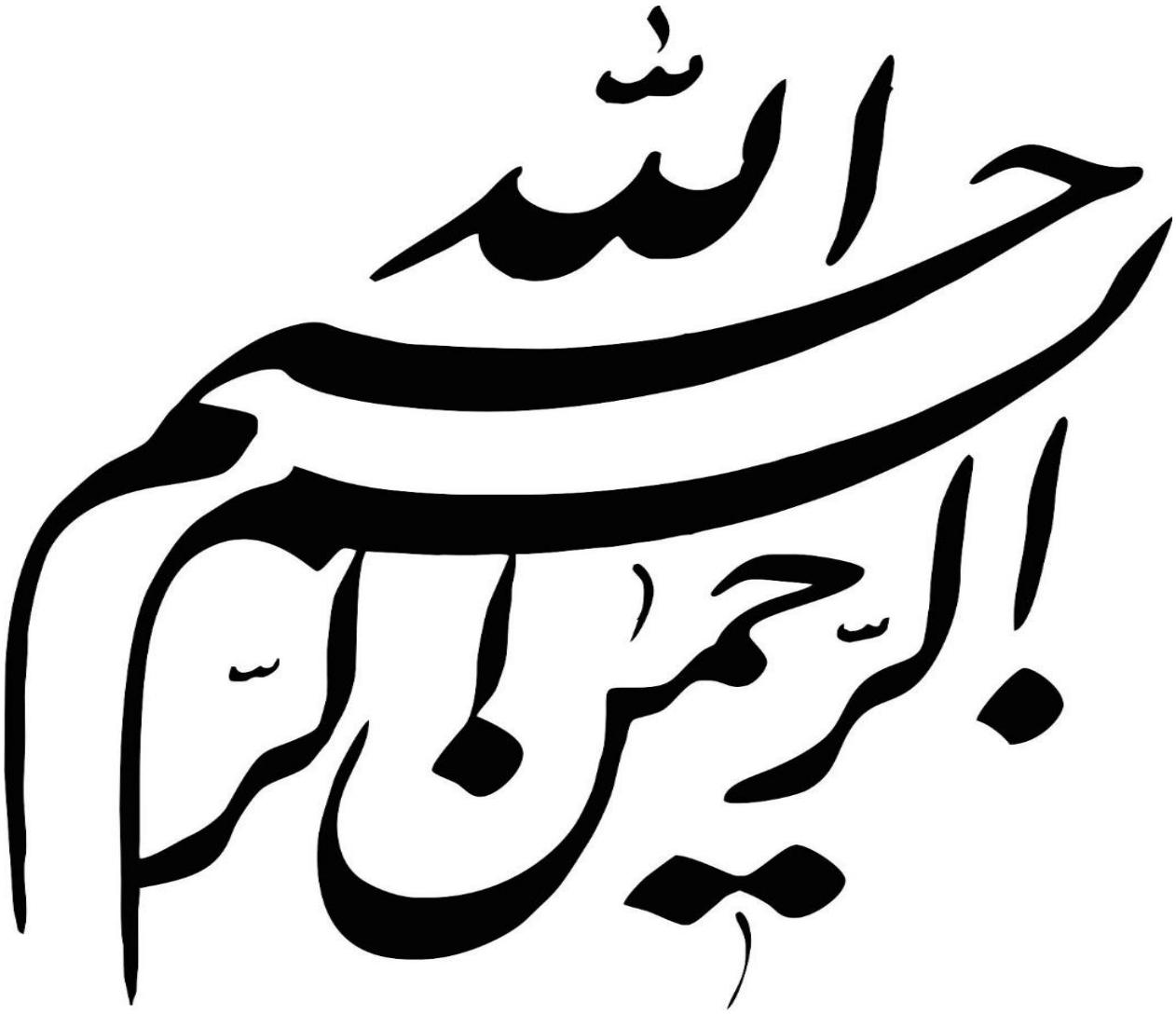

بسم الله
الرحمن الرحیم

دانشکده مهندسی برق و رباتیک

گروه الکترونیک

پایان‌نامه کارشناسی ارشد

طراحی یک سیستم فشرده‌سازی/بازسازی تصاویر متنی با درجه‌ی تفکیک مکانی بالا مبتنی بر فرا تفکیک‌پذیری

سعید مرادی

استاد راهنما:

دکتر هادی گرایلو

شهریور ۱۳۹۵

دانشگاه صنعتی شاهرود

دانشکده مهندسی برق و رباتیک
گروه الکترونیک

پایان‌نامه کارشناسی ارشد آقای سعید مرادی
تحت عنوان: طراحی یک سیستم فشرده سازی/بازسازی تصاویر متنی با درجه ی تفکیک مکانی بالا مبتنی بر فراتفکیک پذیری

در تاریخ ..... توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد مورد ارزیابی و با درجه ..... مورد پذیرش قرار گرفت.

استاد راهنما	امضاء	استاد مشاور	امضاء
دکتر هادی گرایلو			

اساتید داور	امضاء	نماینده تحصیلات تکمیلی	امضاء
نام و نام خانوادگی :		نام و نام خانوادگی :	
نام و نام خانوادگی :			
نام و نام خانوادگی :			
نام و نام خانوادگی :			

تقدیم به پدر و مادرم

که از نگاهشان صلابت

از رفتارشان محبت

و از صبرشان ایستادگی را آموختم..

تشکر و قدردانی

شکر شایان نثار ایزد منان که توفیق را رفیق راهم ساخت تا این پایان‌نامه را به پایان برسانم. از استاد فاضل و اندیشمند جناب آقای دکتر گرایلو به عنوان استاد راهنما که همواره نگارنده را مورد لطف و محبت خود قرار داه‌اند و همه کسانی که در انجام این راه به من کمک نمودند، کمال تشکر را دارم.

