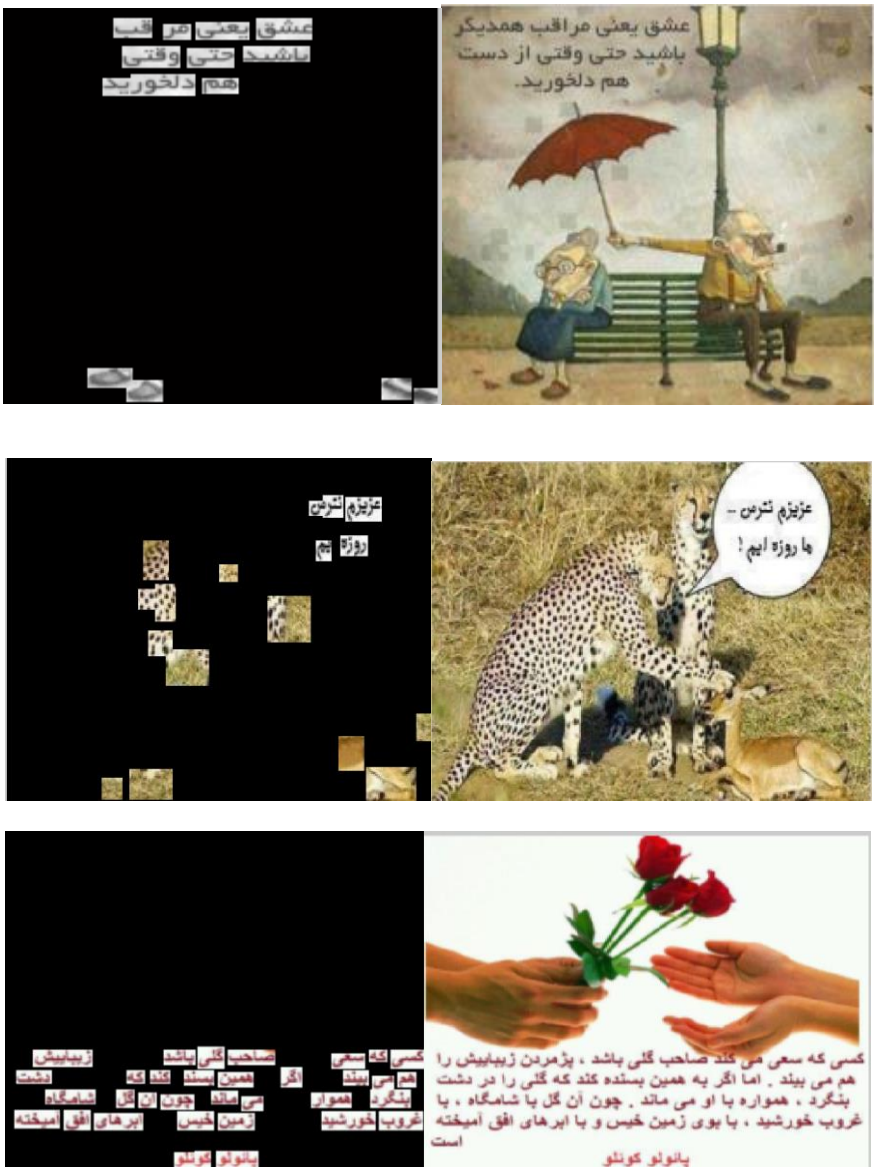
مشاهدات تا چه اندازه در اطراف میانگین قرار دارد. یک عدد کوچک برای واریانس مجموعه داده‌ها نشان دهنده این واقعیت است که داده‌ها در دامنه کوچکی حول میانگین پراکنده شده‌اند و برعکس. همچنین از ویژگی آنتروپی بلوک‌ها نیز به عنوان ویژگی دوم استفاده می‌شود.

با کلاسه بندی دو ویژگی فوق توسط svm بلوک‌های متنی و غیر متنی از هم جدا می‌شوند. مشاهده نتایج نشان‌دهنده آن است که در متون انگلیسی تشخیص به درستی صورت می‌گیرد. اما در متون فارسی در مواقعی که حروف به صورت متصل به هم هستند و همچنین در حروف دندانه‌دار دچار مشکل می‌شود. به این خاطر که عرض قلم در برخی نواحی متنی دارای تغییرات زیادی است، تشخیص به درستی امکان پذیر نمی‌باشد. در جدول (۴-۲) نمونه‌هایی از بلوک‌ها با عرض متفاوت نشان داده شده است.

جدول ۴-۲ : نمونه‌هایی از عرض متفاوت یک حرف در یک بلوک



همچنین بلوک‌های غیر متنی موجود است که ساختاری بسیار مشابه به حروف انگلیسی و فارسی دارند که تشخیص را برای ما مشکل می سازند. در این روش ما ۷۳ درصد تشخیص درست متن و ۶۳ درصد تشخیص درست غیر متن را شاهد بودیم. در شکل(۴-۱۱) نتایج این روش مشاهده می‌شود.



الف ب

شکل۴-۱۱ : مواردی که سیستم به درستی کار کرده الف)تصویر ورودی ب)متن تشخیص داده شده

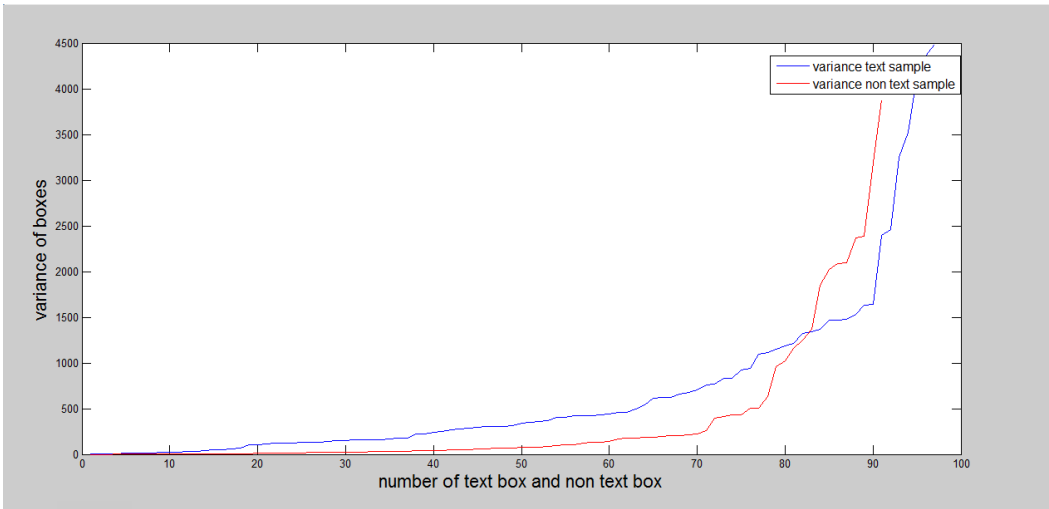
همانطور که مشخص است این روش در تشخیص متون عملکرد مناسبی دارد، اما در صحنه‌ها با پس‌زمینه پیچیده بلوک‌های غیر متنی زیادی را در تصویر نهایی شاهد خواهیم بود. در شکل(۴-۱۲) نمونه‌هایی که سیستم به درستی پاسخ‌گو نبوده است مشاهده می‌شود.

الف ب

شکل ۴-۱۲: نمونه‌هایی که سیستم به درستی پاسخ‌گو نبوده الف)تصویر ورودی ب)متن تشخیص داده شده

۴-۳-۲ پیشنهاد دوم استفاده از واریانس برای استخراج نواحی متنی

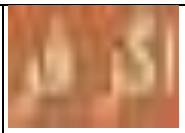
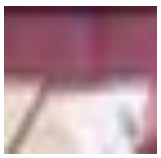
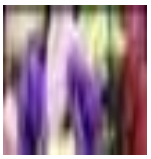
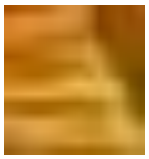
با توجه به اینکه اطراف ناحیه متنی، پس‌زمینه‌ی غیر متنی وجود دارد واریانس بلوک‌های به‌دست آمده مربوط به رنگ قلم و همچنین پس‌زمینه است. لذا با خوشه‌بندی رنگ درون بلوک‌ها ابتدا نواحی درون هر بلوک را به ۳ خوشه تفکیک می‌کنیم. برای هر خوشه واریانس رنگ محاسبه شده، خوشه با کمترین واریانس می‌تواند مربوط به متن باشد. بلوک‌هایی که این واریانس برای آن‌ها کمتر از یک آستانه مشخص است را به عنوان بلوک متنی در نظر گرفتیم.



شکل ۴-۱۳ :مقادیر واریانس برای بلوک‌های نمونه. نمودار قرمز رنگ مقدار واریانس نواحی متنی و نمودار آبی مقدار واریانس نواحی غیر متنی است. محور عمودی مقادیر واریانس برای خوشه با کمترین واریانس برای بلوک‌های متن و غیرمتن را نشان می‌دهد و محور افقی تعداد ۱۰۰ نمونه متن و ۸۵ نمونه غیر متن می‌باشد

همانطور که ملاحظه می‌گردد متاسفانه آستانه پایداری برای فیلتر کردن بلوک‌های غیرمتنی از متن وجود ندارد. در شکل(۴-۱۳) نمودار قرمز رنگ مقدار واریانس نواحی متنی و نمودار آبی مقدار واریانس نواحی غیر متنی است. محور عمودی مقادیر واریانس برای خوشه با کمترین واریانس برای بلوک‌های متن و غیرمتن را نشان می‌دهد و محور افقی تعداد ۱۰۰ نمونه متن و ۸۵ نمونه غیر متن می‌باشد. در جدول (۴-۳) موارد نقض در صحنه‌ها متنی نشان داده شده است.