سعید مرادی

شهریور ۱۳۹۵

تـــعـهـد نـــامـــه

اینجانب **سعید مرادی** دانشجوی مقطع کارشناسی ارشد رشته **مهندسی برق-الکترونیک سیستم** دانشکده **مهندسی برق و رباتیک** دانشگاه صنعتی شاهرود نویسنده پایان نامه **طراحی یک سیستم فشرده‌سازی/بازسازی تصاویر متنی با درجه‌ی تفکیک مکانی بالا مبتنی بر فرا تفکیک‌پذیری** تحت راهنمائی **دکتر هادی گرایلو** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام «دانشگاه صنعتی شاهرود» و یا « **Shahrood University of Technology** » به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافت‌های آن‌ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ

سعید مرادی

چکیده:

در این پایان‌نامه یک سیستم فشرده‌سازی/بازسازی برای تصاویر متنی[1] با درجه‌ی تفکیک مکانی بالا[2] مبتنی بر فرا تفکیک‌پذیری[3] طراحی شده‌است. در تصاویر متنی درجه‌ی تفکیک مکانی از اهمیت بالایی برخوردار است اما این ویژگی باعث می‌شود که حجم ذخیره‌سازی زیاد شود. با توجه به محدود بودن حافظه‌های دیجیتالی و حجم زیاد این تصاویر نیاز به فشرده‌سازی تصاویر متنی امری ضروری به‌نظر می‌رسد. اما مشکلاتی وجود دارد، ۱) فشرده‌سازی می‌تواند موجب اثرات مخربی روی تصاویر متنی شود. ۲) در روش‌هایی مانند JPEG و JPEG2000، میزان فشرده‌سازی یا نرخ بیت[4] خروجی محدود است.

ایده اصلی در این پایان‌نامه این بوده‌است که با ارائه روشی مناسب بتوان میزان فشرده‌سازی را بیشتر و اثرات مخرب روی تصویر فشرده‌شده را نیز کم کرد. برای رسیدن به میزان فشرده‌سازی بیشتر از ایده کاهش ابعاد در تصاویر متنی استفاده شده‌است. البته باید به این موضوع توجه شود که کاهش ابعاد روی کیفیت تصویر بی تأثیر نیست و ممکن است باعث تنزل در کیفیت تصویر شود. بنابراین باید روشی انتخاب شود که بتواند هم ابعاد تصویر را افزایش دهد و هم اثرات مخرب تأثیر گذار بر تصویر را اصلاح کند. به همین دلیل از ایده فرا تفکیک‌پذیری در ترکیب با روش‌های فشرده‌سازی استفاده شده‌است.

از روش‌های فرا تفکیک‌پذیری مبتنی بر درون‌یابی و مبتنی بر یادگیری در مرحله بازسازی استفاده شده است. در روش مبتنی بر درون‌یابی، تصویر وضوح پایین[5] ورودی به سه لایه تقسیم و سپس هر لایه

[1] Textual Images
[2] High Resolution
[3] Super Resolution
[4] Bit Rate
[5] Low Resolution

براساس اهمیت اطلاعاتی آن با یک روش بزرگ‌نمایی شده‌است. در نهایت لایه‌های بزرگ‌نمایی شده با هم ترکیب شده‌اند و تصویر وضوح بالای نهایی را تشکیل داده‌اند.

در روش مبتنی بر یادگیری، ابتدا پایگاه داده‌ای مناسب از تصاویر متنی وضوح بالا و وضوح پایین (نسخه‌های کاهش ابعاد داده شده و فشرده‌شده‌ی تصاویر وضوح بالا) ایجاد شده‌است. برای ساخت واژه‌نامه[1] از وصله‌های ۳*۳ و ۵*۵ استفاده شده است. واژه‌نامه‌ها مبتنی بر ویژگی بوده‌اند و از چهار فیلتر بالاگذر[2] برای داشتن مؤلفه‌های فرکانس بالای تصاویر به‌منظور ساخت واژه‌نامه‌ها استفاده شده است. واژه‌نامه‌ها برای ۵۰۰۰۰ و ۱۰۰۰۰۰ اتم آموزش داده شده‌اند.

پایگاه داده‌ی ایجاد شده شامل تصاویر متنی واقعی با درجه‌ی تفکیک مکانی بالا و متوسط بوده است. برای تصویری با درجه‌ی تفکیک ۶۰۰dpi، روش JPEG[3] و JPEG2000 توانستند به‌ترتیب تا ۰/۰۷ و ۰/۰۱۵ نرخ بیت فشرده‌سازی را انجام دهند؛ روش پیشنهادی برای همان تصویر تا ۰/۰۰۴ نرخ بیت در هر دو روش توانسته نتایج قابل قبولی داشته باشد. نتایج براساس معیارهای بازشناسی متن[4] (OCR)، متوسط امتیاز نظرسنجی[5] (MOS) و پیک سیگنال به نویز[6] (PSNR) ارزیابی و با هم مقایسه شده‌اند. روش پیشنهادی از نظر معیارهای OCR و MOS که در تصاویر متنی مهم هستند جواب بهتری نسبت به روش‌های دیگر داشته‌است اما از نظر PSNR جواب خوبی بدست نیامده‌است.

واژه‌های کلیدی: فشرده‌سازی تصاویر متنی، فشرده‌سازی JPEG، فشرده‌سازی JPEG2000، فشرده‌سازی SPIHT[7]، فرا تفکیک‌پذیری، شناسایی نوری کاراکترها، تصاویر متنی با درجه تفکیک مکانی بالا.

[1] Dictionary
[2] High Pass
[3] Joint Photographic Experts Group
[4] Optical Character Recognition
[5] Mean Opnion Score
[6] Peak Singnal-to-Noise Ratio
[7] Set partitioning in hierarchical trees

فهرست مطالب

فهرست شکل‌ها

فصل اول : مقدمه

۱-۱ مقدمه

امروزه با توجه به اینکه منابع الکترونیکی به صورت گسترده مورد استفاده قرار می‌گیرد، سعی بر آن است که برای نسخه‌های چاپی، نسخه‌های الکترونیکی آن‌ها را نیز ذخیره کنند و در اختیار کاربران قرار دهند. دلیل این امر این است که ذخیره اطلاعات به شکل دیجیتالی، تا حد زیادی مقرون به‌صرفه‌تر از نگهداری آن‌ها در قفسه‌های کتابخانه‌ها است. پس تصاویر متنی در دنیای امروز اهمیت پیدا می‌کند. اما مسئله‌ای که وجود دارد به حافظه لازم جهت ذخیره کردن منابع الکترونیکی که در این مسئله تصاویر متنی هستند برمی‌گردد. چرا که تصاویری که با درجه تفکیک مکانی بالا اسکن می‌شوند دارای حجم فایل زیادی می‌باشند. از این‌رو نیاز به فشرده‌سازی تصاویر متنی اهمیت پیدا می‌کند.