جدول۴-۳: موارد نقض در بلوک متنی

موارد نقض در تشخیص متن				
موارد نقص در تشخیص غیرمتن				

۴-۳-۳ پیشنهاد سوم: استفاده از POI[1] و فیلترگابور

در پیشنهاد بعدی برای حذف بلوک‌های غیر متنی از فیلتر گابور استفاده می شود[۱۹][۲۷] و ابتدا نقاط با بیشترین مقدار پاسخ گابور مشخص می‌شود. این نقاط، نقاط ویژه در متن می‌باشند و سپس به استخراج ویژگی از این نقاط پرداخته می‌شود[۱۹][۲۷].

در ابتدا لبه‌های صحنه را پیدا کرده و برای هر پیکسل از صحنه، فاصله نزدیک‌ترین لبه با آن پیکسل محاسبه می‌شود و به عنوان پارامتر لاندا برای محاسبه پاسخ گابور مورد استفاده قرار می‌گیرد و زاویه بین پاره خط مذکور تا محور x ها تحت عنوان θ برای محاسبه پاسخ گابور مورد استفاده قرار می‌گیرد. فیلتر گابور طبق رابطه (۴-۱) ساخته می‌شود.

در این رابطه $\sigma = 2$ و $\gamma = 0.2$ است.

$$g(x,y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \varphi\right) \qquad (۴-۱)$$

$$y' = x\sin\theta + y\cos\theta \qquad (۴-۲)$$

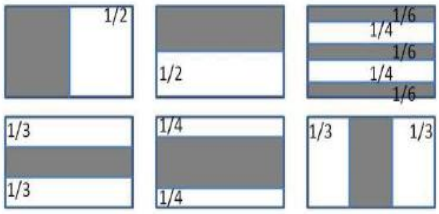
$$x' = x\cos\theta + y\sin\theta \qquad (۴-۳)$$

فیلتر گابور ساخته شده در صحنه کانوال می‌شود و پاسخ گابور در هر پیکسل محاسبه می‌شود. در مرحله بعد فیلتر عمود بر فیلتر قبلی، $(\theta + 90)$، در صحنه کانوال شده و دوباره مقدار کانوال در پیکسل مورد نظر به دست می‌آید. همین روند برای کلیه نقاط صحنه انجام می‌شود و سپس قدر مطلق تفاضل این دو فیلتر محاسبه می‌شود و نقاطی که به صورت محلی دارای ماکزیمم اختلاف باشند نقاط ویژه تصویر نامیده می‌شوند[۱۹].

[1] Pixels of Interest(POI)

در مرجع[۱۹] برای استخراج ویژگی از نقاط ویژه تصویر استفاده شده است. نتایج تجربی بیان‌گر این واقعیت بود که این نقاط نتیجه مطلوبی را نه برای متون فارسی و نه انگلیسی حاصل نکردند و در ادامه ما برای استخراج ویژگی از نقاط مرکزی یک نویسه استفاده می‌کنیم. به این صورت که در هر صحنه نقاط میان لبه‌های نواحی کاندید متن در صحنه به عنوان ویژه تصویر استفاده می‌کنیم. این نقاط در بلوک‌های متنی دارای ساختاری هماهنگ متناظر با جهت‌گیری حرف یا حروف مورد نظر می‌باشد و برای بلوک‌های غیرمتنی به صورت ناهماهنگ و پراکنده در کل بلوک هستند.

برای استخراج ویژگی از هر بلوک ۶ الگو در نظر گرفته می‌شود و همچنین برای ایجاد تعادل در ارزش ویژگی‌ها برای بخش‌های سفید و سیاه یک وزن در نظر گرفته می‌شود[۱۹]، به صورتی که اگر در یک بخش ما سه ناحیه سیاه داریم به هر کدام از آنها وزن ۳/(۰،۵) داده می‌شود و به همین صورت برای بخش‌های دارای دو ناحیه سیاه وزن ۲/(۰،۵) اختصاص داده می‌شود که در شکل (۴-۱۴) قابل مشاهده است.



شکل ۴-۱۴:وزن‌دهی بلوک‌ها[۱۹]

بدین صورت ما ۶ بخش بدون ضریب و ۴ بخش با ضریب را خواهیم داشت که در مجموع تولید ۱۰ بخش را می‌کند.[۱۹]ویژگی‌های مورد نظر که در ادامه به معرفی آنها پرداخته می‌شود برای هر بخش سیاه و سفید در نقاط مطلوب به دست می‌آید و میانگین و جمع ویژگی‌ها برای هر بخش سیاه و سفید برای ما تولید ۲ ویژگی می‌کند و در نتیجه ۲۰=۲*۱۰ ویژگی خواهیم داشت.

در ادامه به معرفی ویژگی‌ها می پردازیم:

۱) گرادیان: در ابتدا با استفاده از یک فیلتر گوسین صحنه مورد نظر مات می‌شود. اندازه و زاویه گرادیان به عنوان دو ویژگی در نقاط مورد علاقه محاسبه شده و جمع و میانگین آن برای هر ناحیه سیاه و سفید در ۱۰ بخش موجود محاسبه می شود. اختلاف بین میانگین اندازه ها و زاویه ها وجمع اندازه ها و زاویه ها در دو قسمت سفید و سیاه در هر بخش ۴۰=۲۰*۲ ویژگی به دست می‌آید.

۲) عرض: عرض قلم به صورت محلی محاسبه می‌شود و واریانس عرض‌ها برای بخش‌های سیاه و سفید محاسبه می‌شود و مانند ویژگی ۱ تفاضل میانگین و جمع واریانس‌ها در بخش سیاه و سفید محاسبه می‌شود و ۲۰ ویژگی را ایجاد می‌کند. همچنین به همین صورت جذر واریانس تقسیم بر میانگین را نیز به عنوان ۲۰ ویژگی دیگر در نظر گرفته می شود.

۳) توزیع عرض: در این بخش بدون در نظر گرفتن بخش‌های سیاه وسفید، وزن‌دهی، عمل می-کنیم و در هر ۱۸ زیر بخش ۶ بلوک موجود با استفاده از روش otsu [۲۸] نسبت مساحت پیش‌زمینه به پس زمینه محاسبه می شود و به مجموع ویژگیها ۱۸ ویژگی اضافه شده است.

تعداد کل ویژگی‌ها به ۹۸ ویژگی می‌رسد.و برای کلاسه‌بندی به ماشین بردار پشتیبان داده می شود و با استفاده از آن بخشهای غیر متنی حذف می شود.

نتایج به دست آمده از این روش بر روی صحنه‌ها اجرا می‌شود و مشاهده می‌شود که در حذف نواحی متنی کار آمد نمی باشد نتایج تجربی به دست آمده بر روی متون انگلیسی به قرار زیر است.

جدول۴-۴: نتایج روش [۱۹] بر روی متون انگلیسی

روش	درصد تشخیص متن	درصد تشخیص غیر متن	درصد کلی
گرادیان	۴۸,۲۳	۵۸,۳۵	۵۵,۲۱
عرض	۶۹,۲۳	۶۵,۲	۶۶,۰۱
توزیع عرض	۷۲,۴۴	۷۰,۳۲	۷۰,۵۴

نتایج تجربی به دست آمده بر روی متون فارسی و انگلیسی به قرار زیر است.

جدول۴-۵: نتایج روش [۱۹] بر روی متون فارسی و انگلیسی

روش	درصد تشخیص متن	درصد تشخیص غیر متن	درصد کلی
گرادیان	۲۰,۲۵	۳۲,۴۸	۲۸,۲۳
عرض	۵۵,۳۲	۵۴,۲۸	۵۴,۸۶
توزیع عرض	۶۰,۳۵	۴۸,۲۱	۵۰,۶۸

با در نظر گرفتن مجموع دو ویژگی ما ۷۰,۲۳ درصد پاسخ صحیح متن و غیر متن را شاهد بودیم و از ویژگی گرادیان صرف نظر کردیم.

۴-۳-۴ روش پیشنهادی چهارم: هیستوگرام گرادیان اندازه و زاویه

با بررسی ساختار حروف انگلیسی و فارسی مشاهده می‌شود که جهت و اندازه گرادیان در زوایای خاص (صفر، ۹۰، ۱۸۰و۲۷۰ درجه) دارای مقادیر بیشتری می‌باشد. در صورتی که در بلوک‌های غیر-متنی اندازه و جهت گرادیان به صورت تصادفی و گوناگون است.به همین خاطر به عنوان ویژگی دیگر، از اندازه گرادیان صحنه در یک بلوک و هیستوگرام آن در ۱۰ بازه استفاده می‌کنیم و به عنوان ۱۰ویژگی به ماشین بردار پشتیبان برای کلاسه بندی متن و غیرمتن می‌دهیم. عدد ۱۰ به‌دست آمده برای تعداد بازه‌های هیستوگرام به صورت تجربی به‌دست آمده است و مشاهده شده است که در ازای تقسیم به ۱۰ بازه نتایج بهتری را شاهد خواهیم بود . ۲۰۰ بلوک شامل ۱۰۰ بلوک متن و ۱۰۰ بلوک غیر متن آموزش داده می‌شود و مشاهده می‌شود و ۸۵ درصد صحنه‌ها درست تشخیص داده می‌شوند.