حجم عظیم تصاویر اسناد متنی که امروزه به‌طور روزانه در مراکزی مانند کتابخانه‌های دیجیتال و مراکز بایگانی اسناد در حال تولید است، سبب می‌شود که نگهداری و مدیریت آن‌ها چالشی مهم محسوب شود. این حجم عظیم از داده‌ها در کنار محدود و پرهزینه بودن حافظه‌های دیجیتالی امروزی، موجب توجه جدی به روش‌های تخصصی فشرده‌سازی تصاویر اسناد متنی شده است. روش‌های فشرده‌سازی به دو دسته‌ی بااتلاف و بدون اتلاف تقسیم می‌شوند. در کاربردهایی که اشاره شد، عمدتاً از روش‌های فشرده‌سازی بااتلاف استفاده می‌شود زیرا هم میزان فشرده‌سازی بسیار بالاتری نسبت به روش‌های بدون اتلاف فراهم می‌کنند و هم این که کیفیت تصویر فشرده شده تا حدی، قابل کنترل است.

در برخی کاربردها، مانند نگهداری دیجیتالی اسناد تاریخی، حفظ کیفیت اولیه‌ی تصویر مهم است؛ بنابراین، در این کاربردها، معمولاً از روش‌های فشرده‌سازی بدون اتلاف و یا روش‌های فشرده‌سازی بااتلاف اما با میزان فشرده‌سازی کم تا متوسط استفاده می‌شود. در مقابل، در برخی کاربردها، کاهش حجم حافظه‌ی ذخیره‌سازی مهم‌تر از کیفیت تصویر فشرده شده است. برای مثال، در مراکز بایگانی دیجیتالی اسناد متنی معمولی و روزمره و نیز برخی کتابخانه‌های دیجیتالی، حساسیت چندانی روی کیفیت تصویر فشرده شده

نمی‌باشد؛ بلکه همین مقدار که خوانایی انسانی/ماشینی تصویر متنی تا حد مناسبی حفظ شود، کافی است. بنابراین با توجه به محدود بودن حجم حافظه‌های موجود و نیز قیمت بالای آن‌ها، این مراکز تمایل زیادی به استفاده از روش‌های فشرده‌سازی بااتلاف و با میزان فشرده‌سازی بالا یا همان کدگذاری با نرخ بیت بسیار پایین دارند. در حالت کلی، تصاویر متنی شامل اجزای متنی (شامل متن و ترسیمات خطی) و اجزای غیرمتنی (شامل پس‌زمینه و قسمت‌های گرافیکی) می‌باشند. در روش‌های فشرده‌سازی با میزان فشرده‌سازی بالا که مورد توجه مراکز مذکور هستند، معمولاً تنها حفظ نسبی اجزای متنی مهم بوده و اگر در این کاربردها، کیفیت اجزای غیرمتنی تا حدی اُفت کند یا دست‌خوش تغییرات شود، اتفاق چندان نامطلوبی تلقی نمی‌شود. از آن جا که در این پایان‌نامه، این نوع از کاربردها مدنظر هستند، در ادامه مطالبی که در استفاده از این دسته از روش‌های فشرده‌سازی باید در نظر گرفته شوند، آورده می‌شود.

روش‌های امروزی موجود برای فشرده‌سازی تصاویر طبیعی را نمی‌توان برای فشرده‌سازی تصاویر اسناد با نرخ فشرده‌سازی زیاد مورد توجه قرار داد، زیرا چنین روش‌هایی از ویژگی‌های مخصوص به متن برای افزایش نرخ فشرده‌سازی استفاده نمی‌کنند. یک تفاوت تصاویر متنی و بیشتر تصاویر طبیعی این است که در تصاویر متنی عمده‌ی اطلاعات در لبه‌ها قرار داشته و نواحی داخلی نویسه‌ها و علائم متنی چندان حاوی اطلاعات نمی‌باشند. اما در بیشتر تصاویر طبیعی، قسمت مهمی از اطلاعات در بدنه‌ی اشیاء موجود در تصویر قرار دارد. به همین دلیل می‌توان گفت یک ویژگی مهم تصاویر متنی این است که طیف این نوع از تصاویر، طیفی میان‌گذر - یا شامل مولفه‌های فرکانسی میانی مهم – است اما طیف تصاویر طبیعی، پایین گذر است [۱]. یک ویژگی مهم دیگر تصاویر متنی این است که در این تصاویر، برخلاف بیشتر تصاویر طبیعی، درجه‌ی تفکیک مکانی بیشتر از درجه‌ی تفکیک مقادیر شدت روشنایی در حفظ کیفیت یا اطلاعات اولیه اهمیت دارد [۲]. تفاوت دیگر بین تصاویر متنی و تصاویر طبیعی این است که تصاویر متنی، برخلاف بیشتر تصاویر طبیعی، شامل جزئیات دیداری مهم مانند شکل و انحنای قلم، رنگ کاغذ، و بافت کاغذ هستند. این مسأله

در مورد تصاویر متنی تاریخی، تصاویر متنی دست‌نویس، و نیز تصاویر شامل جداول و روابط ریاضی اهمیت ویژه‌ای پیدا می‌کند. ویژگی‌های تصاویر متنی آن قدر مهم و متمایز از دیگر انواع تصاویر است که بخش ۶ از استاندارد JPEG2000 ویژه تصاویر متنی بوده و پسوند فایل JPM را به خود اختصاص داده است [۳].