همچنین زاویه گرادیان نیز محاسبه شده و هیستوگرام آن در ۵ بازه محاسبه شده و به صورت مجزا و همچنین ترکیب با ۵ ویژگی قبلی به طبقه بند ماشین بردار پشتیبان داده می‌شود که در قسمت اول ۷۵ درصد و در ترکیب ویژگی‌ها ۸۷ درصد صحنه‌ها درست تشخیص داده می‌شوند.

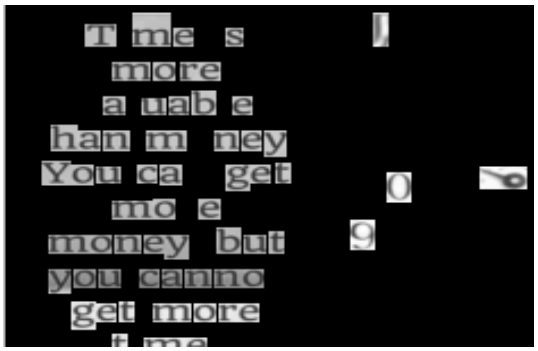
جدول ۴-۶: نتایج روش پیشنهادی

روش	درصد تشخیص متن	درصد تشخیص غیر متن	درصد کلی
اندازه گرادیان	۷۴,۵۴۵۵	۹۵,۴۵۴۵	۸۵,۲۴۵۹
زاویه گرادیان	۷۳,۴۶۹۴	۷۸,۳۳۳۳	۷۵,۴۵۴۵
اندازه+زاویه گرادیان	۷۳,۰۷۶۹	۱۰۰	۸۷,۰۶۹۰

۸۷ درصد از بلوک‌ها به درستی تشخیص داده می‌شوند. اما با بررسی نتایج تجربی مشاهده می‌شود که در میان حروف یک کلمه بعضی از حروف شناسایی نشده‌اند. در ادامه به دنبال روشی برای بازیابی حروف از دست رفته هستیم.

۴-۴ بازیابی حروف از دست رفته

برای بازیابی موارد از دست رفته، امکان اتصال کلیه بلوک‌های متنی موجود در یک سطر وجود ندارد. همانطور که در شکل(۴-۱۵) مشاهده می‌شود، بلوکی غیرمتنی در گوشه‌ای از صحنه وجود دارد که با اتصال بلوک‌های موجود در یک سطر، ایجاد ناحیه ناصحیح می‌کند.



شکل ۴-۱۵: نواحی غیرمتنی در امتداد خطوط متنی

با توجه به شکل(۴-۱۵) در صورت اتصال کلیه بلوک‌های موجود در یک سر خروجی نواحی غیر متنی زیادتری در متن تولید می‌کند.

۴-۴-۱ گروه بندی بلوک‌ها در هر سطر

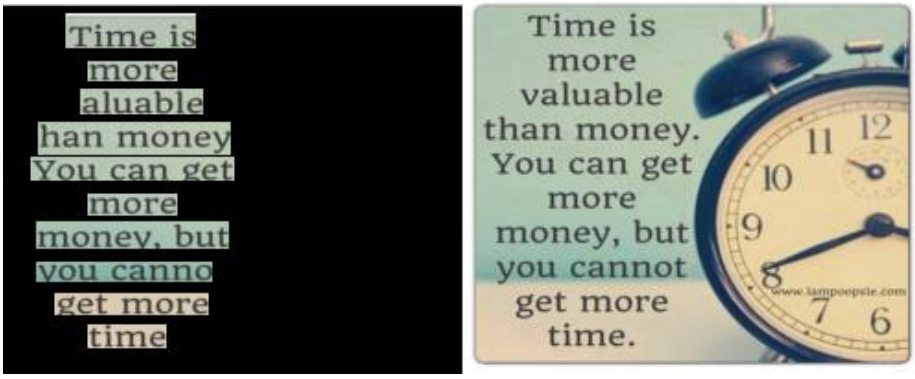
در دید اول تلاش برای ترکیب بلوک‌ها بر مبنای فاصله‌ی آن‌ها از یکدیگر انجام شد. به این صورت که مختصات y و x هر بلوک در سطر به عنوان ویژگی به تخمین‌گر چگالی پنجره پارزن داده شد. با این پیش‌زمینه فکری که برای ما دو قله که شامل قله‌ی بلوک‌های متنی و غیر متنی ایجاد کند و در مرحله بعد بلوک‌های موجود در هر گروه باهم ترکیب شوند. اما در عمل، پارزن هر فاصله کوچک میان بلوک‌ها را به عنوان یک قله در نظر گرفت و این روش، روش کارآمدی نبود.

۴-۴-۲ بررسی فاصله میان بلوکی

در دیدگاه بعدی برای بازیابی از این تاکتیک استفاده می‌شود که، اگر فاصله میان دو بلوک مشخص شده به عنوان بلوک متنی از مقدار مشخصی کمتر باشد این فاصله میانی نیز به عنوان ناحیه متنی در نظر گرفته می‌شود. برای تعیین این میزان آستانه، از عرض بلوک‌های موجود در سطر مورد نظر استفاده می‌شود. به این خاطر که اندازه بلوک‌های در مثال‌های مختلف متفاوت است و تعیین فاصله مشخص که برای همه مثال‌ها پاسخ گو باشد امکان‌پذیر نیست. در روش پیشنهادی این فاصله

دو برابر مقدار میانگین عرض بلوک‌های موجود در هر سطر است که امکان بازیابی دو حرف حذف شده در یک کلمه را امکان پذیر می‌سازد.

نتایج حاصل تا این مرحله در شکل (۴-۱۶) مشاهده می‌شود. که همانطور که مشخص است بخش‌های متنی زیادی به مجموعه اولیه اضافه شده است.

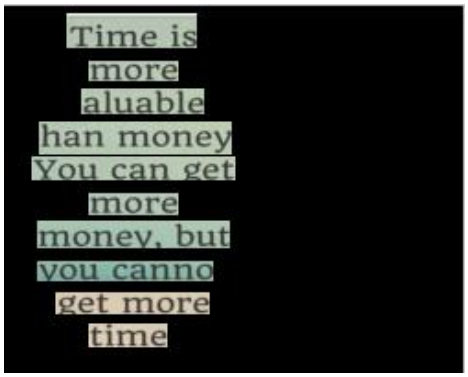


الف ب

شکل ۴-۱۶ :نتیجه این مرحله الف) تصویر ورودی ب) بخش‌های متنی اضافه شده به مجموعه اولیه

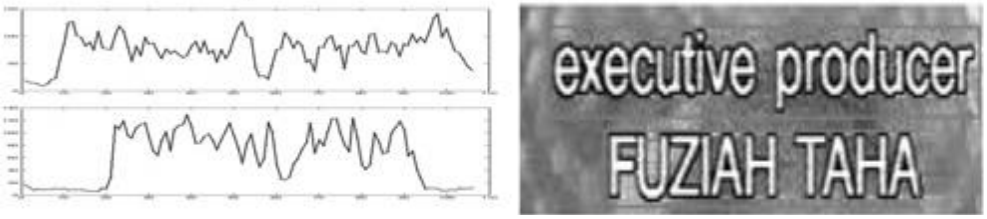
۴-۴-۳ بازیابی نهایی با افکنش

مشاهده نتایج تجربی بیان‌گر این واقعیت است که اگر بلوک‌های ابتدایی و انتهایی در یک سطر در مراحل قبلی شناسایی نشده باشند حرف یا حروف ابتدایی در کلمه اول و یا انتهای در کلمه آخر با روش گفته شده قابل بازیابی نیست. در شکل (۴-۱۷) بلوک متنی مربوط به سه حرف v در کلمه valuable و t در then و tدر حرف آخر not شناسایی نشده است و در مرحله بعدی ، با روش بیان شده، حروف میانی کلمه شناسایی می‌شود اما حروف ابتدا و انتها قابل شناسایی نیست.



شکل ۴-۱۷ : نقصان برخی حروف در متن

برای رفع این مشکل از افکنش افقی بهره می‌بریم [۴] و[۲۶]. که نواحی ابتدایی و انتهایی یک متن را نمایش می‌دهد. شکل(۴-۱۸) ابتدا و انتهای بخش متنی را برای ما مشخص می‌سازد.



شکل ۴-۱۸ : افکنش نواحی متنی[۲۶]

بدین صورت که بعد از مشخص شدن ناحیه متنی با اضافه کردن ۱۰ پیکسل به ابتدا و انتهای ناحیه متنی ناحیه بزرگتری ایجاد می‌شود و افکنش آن در راستای افقی توسط رابطه (۴-۲) زیر محاسبه می‌شود.