روش‌های فشرده‌سازی تصاویر متنی را می‌توان به دو دسته‌ی روش‌های مبتنی بر مدل و روش‌های غیرمبتنی بر مدل تقسیم‌بندی کرد. روش‌های مبتنی بر مدل، از مدل محتوای ترکیبی (MRC[1]) استفاده نموده و تصویر متنی را به لایه‌های مختلفی تجزیه می‌کنند. روش‌های غیرمبتنی بر مدل، عمدتاً به دنبال ارتقاء کارایی فشرده‌سازی یکی از روش‌ها یا استانداردهای نسبتاً عمومی موجود به‌منظور استفاده در کاربرد فشرده‌سازی تصاویر متنی می‌باشند. با توجه به نوع روش فشرده‌سازی تصاویر متنی می‌توان اثرات نامطلوبی را در تصویر فشرده شده مشاهده نمود. این اثرات با افزایش میزان فشرده‌سازی یا با کاهش حجم فایل خروجی حاصل از فشرده‌ساز بیشتر می‌شوند. امروزه از روش‌های مختلفی برای فشرده‌سازی تصاویر متنی استفاده می‌شود که روش‌های فشرده‌سازی مبتنی بر تبدیل موجک جزو روش‌های مؤثر محسوب می‌شوند [۴]. در تصاویر پس از فشرده‌سازی بخشی از جزئیات یا لبه‌ها در تصویر از بین می‌رود و هم‌چنین در برخی روش‌ها اثر بلوکی نیز در تصویر دیده می‌شود [۵]. در واقع در فشرده‌سازی بر اثر خطای کوانتیزاسیون و اعوجاج ناشی از فشرده‌سازی کیفیت تصویر کم می‌شود. دانستن اطلاعاتی در رابطه با اثرات فشرده‌سازی روی تصاویر متنی می‌تواند به بهبود کارایی روشی که برای بازسازی تصاویر فشرده شده مورد استفاده قرار می‌گیرد کمک کند.

از ویژگی‌های مهم تصاویر متنی درجه تفکیک مکانی بالای این تصاویر می‌باشد؛ مشکلی که این ویژگی تصاویر متنی ایجاد می‌کند زیاد شدن حجم فایل تصویر با افزایش درجه تفکیک مکانی می‌باشد. بنابراین نیاز به فشرده‌سازی تصاویر متنی امری ضروری به نظر می‌رسد. روش‌های متفاوتی برای فشرده‌سازی

[1] Mixed Raster Content (MRC)

تصاویر متنی وجود دارد که هر کدام از آنها در نرخ بیت‌های پایین اثرات مخربی مانند اثر بلوکی و نشتی را روی تصاویر متنی ایجاد می‌کنند. روش‌های فشرده‌سازی مانند JPEG و JPEG2000 تا نرخ بیت مشخصی می‌توانند یک تصویر را فشرده کنند. سوالی که پیش می‌آید این است که برای نرخ بیت‌های کمتر چه راه حلی وجود دارد؟ راه حلی که در این پایان‌نامه مطرح شده‌است کاهش ابعاد تصویر با درجه‌ی تفکیک مکانی بالای ورودی با یکی از روش‌های موجود مانند دومکعبی و در ادامه فشرده‌کردن تصویر کاهش ابعاد داده شده بوده‌است. البته باید به این نکته توجه نمود که کاهش ابعاد می‌تواند کیفیت تصویر را دچار تنزل نماید. به‌طور مثال ممکن است که منجر به ماتی لبه‌ها شوند. از آنجا که لبه‌ها در تصاویر متنی مهم‌ترین بخش هستند این اثر برای کاربر خوشایند نیست.

با توجه به اثرات مخرب ناشی از فشرده‌سازی تصاویر متنی و کاهش ابعاد تصویر ورودی برای رسیدن به نرخ بیت‌های کمتر، در مرحله بازسازی نیاز به روشی است که بتواند به‌طور همزمان هم اثرات مخرب را بهبود دهد و هم ابعاد تصویر را بزرگ کند. برای این منظور از روش‌های فرا تفکیک‌پذیری در مرحله بازسازی استفاده شده است. طبق اطلاعات ما تا کنون کاری در زمینه ترکیب روش‌های فرا تفکیک‌پذیری با روش‌های فشرده‌سازی تصاویر متنی انجام نشده است.

برای رسیدن همزمان به بهبود و افزایش ابعاد تصاویر روش‌های زیادی ارائه شده‌اند. یکی از روش‌های معروف، روش فراتفکیک‌پذیری می‌باشد. فراتفکیک‌پذیری فرآیند به‌دست‌آوردن یک یا چند تصویر با تفکیک-پذیری بالا از روی یک یا چند تصویر با تفکیک‌پذیری پایین است.

۱-۲ فرا تفکیک‌پذیری

درون‌یابی تصویر موضوعی جذاب در زمینه‌ی پردازش تصاویر دیجیتال است. در سیستم‌های دیجیتال یک سیگنال توسط تعداد محدودی از نمونه‌ها ارائه می‌شود. درون‌یابی این سیگنال‌ها توسط نمونه‌هایی که مشخص هستند، انجام می‌گیرد. چنین مسئله‌ای یک مسئله بدحالت است چون سیگنال‌هایی را می‌توان

یافت که نمونه‌های یکسانی داشته باشند. در بیشتر موارد فرض بر این است که سیگنال‌ها باند محدود هستند. درون‌یابی در صورتی امکان‌پذیر است که فرآیند نمونه‌برداری داده‌ها از شرط نایکوئیست پیروی کند. این شرط ممکن است که همیشه برقرار نباشد. فن‌های فرا تفکیک‌پذیری برای افزایش وضوح فضایی استفاده می‌شوند. در واقع مقادیر نمونه‌های جدید توسط همسایه‌ها تخمین زده می‌شود، به‌طوریکه سیگنال جدید جزئیات بیشتر و اطلاعات بهتری را فراهم می‌آورد.

الگوریتم‌های فراتفکیک پذیری را می‌توان براساس معیارهای متفاوتی طبقه‌بندی کرد. این معیارها شامل حوزه کاربردی، تعداد تصاویر در دسترس با وضوح پایین و روش بازسازی حقیقی است. به طور مثال در [۶] طبقه‌بندی وسیعی از الگوریتم‌های فرا تفکیک‌پذیری را ابتدا براساس حوزه کاربرد آنها (مکان و فرکانس) انجام داده است.