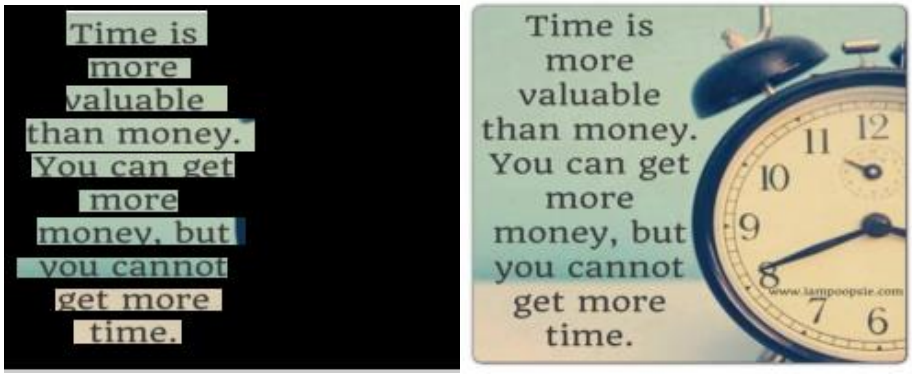
$$V_{pr}(x) = \sum_{y=1}^{w} I(x,y) \qquad (۴-۴)$$

در ادامه برای نرمال کردن آن بر بیشترین مقدار(M) طبق رابطه(۴-۳) تقسیم می‌شود.

$$V_p(x) = {V_{pr}(x)}/{M} \qquad (۵-۴)$$

اگر ناحیه اضافه شده دارای افکنش مشابه با ناحیه متن اصلی بود به ناحیه قبلی اضافه می‌شود. و بدین صورت حرف و یا حرف‌های احتمالی از دست رفته قابل بازیابی می‌باشند.

در شکل(۴-۱۹) نتایج نشان داده شده است.



الف ب

شکل ۴-۱۹ :الف)تصویر ورودی ب) نتیجه نهایی استخراج بلوک‌های متنی در الگوریتم پیشنهادی

فصل پنجم

نتایج شبیه سازی

۵-۱ مجموعه داده و نتایج تجربی

در این بخش به معرفی سه مجموعه داده به‌کار رفته و ارائه نتایج تجربی مربوطه می‌پردازیم.

الگوریتم پیشنهادی در محیط Matlab پیاده‌سازی شده است.

۵-۱-۱ مجموعه داده ICDAR 2003/2005

این مجموعه تصاویر اولین بار در کنفرانس بین المللی تشخیص اسناد در سال ۲۰۰۳ مورد استفاده قرار گرفت [۲۹] و هم اکنون به طور گسترده برای تشخیص متن در صحنه‌ها ی طبیعی مورد استفاده قرار می‌گیرد. در شکل ۵-۱ نمونه‌هایی از این مجموعه تصاویر مشاهده می‌شود.

شکل ۵-۱ : نمونه هایی از تصاویر در پایگاه داده ICDAR 2003/2005

این مجموعه شامل ۲۵۸ صحنه آموزشی و ۲۵۱ صحنه آزمون است و مجموعه‌ای از صحنه‌های اشیا خانگی، نشانه‌های راهنمایی، مغازه، پوستر و جلد کتاب در ابعاد ۹۳*۳۰۷ تا ۹۶۰*۱۲۸۰ می-باشد. این مجموعه از آدرس [۳۰] قابل دانلود است.

نتایج تجربی سیستم تشخیص متن این پایان‌نامه در شکل‌های زیر مشاهده می‌شود. شکل (۵-۲) صحنه‌هایی دارای نویسه‌ها با عرض‌های متفاوت می‌باشد که سیستم توانایی تشخیص کلمات با عرض-های متفاوت را داشته است. در شکل(۵-۲) تصویر الف تصویر ورودی و ب متن تشخیص داده شده است.

الف ب

شکل ۵-۲:نتیجه اعمال روش بر روی تصویر نمونه با اندازه قلم مختلف در یک صحنه الف)تصویر ورودی و ب)متن تشخیص داده شده

شکل‌های (۵-۳) صحنه‌هایی با پس‌زمینه پیچیده و رنگ‌های مختلف می‌باشد.

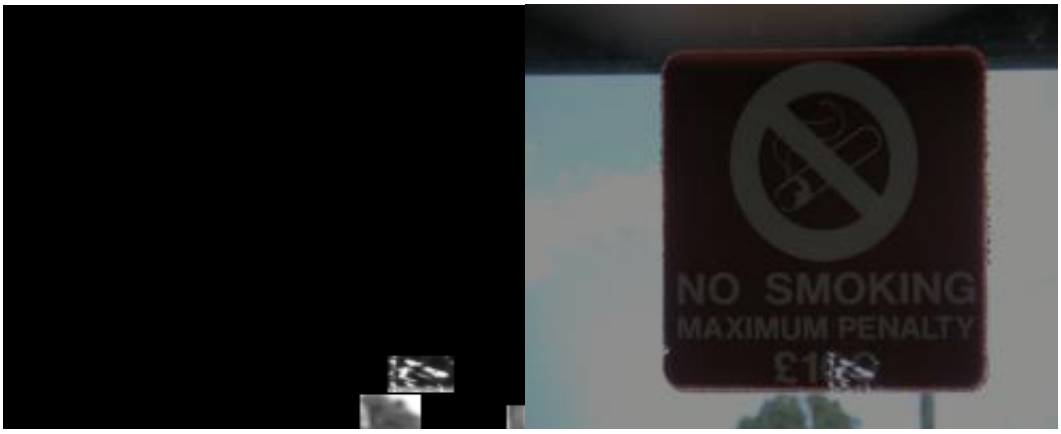
الف ب

شکل ۵-۳ : صحنه با پس‌زمینه پیچیده الف)تصویر ورودی و ب) متن تشخیص داده شده

در شکل‌های (۵-۴)، (۵-۵) مواردی از این مجموعه داده آورده شده است که از سیستم پاسخ مناسبی دریافت نشده است. از جمله عوامل در عدم تشخیص متن می‌توان به اندازه کوچک قلم، اختلاف کم رنگ پس‌زمینه و متن و متن‌های مورب اشاره کرد.

شکل ۵-۴:مثالی از عدم موفقیت روش پیشنهادی به خاطرمورب بودن متن الف)تصویر ورودی وب)متن تشخیص داده شده

در شکل(۵-۵) صحنه دارای اختلاف کم رنگ با پس‌زمینه خود می‌باشد. سیستم توانایی تشخیص تا اختلاف رنگ مشخصی را داشته است. با نزدیک شدن سطح رنگ و به طور هم‌زمان کوچک شدن اندازه در سطر دوم متن امکان تشخیص دقیق لبه‌های متنی برقرار نشده است و متن از دست رفته است.



الف ب

شکل ۵-۵: مثال‌هایی از عدم موفقیت روش پیشنهادی به خاطر اختلاف کم رنگ یا روشنایی قلم و پس‌زمینه

الف)تصویر ورودی و ب)متن تشخیص داده شده

۵-۱-۲ مجموعه داده Microsoft Street View Text Detection Dataset

این مجموعه داده توسط اشتین و همکارانش در سال ۲۰۱۰ منتشر شد.[۳۱] این مجموعه شامل ۳۰۷ صحنه از نمای خیابان در قطع ۷۶۸*۱۰۲۴ تا ۱۰۲۴*۱۳۶۰ می‌باشد. از آن‌جا که اکثر صحنه‌ها حاوی گیاهان،الگوهای تکراری، متن در اندازه‌های کوچک است نسبت به مجموعه ICDAR ۲۰۰۵/۲۰۰۳ بسیار پیچیده‌تر و تشخیص مشکل‌تر است. این مجموعه از آدرس [۳۲] قابل دانلود است.

شکل‌های زیر برخی از نتایج تشخیص متن در این مجموعه را نشان می‌دهند. این مجموعه در مقایسه با مجموعه ICDAR 2003/2005 دارای صحنه‌های پیچیده‌تر شهری می‌باشد.

شکل(۵-۶) ساختاری متفاوت در نوشتار دارد. قسمتی از متن در پس‌زمینه‌ای روشن‌تر از متن است و قسمت دیگر متن تیره‌تر از پس‌زمینه است. قسمتی از متن از دست رفته است.

ب الف

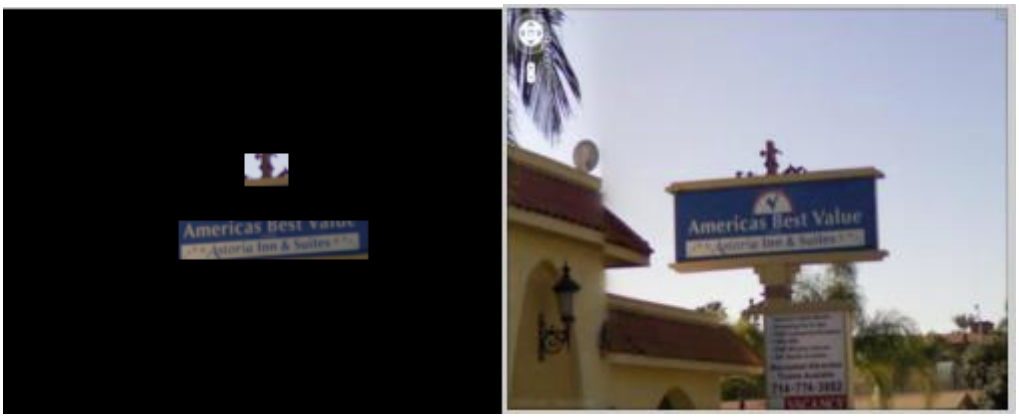
شکل ۵-۶: تفاوت در ساختار متنی الف)تصویر ورودی و ب)متن تشخیص داده شده

شکل(۵-۷) صحنه شهری پیچیده‌ای را نمایش می‌دهد. با وجود تضاد کم رنگی در متن و پس-زمینه، متن به درستی تشخیص داده شده است.

الف ب

شکل ۵-۷ : صحنه با پس‌زمینه شهری پیچیده الف)تصویر ورودی و ب)متن تشخیص داده شده

در شکل(۵-۸) با توجه به پس زمینه پیچیده و متون با قلم کوچک و حروف به هم‌چسبیده بسیاری از اجزای متنی از دست رفته و اما در صحنه نواحی غیر متنی به اشتباه انتخاب نشده است.



الف ب

شکل ۵-۸: حذف نواحی متنی با قلم بسیار ریز الف)تصویر ورودی و ب)متن تشخیص داده شده

شکل‌های(۵-۹) دارای متون به هم چسبیده می‌باشند. هاله‌ی سفید رنگی که متناسب با رنگ متن است بر روی قسمت متنی تشخیص داده نشده وجود دارد. دو کلمه welcome و ford تشخیص داده نشده‌اند. چون لبه‌های صحنه قابل شناسایی نبوده و بلوک متنی ایجاد نشده است.

الف ب

شکل ۵-۹: صحنه با حروف به هم چسبیده الف)تصویر ورودی و ب)متن تشخیص داده شده

شکل(۵-۱۰) ساختار متفاوتی در نوشتار دارد. قسمتی از متن در پس‌زمینه‌ای روشن‌تر نسبت به متن قرار دارد و در قسمت دیگر متن تیره‌تر از پس‌زمینه است. در نتیجه قسمتی از متن از دست می-رود.

الف ب

شکل ۵-۱۰: ساختار متفاوت رنگی در نگارش الف)تصویر ورودی و ب)متن تشخیص داده شده

۵-۱-۳ پایگاه داده به زبان فارسی

مجموعه بعدی که نتایج الگوریتم بر روی آن آزمایش می‌شود شامل ۲۰۰ عکس از مناظر طبیعی، سطح کتاب، صحنه‌ها با پس‌زمینه پیچیده و تنوع رنگ و قلم‌های مختلف در ابعاد ۲۰۰*۸۵۳ تا ۵۲۶* ۱۲۵۰ با متون فارسی می‌باشد که خودمان تهیه کرده ایم.

شکل(۵-۱۱) ترکیبی از دو نوشتار فارسی و انگلیسی در یک صحنه با پس‌زمینه پیچیده است که سیستم به طور هم‌زمان توانایی استخراج متون فارسی و انگلیسی را داشته است.

الف ب

شکل ۵-۱۱: ترکیب دو نوشتار الف)تصویر ورودی و ب)متن تشخیص داده شده

در شکل(۵-۱۲) کلیه آدمک‌ها به عنوان بلوک‌های متنی در نظر گرفته شده بودند و مانند متون دارای ساختاری با عرض و مراکز یکسان به حساب می‌آمدند و در مرحله استخراج ویژگی امکان حذف آن‌ها بر قرار شد.

الف ب

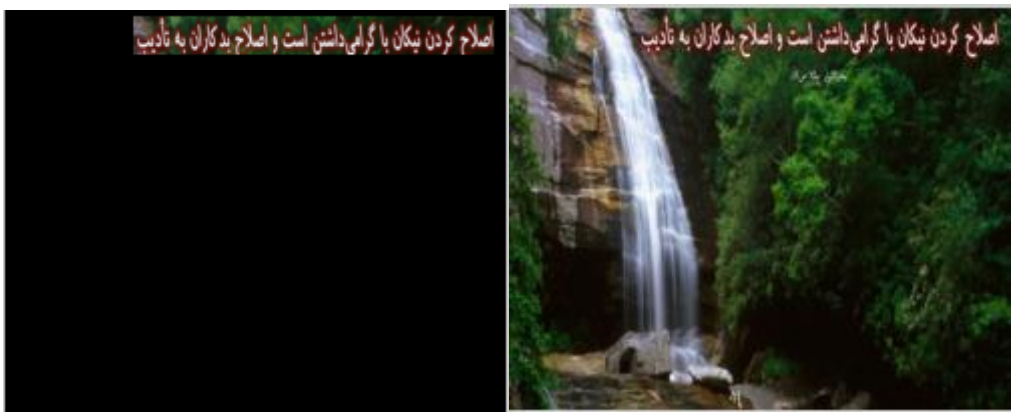
شکل ۵-۱۲: ساختار غیر متنی مشابه با متن در پس‌زمینه الف)تصویر ورودی و ب)متن تشخیص داده شده

شکل (۵-۱۳) دارای قلم ریز است و مشاهده می‌شود در مواردی که رنگ پس‌زمینه و پیش‌زمینه دارای اختلاف ناچیز باشد امکان تشخیص متن وجود ندارد. که در این متن " بهای جان توست" این ویژگی را داشته و به همراه آن چون فاصله بین "ادب " و "پس" بیشتر از حد معمول کلمات به حساب می‌آید بخش "عمل و ادب" نیز به اشتباه جز ناحیه غیر متنی به حساب آمده و از صحنه نهایی حذف می‌شود.

الف ب

شکل ۵-۱۳: قلم ریز و اختلاف کم رنگ پس‌زمینه و متن الف)تصویر ورودی و ب)متن تشخیص داده شده

شکل (۵-۱۴) صحنه‌هایی با پس‌زمینه پیچیده در رنگ‌های مختلف می‌باشند و همچنین حروف در متن دارای حاشیه‌ای با رنگ متفاوت نسبت به رنگ نوشتار داخلی خود هستند. در روش‌های مبتنی بر رنگ این اختلاف نوشتاری در متون و همچنین تنوع رنگ پس‌زمینه امکان تشخیص کامل متن را برای ما از بین خواهد برد. [۸] در اینجا همان‌طور که مشاهده می‌شود متن بدون هیچ کاستی به درستی تشخیص داده شده است.



شکل ۵-۱۴ : پس‌زمینه پیچیده با قلم ریز الف)تصویر ورودی و ب)متن تشخیص داده شده

در شکل(۵-۱۵) صحنه دارای روشنایی یکنواخت نمی‌باشد. در این صورت پیدا کردن لبه‌ها و در ادامه بلوک متنی با مشکل مواجه می‌شود ولی به علت اختلاف رنگی بالا میان متن و پس‌زمینه، بخش‌های متنی به درستی شناسایی شده است.

الف ب

شکل ۵-۱۵: صحنه با نور غیر یکنواخت الف)تصویر ورودی و ب)متن تشخیص داده شده

شکل(۵-۱۶) صحنه‌هایی با پس‌زمینه پیچیده را نشان می‌دهد که دارای اندازه قلم بیشتری نسبت به موارد قبلی است و در این صحنه‌ها نیز تشخیص به درستی صورت گرفته است. با توجه به شکل (۵-۱۶)، اگر سطری از متن دارای تعداد بلوک متنی خیلی کمتر، نسبت به سطرهای دیگر باشد تشخیص آن بلوک‌ها با مشکل مواجه می‌شود. در سطر آخر این صحنه ۲ بلوک متنی شناسایی شد که

در مقایسه با تعداد بلوک‌های دو سطر قبلی، این دو بلوک در مرحله فیلترینگ حذف شد و بخشی از متن از دست رفته است.

الف ب

شکل ۵-۱۶:صحنه در پس‌زمینه پیچیده الف)تصویر ورودی و ب)متن تشخیص داده شده

۵-۲ معیارهای ارزیابی

هدف از ارزیابی عملکرد یک سیستم استخراج متن، اندازه‌گیری تفاوت بین خروجی متن که انتظار می‌رود و خروجی متن واقعی سیستم است. روش‌های خوب ارزیابی عملکرد می‌توانند اطلاعات با ارزشی نه تنها برای ارزیابی سیستم و مقایسه، بلکه برای انتخاب سیستم مناسب و بهبود آن ارائه کنند.

Precision و Recall دو معیار قوی برای ارزیابی سیستم‌های تشخیص متن می‌باشند. لوکاس و همکارانش[۲۷] در robust reading competition of ICDAR 2003 روش ارزنده‌ای را برای ارزیابی سیستم مطرح کردند که در آن Recall نسبت تعداد بلوک‌های متنی درست تشخیص داده شده به تعداد کل بلوک‌های متنی در صحنه است و Precision نسبت تعداد بلوک‌های متنی درست تشخیص داده شده به مجموع تعداد بلوک‌های متنی درست تشخیص داده شده و موارد غیر متنی که به اشتباه متنی تشخیص داده شده می‌باشد.[۲۷]

$$\mathrm{Re}call = \frac{N_r}{N_g} \qquad (۱-۵)$$

$$\mathrm{Pr}ecision = \frac{N_r}{N_d} \qquad (۲-۵)$$

که در روابط بالا N_r تعداد بلوک‌های متنی درست تشخیص داده شده و N_g تعداد کل بلوک‌های متنی در صحنه و N_d تعداد کل بلوک‌های تشخیص داده شده می‌باشند.