۱-۲-۱ حوزه فرکانس

الگوریتم‌های فرا تفکیک‌پذیری در حوزه فرکانس ابتدا تصویر یا تصاویر با وضوح پایین را به حوزه فرکانس می‌برند و در این حوزه تصویر با وضوح بالا را تخمین می‌زنند سرانجام تصویر با وضوح بالای ساخته شده را به حوزه مکان انتقال می‌دهند. براساس تبدیلات انجام شده برای تبدیل تصاویر به حوزه فرکانس، این الگوریتم‌ها به طور کلی به دو گروه؛ روش‌های مبتنی بر تبدیل فوریه و مبتنی بر تبدیل موجک تقسیم می‌شوند.

- روش‌های مبتنی بر تبدیل فوریه

در سال ۱۹۷۴ Gerchberg [۷] و سپس Santis و Gori [۸] اولین الگوریتم‌های فرا تفکیک‌پذیری را معرفی کردند. این الگوریتم‌ها به‌صورت روش‌های تکرار[1] مبتنی بر تبدیل فوریه در حوزه فرکانس بودند

[1] Iterative

[۹]. این الگوریتم‌ها اگرچه دوباره در شکل‌های بدون تکرار مبتنی بر تجزیه مقدار منحصر به فرد[1] معرفی شدند [۱۰] اما نتوانستند شهرت روش Tsai و Huang [۱۱] را به دست بیاورند.

سیستم Tsai و Huang [۱۱] اولین الگوریتم فرا تفکیک‌پذیری چند تصویر در حوزه فرکانس بود. این الگوریتم برای کار بر روی تصاویر به دست آمده از ماهواره‌ی Landsat ۴ توسعه یافت. این ماهواره مجموعه‌ای از تصاویر مشابه را که به صورت کلی نسبت به هم دارای جابجایی بودند، از یک ناحیه از زمین تولید می‌کرد. در الگوریتم معرفی شده توسط Tsai و Huang این جابجایی بین پیکسلی در تصاویر وضوح پایین ارسالی از ماهواره در حوزه تبدیل فوریه به‌صورت ویژگی جابجایی زمانی تبدیل فوریه در نظر گرفته شد و براساس آن مدلی برای فرا تفکیک‌پذیری تعریف شد.

- روش‌های مبتنی بر تبدیل موجک

تبدیل موجک بعنوان جایگزینی برای تبدیل فوریه به‌صورت گسترده در الگوریتم‌های فرا تفکیک-پذیری در حوزه فرکانس استفاده شده است. به‌طور معمول این تبدیل به منظور تجزیه تصویر ورودی به زیر تصویرهای ساختاری مشابه مورد توجه استفاده کنندگان قرار گرفت. این کار منجر می‌شود که از خود تشابهی بین نواحی همسایگی محلی استفاده کند [۱۳-۱۲]. برای مثال در [۱۳] تصاویر ورودی ابتدا به زیر باندهایش تجزیه می‌شود. سپس تصویر ورودی و زیر باندهای فرکانس بالای آن هر دو درون‌یابی می‌شوند. سپس نتایج حاصل از تبدیل فوریه ایستا[2]ی زیر باندهای فرکانس بالا به‌منظور بهبود زیر باندهای درون‌یابی شده مورد استفاده قرار می‌گیرند. در نهایت خروجی با تفکیک‌پذیری بالا با ترکیب تمام این زیر باندها با استفاده از عکس تبدیل موجک گسسته[3] ایجاد می‌شود.

[1] Singular Value Decomposition
[2] Stationary Wavelet Transform (SWT)
[3] Discrete Wavelet Transform (DWT)

۱-۲-۲ حوزه مکان

براساس تعداد مشاهدات با تفکیک‌پذیری پایین موجود، الگوریتم‌های فرا تفکیک‌پذیری می‌توانند به دو دسته الگوریتم‌های مبتنی بر تک تصویر و چند تصویر تقسیم شوند.

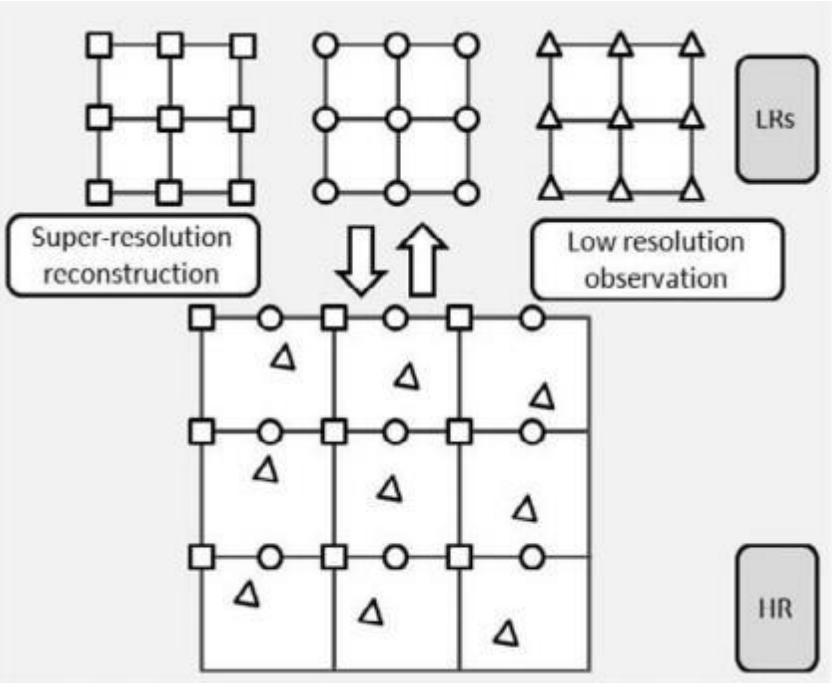
- الگوریتم‌های مبتنی بر چند تصویر

در این روش‌ها با استفاده از ترکیب اطلاعات چند تصویر وضوح پایین یک تصویر وضوح بالا به‌دست می‌آید. به عبارتی مؤلفه‌های فرکانس بالا افزایش پیدا می‌کنند [۱۷-۱۴]. ایده‌ی اصلی در این روش‌ها این است که از اطلاعات مفید مربوط به چند فریم که از یک صحنه گرفته شده‌اند برای تولید تصویر وضوح بالا استفاده می‌شود. اطلاعات مفیدی که در این تصاویر به وجود می‌آید توسط جابجایی‌های زیر پیکسلی بین فریم‌ها معرفی می‌شوند. این جابجایی‌های زیر پیکسلی معمولاً به دلیل حرکات غیرقابل کنترل بین سیستم تصویربرداری و صحنه همانند حرکات شی و یا به‌وسیله‌ی حرکات قابل کنترل همانند حرکت سیستم تصویربرداری ماهواره‌ای که با سرعت مشخص و از پیش تعریف شده‌ای به دور زمین می‌چرخد، به وجود می‌آید. بنابراین فراتفکیک‌پذیری چند فریمی در صورتی امکان‌پذیر است که بین پیکسل‌های موجود در فریم‌ها شیفت‌های زیر پیکسلی وجود داشته باشد تا بتواند مشکل بدحالت بودن این مسئله را حل کند [۱۴]. شکل ۱-۱ دیاگرام ساده‌ای از ایده‌ی بازسازی فرا تفکیک‌پذیری را نشان می‌دهد.

- الگوریتم‌های مبتنی بر تک تصویر

دسته‌ی دیگری از روش‌های فراتفکیک‌پذیری، حالت تک تصویر می‌باشد، به‌طوری‌که تنها یک تصویر وضوح پایین برای تولید تصویر وضوح بالا مورد نیاز است. این روش در کاربردهایی که چند تصویر وضوح پایین در دسترس نباشد بسیار مؤثر خواهد بود. دقت الگوریتم‌های فرا تفکیک‌پذیری چند فریمی به دقت تخمین حرکت تصاویر وضوح پایین وابسته است. این محدودیت در کاربردهای واقعی استفاده از الگوریتم‌های

فرا تفکیک‌پذیری چند فریمی را با مشکل مواجه می‌کند. زیرا ممکن است اشیای مختلف در صحنه‌های یکسان حرکت‌های پیچیده و



شکل ۱-۱ ایده اصلی فراتفکیک‌پذیری چند فریمی. شیفت‌های زیر پیکسلی اطلاعات تکمیل کننده فریم‌های وضوح پایین می‌باشند و به کمک آن‌ها تصویر وضوح بالا ساخته می‌شود.

متفاوتی داشته باشند. ازاین‌رو الگوریتم‌های فرا تفکیک‌پذیری تک تصویر برای غلبه بر این مشکلات به وجود آمده‌اند. در الگوریتم‌های فرا تفکیک‌پذیری تک تصویر برای بازیابی اطلاعات فرکانس بالا به ترم پیشین (همانند استفاده از یک پایگاه داده خارجی و یا استفاده از هرم تصویر ورودی) نیاز است. این روش‌ها در فصل بعد به تفصیل بیان می‌شوند.

۱-۳ کاربردهای فرا تفکیک‌پذیری

فرا تفکیک‌پذیری در بسیاری از زمینه‌ها کاربرد دارد:

۱- نظارت بر ویدیو [۱۸, ۱۹]: بزرگ‌نمایی منطقه موردنظر (ROI) در یک ویدیو برای شناسایی افراد یا خواندن پلاک خودرو.

۲- سنجش از راه دور [۲۰]: به کمک چندین تصویر وضوح پایین که از یک منطقه گرفته می‌شود یک تصویر وضوح بالا می‌سازند.

۳- پردازش تصاویر پزشکی (CT, MRI, …) [۲۱]: با استفاده از چندین تصویر وضوح پایین یک تصویر وضوح بالا به وجود می‌آورند.

۴- شناسایی و تشخیص هدف [۲۲]

۱-۴ هدف پایان‌نامه

در این پایان‌نامه سعی بر این است که بتوان در فشرده‌سازی تصاویر متنی تا جایی که امکان‌پذیر باشد به میزان فشرده‌سازی بالاتر یا حجم فایل کمتر و یا نرخ بیت کمتری دست یافت. با توجه به اینکه تصاویر مورد بررسی در این پایان‌نامه تصاویر متنی با درجه‌ی تفکیک مکانی بالا هستند و این ویژگی در تصاویر متنی سبب حجم فایل زیاد این تصاویر می‌شود، سعی شده تا این نوع از تصاویر تا جایی که امکان دارد در نرخ بیت کمتری ذخیره شوند. برای رسیدن به این منظور ابتدا ابعاد تصویر ورودی کاهش داده و سپس تصویر کوچک شده با یکی از روش‌های فشرده‌سازی موجود فشرده می‌شود و تصویری با میزان فشرده‌سازی بیشتر بدست می‌آید. با این کار می‌توان به حجم فایل‌های خیلی کم نیز دسترسی پیدا نمود. اما اعمال کاهش ابعاد و فشرده‌سازی تصاویر متنی خود باعث ایجاد اثرات مخربی مانند ماتی، اثر بلوکی و یا نشتی روی تصویر می‌شوند.

پس از فشرده‌سازی تصویر متنی ورودی و رسیدن به نرخ بیت مورد نظر و ذخیره‌سازی این تصویر با توجه به اثرات فشرده‌سازی و کاهش ابعاد تصویر، خوانایی تصویر متنی بسیار کم می‌شود. از آنجا که در

تصاویر متنی، موارد استفاده یا برای تشخیص نوری کاراکترها[1] و یا استفاده انسان از این نوع تصاویر می‌باشد، در مرحله بازسازی تلاش بیشتر روی بهبود تصاویر فشرده‌شده براساس این دو معیار متمرکز بوده‌است. محدودیت در کم کردن میزان فشرده‌سازی تصاویر متنی را نیز این دو معیار مشخص می‌کنند. به‌عبارت دیگر تا جایی فشرده‌سازی انجام داده‌شده که تصاویر متنی جواب‌های قابل قبولی از نظر معیارهای خوانایی انسانی و ماشینی داشته باشند.