عملکرد کلی توسط پارامتر f که می‌تواند ترکیبی از دو معیار بالا باشد ارائه می‌شود[۲۸]:

$$f = \frac{2*p*r}{p+r} \qquad (۳-۵)$$

در جدول(۵-۱) الگوریتم پیشنهادی با استفاده از ۳ معیار بیان شده(Precision ، Recall، f) با کارهای گذشته مقایسه می‌شود.کارهای انجام گرفته بر روی مجموعه داده ICDAR 2003/2005 و

مجموعه داده Microsoft Street View Text Detection Dataset می‌باشند. مقادیر ماکزیمم در کارهای پیشین برای هر یک از معیارها در جدول مشخص شده است.

جدول۵-۱: ارزیابی سیستم با روش‌های پیشین بر روی مجموعه داده ICDAR 2003/2005 و مجموعه داده Microsoft Street View Text Detection Dataset

Method	P	R	F
Epshteinet al[۳۱]	.73	.60	.66
Becker[۲۹]	.63	.65	.64
Lukas et al[۲۹]	.59	.55	.57
Ashida[۳۱]	.55	.46	
Liu [۳۴]	.63	.65	.64
Yi et all[۳۵]	.71	.62	.62
Chen et all[۳۷]	.60	.60	.58
Zhu et all[۳۷]	.33	.40	.33
Kim et all[۳۷]	.22	.28	.22
Ezaki et all[۳۷]	.18	.36	.22
Minetto [۳۸]	.63	.61	.61
Fabrizio [۳۹]	.46	.39	.43
Zhanh [۳۱]	.67	.46	.55
Our method	.81	.71	.73

در ادامه برای ارزیابی دقیق‌تر از معیار تشخیص در سطح کلمه و نویسه نیز بهره بردیم. برای این ارزیابی از سه مجموعه معرفی شده (مجموعه داده ICDAR 2003/2005 و مجموعه داده Microsoft Street View Text Detection Dataset و پایگاه داده به زبان فارسی) استفاده کردیم. در مجموع ۲۰۰ صحنه از مجموعه سه دیتا بیس انتخاب شده است. این صحنه‌ها دارای ۱۴۷۹ کلمه (فارسی و انگلیسی) و تعداد ۴۷۲۸ نویسه و زیر کلمه می‌باشند. جداول زیر تعداد و درصد شناسایی را در دو سطح کلمه و زیر کلمه نمایش می‌دهند. نتایج حاصل برای دو بخش از کار می‌باشند .جدول ۵-۲

نتایج بعد از مرحله استخراج ویژگی و خروجی svm است. و جدول ۵- ۳ نتایج خروجی نهایی سیستم تشخیص متن را نمایش می دهد.

جدول ۵-۲: نتایج بعد از مرحله استخراج ویژگی و خروجی ***svm***

سطح	تعداد تشخیص درست	تعداد کل	درصد
بلوک	۱۷۵۳	۲۱۵۶	۸۱,۳
کلمه	۱۲۱۲	۱۴۷۹	۸۱,۹۴
زیر کلمه	۳۷۹۷	۴۷۲۸	۸۰,۳

جدول ۵-۳: نتایج خروجی نهایی سیستم پیشنهادی

سطح	تعداد تشخیص درست	تعداد کل	درصد
بلوک	۱۷۵۳	۲۱۵۶	۸۱,۳
کلمه	۱۲۷۲	۱۴۷۹	۸۶,۰۱
زیر کلمه	۴۰۱۰	۴۷۲۸	۸۴,۸۱

فصل ششم

نتایج و کارهای آینده

۶-۱ نتیجه گیری

در این پایان‌نامه درفصل اول مقدمه‌ای برای آشنایی با موضوع پایان‌نامه، دلایل انجام آن، کاربردها و چالش‌های پیش رو بیان شد. فصل دوم به بیان پیشینه‌ای از کارهای انجام شده،دسته‌بندی کارهای گذشته در ۶ گروه و توجه به ویژگی‌ها و نقاط ضعف و قوت هر گروه پرداخته شد و در فصل سوم توضیح مختصری در مورد مفاهیم به کار رفته در پژوهش گرد آوری شده است.در فصل چهارم روش پیشنهادی به تفصیل آورده شد. سیستم تشخیص متن بر مبنای ارتباط میان اجزای متنی است که با در نظر گرفتن برخی ویژگی‌های متنی به حذف اجزای غیر متنی می‌پردازد. در ابتدا به استخراج نواحی کاندید متن با استفاده از سه روش زیر پرداخته شد:

۱) تفکیک نواحی صحنه با توجه به رنگ موجود در آن‌ها

دراین زمینه به گروه‌بندی پیکسل‌های لبه با توجه به رنگ همسایه‌های مجاور پرداخته شد. این روش با استفاده از الگوریتم EM به استخراج پارامتر های مخلوط گوسی(GMM)در فضای رنگ لبه-های صحنه پرداخته است[۱۹]. نتایج تجربی نشان داد که این روش دراستخراج لبه‌های متنی و در ادامه استخراج ناحیه متن مناسب نیست. به این خاطر که نواحی متن از هم جدا شده و در ۵ لایه تقسیم می‌شدند و همچنین در هر بار انجام الگوریتم به خاطر تصادفی بودن فرایند خوشه‌بندی به کمک GMM، نتایج متفاوتی در ۵ لایه برای عکسی یکسان مشاهده می‌شد. به همین خاطر از این روش صرف نظر شد

۲) استخراج نواحی متنی از تبدیل موجک سطح یک

بعد از استخراج لبه‌ها با تبدیل موجک، لبه‌های هر زیر باند با استفاده از یک آستانه فیلتر شد. این آستانه با توجه به ماکزیمم مقدار هیستوگرام و اندازه هر زیر باند تعیین می‌شد. لبه‌های منفرد در هر زیر باند حذف شده و اگر یک پیکسل در هر سه زیر باند پیکسل متنی بود، به عنوان پیکسل متنی

انتخاب می شد. در ادامه با چند عملیات ریخت شناسی بازکردن و بسته کردن در راستای افقی و عمودی ناحیه متنی به دست آمد. نتایج تجربی نشان دهنده آن بود که این روش در صحنه‌ها با پس-زمینه ساده یا پیچیدگی کم پاسخ مناسبی می‌دهد ولی در صحنه‌ها با پس زمینه پیچیده با لبه‌های بسیار پاسخ مناسبی را شاهد نخواهیم بود.

۳) استفاده از جهت گرادیان روشنایی

در ادامه برای یافتن روشی کارآمد با استفاده از لبه یابCanny لبه‌های صحنه مشخص شد.لبه-های صحنه استخراج شده در هر سطر مورد پردازش قرار گرفت. لبه‌هایی که فاصله آن‌ها از یک آستانه کمتر بود به عنوان دو لبه متنی شناخته شده و ناحیه ما بین آن‌ها به عنوان ناحیه متنی مارک شد. از جمله معایب این روش در مواردی که حروف نزدیک به هم باشند ناحیه بین حروف نیز به عنوان متن انتخاب می‌شودو همچنین در مورد متون فارسی لبه‌های دو طرف حروفی مانند "ب" ، "پ"،"ت" و امثال آن در قسمت انتهایی زیاد است و این نواحی تحت عنوان متن شناسایی نمی‌شوند.برای بهبود سیستم شناسایی، روش زیر ارایه شد.

در ابتدا گرادیان تصویر استخراج شد. با توجه به این واقعیت که در لبه‌های متنی صحنه تغییر در روند مقادیر گرادیان(صعودی و نزولی) را شاهد هستیم و با قرار دادن یک مقدار آستانه لبه‌های صحنه مشخص شد. این فرایندر هر دو راستایxو y بر روی صحنه اعمال شد. و با هر بار تغییر در جهت گرادیان و برقراری شرط فاصله، مقادیر ۱ و۱- به لبه مورد نظراختصاص داده شد. در هر سطر صحنه این توالی بررسی شد و مرکز بین جفت پیکسل‌های مذکور مشخص شد. در مرحله بعدی اجزای متصل در صحنه به دست آمده را استخراج کرده و پنجره محیط بر هرناحیه متصل استخراج شد که بلوک بلوک‌های متنی و غیر متنی را برای ما تولید کردند. در گام بعد اقدام به گروه بندی بلوک‌های موجود در صحنه به کمک دو ویژگی کردیم. ویژگی اول مختصات سطری مرکز بلوک‌ها و ویژگی دوم ارتفاع پنجره‌های کاندید متن است که مورد توجه قرار می‌گیرد. به این خاطر که نواحی متنی دارای بلوک-

هایی با ارتفاع‌های نزدیک به هم هستند.از هر بلوک کاندید این دو ویژگی استخراج شده و چگالی بلوک‌ها با استفاده از پنجره پارزن در فضای دو ویژگی استخراج شده، تخمین زده شد. با آستانه گذاری بر روی خوشه‌ها، خوشه‌هایی که قله آن‌ها بزرگتر از آستانه تعیین شده باشد به عنوان خوشه متنی کاندید گردید.قله‌ای که دارای پیک بزرگتری است نشان دهنده این واقعیت است که فراوانی بلوک‌ها با سطری مشابه و ارتفاع‌های نزدیک به هم در ناحیه‌ای از فضای ویژگی زیاد است. این سطرها به عنوان سطرهایی که مراکز بلوک‌های متنی در آن‌ها قرار دارد در نظر گرفته شدند.