پس از رسیدن به نرخ بیت مورد نظر نوبت به مرحله بازسازی می‌رسد. برای بازسازی از روش‌های فراتفکیک‌پذیری استفاده شده‌است. زیرا می‌خواهیم به‌طور همزمان ابعاد تصویر را بزرگ و اثرات مخرب روی تصویر را نیز اصلاح کنیم.

در مرحله بازسازی مشکلاتی وجود دارد. می‌توان مشکلات را به دو گروه تقسیم کرد؛ ۱) مشکلات ناشی از اثرات فشرده‌سازی و کاهش ابعاد ۲) زمان بازسازی.

پس می‌توان چنین بیان کرد که می‌خواهیم میزان فشرده‌سازی را تا حد امکان افزایش، اثرات مخرب ناشی از فشرده‌سازی و کاهش ابعاد روی تصاویر متنی را بهبود و در مرحله بازسازی زمان بازسازی را بهبود دهیم.

۱-۵ ساختار پایان‌نامه

در فصل ۲ مروری بر روش‌های انجام گرفته در زمینه‌ی فشرده‌سازی تصاویر متنی و فرا تفکیک-پذیری تک تصویر ارائه می‌شود. در فصل ۳ تئوری‌های مورد نیاز برای انجام فرآیند فشرده‌سازی و فرا تفکیک‌پذیری مطرح می‌شود. در فصل ۴ روش پیشنهادی مطرح شده‌است. در فصل ۵ نتایج تجربی نشان

[1] Optical Character Recognition

داده‌شده‌است. در نهایت در فصل ۶ پایان‌نامه نتیجه‌گیری شده و پیشنهادهایی برای کارهای آتی ارائه شده-است.

فصل دوم:

مروری بر کارهای انجام شده

در سال‌های اخیر تلاش‌های زیادی در زمینه‌ی پردازش تصویر صورت گرفته است. علت اصلی این تلاش‌ها این است که اکثریت داده‌هایی که از طریق مشاهده یا فن‌های پردازش تصویر به دست انسان‌ها می‌رسند در زمینه‌های گسترده‌ای هم‌چون تصاویر پزشکی، نظارت و کنترل از راه دور کاربرد دارند؛ بنابراین در بسیاری از زمینه‌ها نیاز به داشتن تصویری با وضوح بالا رو به افزایش است. تصاویری که در این پایان نامه مورد بررسی قرار می‌گیرند تصاویر متنی هستند. از ویژگی‌های مهم این تصاویر همان‌طور که قبلا نیز گفته شد، درجه تفکیک مکانی بالای این تصاویر می‌باشد؛ مشکلی که این ویژگی تصاویر متنی ایجاد می‌کند زیاد شدن حجم فایل تصویر با افزایش درجه تفکیک مکانی می‌باشد. پس نیاز به فشرده‌سازی تصاویر متنی امری ضروری به نظر می‌رسد. از طرف دیگر پس از فشرده‌سازی تصاویر متنی با توجه به اثرات مخرب ناشی از فشرده‌ساز و کاهش ابعاد تصویر ورودی برای رسیدن به حجم فایل‌های کم، در مرحله بازسازی نیاز به روشی است که بتواند به‌طور همزمان هم اثرات مخرب را بهبود دهد و هم ابعاد تصویر را بزرگ کند. برای این منظور از روش‌های فرا تفکیک‌پذیری در مرحله بازسازی استفاده شده‌است.

۲-۱ فشرده‌سازی تصاویر متنی

در ادامه، برخی روش‌های تخصصی یا نیمه‌تخصصی فشرده‌سازی تصاویر متنی بررسی می‌شوند. این بررسی در دو زیربخش روش‌های نسبتاً عمومی و روش‌های مبتنی بر مدل محتوای ترکیبی انجام می‌شود.

۲-۱-۱ روش‌های نسبتا عمومی

روش‌های فشرده‌سازی مبتنی بر تبدیل موجک جزو روش‌های امروزی و موثر محسوب می‌شوند **[۴]**. امروزه مشخص شده است که ترکیب تبدیل موجک و کدگذارهای بخش‌بندی مجموعه[1] (SPC) **[۲۳، ۲۴]** در حالت کلی، در بیشتر کاربردها کارایی بیشتری نسبت به دیگر انواع تبدیل‌ها (مانند DCT) و کدگذاری‌ها (مانند هافمن و حسابی وفقی) دارد **[۲۳، ۲۴، ۴]**. کدگذارهای SPC عموماً از ویژگی وجود شباهت در

[1] Set Partition Coding (SPC)

زیرباندهای هم‌راستا در تجزیه‌ی هرمی موجک بهره می‌برند. عملکرد این روش‌ها عمدتاً مبتنی بر روشی قانون‌مند برای جستجوی مهم‌ترین ضرایب موجک و تصمیم‌گیری در مورد کدگذاری مناسب آنها است طوری که واحد کدگشا، با در دست داشتن نتایج این جستجو و تصمیم‌گیری، قادر به بازسازی ماتریس ضرایب موجک و سپس، اعمال عکس تبدیل موجک باشد **[۲۵]**. کدگذارهای SPC روی ماتریس حاصل از تجزیه‌ی موجک کار کرده و قادر به اولویت‌بندی کدگذاری ضرایب برحسب میزان تاثیر آنها در تصویر بازسازی شده می‌باشند؛ هر ضریبی که اندازه‌ی بزرگتری داشته باشد، زودتر کدگذاری می‌شود. بنابراین، در این نوع از کدگذاری با نرخ بیت بسیار کم، ضرایبی که اندازه‌ی بزرگتری دارند شانس بیشتری برای کد شدن دارند. این دسته از کدگذارها دارای چندین ویژگی مطلوب می‌باشند. اول این که رشته‌ی بیتی خروجی به گونه‌ای تهیه و ارسال می‌شود که در هر زمان که رشته‌ی بیتی دریافتیِ واحد کدگشا به هر دلیلی مانند محدود و ثابت بودن پهنای باند خط انتقال یا وجود نویز قطع شود، با بیت‌های دریافتی تا آن لحظه، بتوان تقریبی از تصویر اولیه به دست آورد. دوم، از آن جا که این کدگذارها ابتدا سعی در ارسال مهم‌ترین (یا بزرگترین) ضرایب دارند، بنابراین، در هر لحظه‌ای که به دلایل مختلف، رشته‌ی بیتی خروجی کدگذار (یا رشته‌ی بیتیِ دریافتی کدگشا) قطع شود، مطمئن هستیم که مهمترین اطلاعات قابل ارسال تا آن لحظه، ارسال شده است و بنابراین، بالاترین حد PSNR ممکن با توجه به نرخ بیت گیرنده، حاصل گردیده است. حُسن سومِ این دسته از کدگذارها، سادگی نسبی پیاده‌سازی آنها در مقایسه با کدگذارهای دیگر مانند روش‌های مبتنی بر مدل MRC است. حسنِ چهارم، قابلیت کدگذاری جاسازی شده‌ی است؛ به این معنا که اگر به آنها همچنان اجازه‌ی کار داده شود، کیفیت‌های بهتر و بهتری از تصویر اولیه می‌توان به دست آورد. به عبارت دیگر، اگر چند کاربر مایل به دریافت تقریب‌های مختلفی از تصویر اولیه با کیفیت‌های مختلف باشند، رشته‌ی بیتی آنها زیرمجموعه‌ای از هم بوده و لازم نیست برای هرکدام رشته‌ی بیتی خروجی مجدداً از ابتدا تهیه و ارسال شود. پنجم، از آن جا که در تصاویر متنی، درجه‌ی تفکیک مقادیر شدت روشنایی