در مرحله بعدی استخراج ویژگی از بلوک‌های موجود برای حذف بلوک‌های غیر متنی در ۴ دیدگاه زیر انجام شد:

۱) عرض قلم در بلوک‌ها

با این استراتژی پیش رفتیم که عرض قلم در بلوک متنی یکسان است و برای تشخیص از معیار-های پراکندگی استفاده شد. معیارهای پراکندگی واریانس وآنتروپی بلوک‌ها بود. با کلاسه بندی دو ویژگی فوق توسط svm بلوک‌های متنی و غیر متنی از هم جدا شدند.مشاهده نتایج نشان‌دهنده آن بود که در متون انگلیسی تشخیص به درستی صورت می‌گیرد. اما در متون فارسی در مواقعی که حروف به صورت متصل به هم هستند و همچنین در حروف دندانه‌دار دچار مشکل می‌شد. به این خاطر که عرض قلم در برخی نواحی متنی دارای تغییرات زیادی است، تشخیص به درستی امکان پذیر نبود. همچنین بلوک‌های غیر متنی که ساختاری بسیار مشابه به حروف انگلیسی و فارسی دارندتشخیص را برای ما مشکل می ساختند. در این روش ما ۷۳ درصد تشخیص درست متن و ۶۳ درصد تشخیص درست غیر متن را شاهد بودیم.

۲) استفاده از واریانس

با توجه به اینکه اطراف ناحیه متنی، پس‌زمینه‌ی غیر متنی وجود داردواریانس بلوک‌های به‌دست آمده مربوط به رنگ قلم و همچنین پس‌زمینه است. لذا با خوشه‌بندی رنگ درون بلوک‌ها، ابتدا نواحی درون هر بلوک را به ۳ خوشه تفکیک کردیم. برای هر خوشه واریانس رنگ محاسبه شده، خوشه با کمترین واریانس می‌توانست مربوط به متن باشد. بلوک‌هایی که این واریانس برای آن‌ها کمتر از یک آستانه مشخص بود را به عنوان بلوک متنی در نظر گرفتیم. متاسفانه آستانه پایداری برای فیلتر کردن بلوک‌های غیرمتنی از متن وجود نداشت.

۳) استفاده از نقاط مرکزی

در مرجع[۱۹] برای استخراج ویژگی از نقاط ویژه تصویر استفاده شده بود که توسط فیلتر گابور به دست می‌آمد. نتایج تجربی بیان‌گر این واقعیت بود که این نقاط نتیجه مطلوبی را نه برای متون فارسی و نه انگلیسی حاصل نکردند و در ادامه ما برای استخراج ویژگی از نقاط مرکزی یک نویسه استفاده کردیم. به این صورت که در هر صحنه از نقاط میان لبه‌های نواحی کاندید متن در صحنه، به عنوان نقاط ویژه تصویر استفاده کردیم. این نقاط در بلوک‌های متنی دارای ساختاری هماهنگ متناظر با جهت‌گیری و انحنای حرف یا حروف مورد نظر بودند و برای بلوک‌های غیرمتنی به صورت ناهماهنگ و پراکنده در کل بلوک دیده می‌شدند. در این روش از ویژگی عرض و توزیع عرض استفاده کردیم و با در نظر گرفتن مجموع دو ویژگی ما ۷۰,۲۳ درصد پاسخ صحیح متن و غیر متن را شاهد بودیم.

۴) استفاده از جهت و اندازه گرادیان

با بررسی ساختار حروف انگلیسی و فارسی مشاهده شد که جهت و اندازه گرادیان در زوایای خاص (صفر، ۹۰، ۱۸۰و۲۷۰ درجه) دارای مقادیر بیشتری می‌باشد. در صورتی که در بلوک‌های غیرمتنی اندازه و جهت گرادیان به صورت تصادفی و گوناگون است.به همین خاطر به عنوان ویژگی دیگر، از

اندازه گرادیان صحنه در یک بلوک و هیستوگرام آن در ۱۰ بازه استفاده کردیم و به عنوان ۱۰ویژگی به ماشین بردار پشتیبان برای کلاسه بندی متن و غیرمتن داده می‌شود.

همچنین زاویه گرادیان نیز محاسبه شده و هیستوگرام آن در ۵ بازه محاسبه شده و به صورت مجزا و همچنین ترکیب با ۱۰ ویژگی قبلی به طبقه بند ماشین بردار پشتیبان داده شد که در قسمت اول ۷۵ درصد و در ترکیب ویژگی‌ها ۸۷ درصد صحنه‌ها درست تشخیص داده می‌شوند. ۸۷ درصد از بلوک‌ها به درستی تشخیص داده می‌شوند، اما با بررسی نتایج تجربی مشاهده شد که در میان حروف یک کلمه، بعضی از حروف شناسایی نشده‌اند. در ادامه برای بازیابی حروف از دست رفته، ۲ روش زیر انجام شد:

۱) گروه بندی بلوک‌ها در هر سطر

تلاش برای ترکیب بلوک‌ها بر مبنای فاصله‌ی آن‌ها از یکدیگر انجام شد. به این صورت که مختصات y و x هر بلوک در سطر به عنوان ویژگی به تخمین‌گر چگالی پنجره پارزن داده شد. با این پیش‌زمینه فکری که برای ما دو قله که شامل قله‌ی بلوک‌های متنی و غیر متنی است ایجاد کند و در مرحله بعد بلوک‌های موجود در هر گروه باهم ترکیب شوند. اما در عمل، پارزن هر فاصله کوچک میان بلوک‌ها را به عنوان یک قله در نظر گرفت و این روش، روش کارآمدی نبود.

۲) بررسی فاصله میان بلوکی

در دیدگاه بعدی برای بازیابی از این تاکتیک استفاده شد که، اگر فاصله میان دو بلوک مشخص شده به عنوان بلوک متنی از مقدار مشخصی کمتر باشد این فاصله میانی نیز به عنوان ناحیه متنی در نظر گرفته شود. برای تعیین این میزان آستانه، از عرض بلوک‌های موجود در سطر مورد نظر استفاده شد. به این خاطر که اندازه بلوک‌های در مثال‌های مختلف متفاوت است و تعیین فاصله مشخص که برای همه مثال‌ها پاسخ گو باشد امکان‌پذیر نبود. در روش پیشنهادی این فاصله دو برابر مقدار میانگین

نتیجه آزمایش برروی صحنه‌هایی شامل متون فارسی ومتون لاتین بررسی شد عرض بلوک‌های موجود در هر سطر است که امکان بازیابی دو حرف حذف شده در یک کلمه را امکان پذیر می‌سازد.

در ادامه با مشاهده نتایج تجربی به این نتیجه رسیدیم که اگر بلوک‌های ابتدایی و انتهایی در یک سطر در مراحل قبلی شناسایی نشده باشند حرف یا حروف ابتدایی در کلمه اول و یا انتهای در کلمه آخر با روش گفته شده قابل بازیابی نیست.برای رفع این مشکل از افکنش افقی بهره می‌بریم.در فصل پنجم روش پیشنهادی را بر روی سه مجموعه داده مورد بررسی قرار دادیم. ICDAR 2003/2005 locating dataset text شامل ۲۵۸ تصاویر آموزشی و ۲۵۱تصاویرآزمون است و Street view text dataset detection شامل ۳۰۷ تصاویرنمای خیابان می‌باشد و پایگاه داده به زبان فارسی شامل ۲۰۰ عکس از مناظر طبیعی، سطح کتاب، صحنه‌ها با پس‌زمینه پیچیده و تنوع رنگ و قلم‌های مختلف با متون فارسی می‌باشد.

نتیجه آزمایش برروی صحنه‌هایی شامل متون فارسی ومتون لاتین بررسی شد. ارزیابی سیستم در سه سطح بلوک و کلمه و زیرکلمه انجام گرفت. در سطح زیرکلمه ۸۴ درصد و در سطح کلمه ۸۶ درصد تشخیص درست را شاهد بودیم. در سطح بلوک سه معیار ارزیابی recall، prsicion وf به ترتیب ۰,۷۳ ، ۰,۸۱ و ۰,۸۶ بود.

براساس تشخیص وارزیابی نتایج حاصل ازسه مجموعه داده‌بیان شده، روش تشخیص متن پیشنهادشده می‌تواندرصحنه‌هایی با قلم‌ها،اندازه‌ها، رنگ مختلف متن کارآمد باشد.

۶-۲ پیشنهادات

۱) ما متون رابه صورت افقی درنظرگرفتیم که برای اغلب صحنه‌ها با متون انگلیسی وفارسی پاسخگو خواهد بود ولی درمواردی که متن به صورت عمودی باشد روش دچار مشکل خواهد شد. برای رفع این مشکل می‌توان مختصات y بلوک‌ها نیز در تخمین‌گر پارزن مورد توجه قرار گیرند.

۲) آستانه‌های استفاده شده درروش پیشنهادی به صورت تجربی به دست آمده اند. برای تخمین این پارامتر ها می توان ازالگوریتم‌های یادگیری نظیر شبکه‌های عصبی برای رسیدن به پاسخ مطلوب‌تر استفاده کرد.

۳) اعمال الگوریتم پیشنهادی بر روی نوشتار به زبان‌های دیگر.

۴) تلاش برای از بین بردن گسستگی‌های احتمالی میان لبه‌های متنی به این خاطر که هر چه لبه‌های متنی بهتر شناسایی شود در ادامه الگوریتم کارایی بهتری خواهد داشت.

مراجع

[1] Botuman.k, Delp.E "A Low Complexity Method For Detection of Text Area in Natural Images," ICASSP, 2010.

[2] Khodadadi M., Behrad A, "Text Localization, Extraction and Inpainting in Color Images", 20th Iranian Conferance on Electrical Engineering, (ICEE2012), May 15-17,2012

[3] Woo.H,. Sook.Y "Unsupervised Event Extraction from Biomedical Text Based on Event and Pattern Information" CICling 2004,LNCS 2645,PP.533-536,2004

[4] Gllavata J., R. Ewerth. And B. Freisleben, "Text Detection in Images Base on Unsupervised Classification of High Frequency Wavelet Coefficients", Proceedings of International Conterence on Pattern Recognition, 2004.

[5] Wang.L, Wei.L, and Shivakumar," Selectivity Estimation for Extraction Operators over Text Data"Technical Report No. UCB/EECS-2010-107 July 2, 2012

[6] Park J. "Automatic Detection and Recognition of Shop Name in Outdoor Signboard Images ", IEEE International Symposium on ISSPIT 2008.

[۷] داراب م ، ۱۳۹۱، پایان‌نامه دوره کارشناسی ارشد،تشخیص متن فارسی از مناظر طبیعی ، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیر کبیر

[8] Ma l., Wang Ch, Xiao B. "Text Detection in Natural Image Based on Multi Scale Edge Detection and Classification", 3rd International Congress on Image and Signal Processing(CISP2010), 2010

[9] Lee J.and et, "Adaboost for Text Detection in Natural Scene ", International Conference on Document Analysis and Recognition,2011.

[10] Park J., G. Lee, E. Kim , J.Lim. S.Kim, and H. Yang, "Automatic Detection and Recognition of Korean Text in Outdoor Signboard Images", Pattern Recognition Letters, Vol.31,2010.

[11] Fu L., W Wang, Y. Zhan. "A Robuts Text Segmentation Approaach in Complex Background Based on Multiple Constraints", Proceedings of the 8th Pacific Rim conference on Advances in multimedia information processing, 2005.

[12] Mancas-Thilou C., B.Gosselin, "Spatial and Color Spaces ombination for Natural Scene Text Extraction",IEEE International Conference on Image Processing, pp.985-988, 2006.

[13] Mancas-Thilou C., B.Gosselin, "Color Text Extraction with Selective Metric-based Clustering", Computer Vision AND Image Understanding, pp.97-107, 2007.

[14] Shivakumara P., T.Q.Phan and C.L.Tan, "New Fourier-Statistical Features in RGB Space for Video Text Detection", IEEE Transactions on Circuits and Systems for Video Technology, 2010.

[۱۵] نظام آبادی ح، عباسی م، سریزدی س ،"استخراج متن فارسی از تصاویر ثابت در حوزه موجک"، پنجمین کنفرانس بینایی ماشین و پردازش تصویر، مرداد ۱۳۸۶

[16] Yi C. and Y.L.Tian, "Text String Detection from Natural Scenes by Structure-based Partition and Grouping", IEEE Transactions on Image Processig, 2011

[17] Dinh, T.N., Park, J., and Lee, G. "Text Localization Using Imaage Cues and Text Line Information", Proceedings of International Conference on Image Pricessing, pp.2261-2264, 2010.

[۱۸] سوزنچی کاشانی ز ، یعقوبی م ،" تشخیص متن در تصاویر بر مبنای دسته‌بندی بدون سرپرستی ویژگیهای حاصل از لبه‌های تصویر" ، دومین همایش ملی کامپیوتر، برق و فناوری اطلاعات، همدان، اول اسفند ۱۳۸۷.

[19] Yi, C, Y.Tian " Localizing Text in Scene Images by Boundary Clustering, Stroke Segmentation, And String Fragment Classification" IEEE Transaction on Image Processig, VOl. 21,NO September 2012.

[۲۰] پرهیزکاری م،۱۳۹۱، پایان‌نامه دوره کارشناسي ارشد ،تشخیص اثر انگشت به روش جدیدی مبتنی بر قطاع بندی، مهندسي برق-الکترونیک، دانشگاه شاهرود

[21] http://fa.wikipedia.org/wiki/

[22] http://adishict.com/AI&Robtic/PT/DR_PR_methods_.pdf

[23] Duda R. O, Hart P. E and Stork D. G, (2000), "Pattern Classification", second edition.

[24] Ramalingam and S. Krishnan, (2006), "Gaussian Mixture Modeling of Short-Time Fourier Transform Features for Audio Fingerprinting", IEEE transactions on information forensics and security, 1, 4.

[25] X. T. Bhanu and B. Y. Lin, " text detection based on learned features," presented at the Center for Res. in Intelligent Syst., Univ. of California, Riverside,CA, USA, 2005

[26] G.Liu .,Qian.",Effective and efficient video text extraction using key text point," IET Image Process.,2011,vol.5,Iss,pp.671-683

[27] J.Weinman, E.Leanred,and A.Hanson,"Scene text recognition using similarity and a lexicon with sparse belief propagation," IEEE Trans. Pattern Anal. Mach. Intell.,vol.31,no.10,pp.1733-1746, Oct.2009

[28] N Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst. Man Cybern., vol. 9, no. 1, pp. 62–66, Jan. 1979.

[29] Lucas S. M., Panaretos, L.Sosa, A.Tang, S.Wong, and R.Young. "ICDAR 2003 Robust Reading Competitions " , Proceedings of International Conference on Document Analysis and Recognition, pp.682-687 , 2003

[30] http://algoval.essex.ac.uk/icdar/Datasets.html

[31] Epshtein B., E. Ofek, and Y. Wexler, "Detecting Text in Natural Scenes with Stroke Width Transform", Proceedings of IEEE Confrence on Computer Vision and Pattern Recognition, 2010.

[32] http://vision.ucsd.edu/~kai/svt /

[33] Icdar 2005 text location competition results. In Proc. ICDAR.volume 1. Pages 80-84, 2005.

[34] X.Liu .Effectively Localize Text in Natural Scene Images.In Proc.ICPR, November 11-15,2012

[35] C. Yi and Y. Tian , Text string detection from natural scenes by structure-based partition and grrouping, IEEE Trans.IP,2011.

[36] X.Chen and A.Yuille. Detecting and reading text in natural scenes. In Proc CVPR.2004

[37] Kim J. N., and Choi T. S., (1998), "A fast three step search algorithm with minimum checking points using unimodal error surface assumption", IEEE Transactions on consumer electronics, Vol. 44, No. 3, pp. 638-648

[38] R.Minetto, N.Thome,M.Cord, J.Fabrizio, and B.Marcotegui, "Snoopertext: A multire solution system for text detection in complex visual scens.," in ICIP, pp.3861-3864.2010.

[39] J.Fabrizio,M.Cord, and Marcotegui, "Text extraction from street level images", in CMRT,2009, pp.199-204.

Abstract

Nowadays, extraction of text in natural scenes is addressed by many researchers. Scene content can be classified into two important categories: perceptual content and semantic content. Perceptual content includes color characteristics, shape and scene texture. semantic content includes text, human's face, behaviors and actions. Among various information of scene, textual information has more importance, because they are intelligible by human and computer, and provides possibility of describing the scene content.

In this thesis, we propose a method to extract text regions from images of natural scene with complex background. The algorithm contains four main steps. At first step we extract candidate regions for text using gradient cue. At next step we cluster extracted regions according to this fact that the consisting features of a text row in an image have approximately same height and direction. Finally we exploit features of histogram of gradient magnitude and gradient orientation in extracted regions to filter out non-text regions. For this purpose, we train a classifier based on support vector machine (SVM). This classifier is trained by histogram features of gradient magnitude and gradient orientation of text regions and non-text regions. Then, the results will be improved by determining distance criterion based on width of detected textual regions and horizontal projection. The evaluation results obtained from applying the proposed method on scenes which have English and Persian text with different fonts and simple and complex background. According to evaluation of experimental results of text detection on three datasets: ICDAR 2003/2005, Microsoft Street View Text Detection and a collection of Farsi text images. The proposed text detection method can cope with variation in type of font, size, color and slightly direction. In comparison to other methods this result is very promising.

Keywords- Text extraction; Gradient histogram; clustering; support vector machine (SVM)

Shahrood University of Technology

Faculty of Electrical and Robotic

Extracting Text from Natural Scene Images

Maryam Sabzevari

Supervisor:

Dr. Alireza Ahmadyfard

Date: February 2014