در مقایسه با درجه‌ی تفکیک مکانی از اهمیت کمتری برخوردار است؛ بنابراین، کدگذارهای SPC برای کدگذاری این نوع از تصاویر با نرخ بیت پایین مناسب می‌باشند زیرا درجه‌ی تفکیک مکانی را حفظ کرده و بهترین درجه‌ی تفکیک ممکن برای مقادیر پیکسل‌ها را – در یک نرخ بیت ثابت- فراهم می‌کنند. معمولاً نتایج این جستجو به کمک کدگذارهایی مانند کدگذار حسابی، کدگذاری شده تا دنباله‌ی بیتی خروجی تولید شود.

از جمله مهم‌ترین روش‌های کدگذاری SPC، روش کدگذاری SPIHT است **[۲۶]**. از این کدگذار تاکنون در فشرده‌سازی تصاویر مختلفی مانند تصاویر هوایی، تصاویر پزشکی، و تصاویر عمومی و نیز برای سیگنال-هایی مانند سیگنال‌های قلبی و مغزی استفاده شده است **[۲۴]**.

روش‌هایی که در دسته‌ی فعلی مورد بررسی قرار دارند ممکن است استانداردهای موجود را برای فشرده‌سازی تصاویر متنی به کار گرفته و عملکرد آن‌ها را بهبود داده باشند. برای مثال در **[۲۷]** یک روش قطعه‌بندی بلوکی تصاویر متنی به منظور فشرده‌سازی این نوع از تصاویر معرفی شده است. در این روش از استاندارد JPEG برای کدگذاری استفاده شده است. روش پیشنهاد شده در **[۲۸]** به دنبال کاهش یا حذف اثرات نامطلوب ایجاد شده از اعمال استاندارد JPEG که ویژه‌ی فشرده‌سازی تصاویر طبیعی است، روی تصاویر متنی می‌باشد. استانداردهایی که تاکنون در کاربرد فشرده‌سازی تصاویر متنی استفاده شده‌اند، محدود به استانداردهای فشرده‌سازی تصاویر نمی‌شوند. برای نمونه، در [۲۹] از کدک ویدیویی H.264/AVC برای فشرده‌سازی تصاویر متنی استفاده شده است.

در تصاویر متنی، کیفیت لبه‌های بازیابی شده در تصاویر بازسازی شده (پس از فشرده‌سازی) مهم است. روش‌های عمومی فشرده‌سازی تصاویر و نیز برخی روش‌های تخصصی فشرده‌سازی تصاویر متنی به این نکته توجهی نشان نداده‌اند. برای مثال در فشرده‌سازی به کمک استاندارد JPEG، اثر بلوکی و حذف فرکانس‌های میانی به بالا مشاهده می‌شود. لبه‌های تصاویر متنی عمدتاً متناظر با فرکانس‌های میانی می‌باشند. اندازه‌ی

این مولفه‌ها گرچه ممکن است در مقایسه با اندازه‌ی مولفه‌های فرکانس پایین نسبتاً کوچک باشد، اما اهمیت آنها در بازسازی کیفیت لبه قابل توجه است. در خروجیِ روش JPEG که مبتنی بر تبدیل DCT است، به دلیل بلوک‌بندی تصویر، اثر بلوکی شدن مشاهده می‌شود. همچنین این روش به دلیل کوانتیزاسیون انجام شده در آن، عملکردی معادل با یک فیلتر پایین‌گذر دارد [۱]؛ در واقع، در تبدیل DCT عموماً ضرایب فرکانس میانی و بالا در مقایسه با ضرایب فرکانس پایین، کوچک هستند و در اثر تقسیم بر ضرایب ماتریس کوانتیزاسیون و گرد شدن، بسیار کوچک یا صفر می‌شوند. بنابراین، حذف مولفه‌های فرکانس میانی باعث اُفت کیفیت لبه‌ها می‌شود.

در استاندارد JPEG2000 از تبدیل موجک و کدگذار EBCOT که نوعی کدگذار بخش‌بندی مجموعه است [۳۰]، استفاده شده است. موجک‌ها از ویژگی حضور محلی و نیز ویژگی متراکم‌سازی انرژی برخوردار هستند؛ بنابراین در JPEG2000، میزان فشرده‌سازی بیشتر شده و کیفیت لبه‌ها نیز بهتر حفظ می‌شود؛ اما متأسفانه به دلیل وقوع اثر نشتی در نرخ بیت‌های پایین، کیفیت لبه‌ها و نیز کیفیت دیداری تصویر، اُفت پیدا می‌کنند [۱].

۲-۱-۲ روش‌های تخصصی مبتنی بر مدل محتوای ترکیبی

در روش‌های تخصصی‌تر فشرده‌سازی تصاویر متنی، از مدل محتوای ترکیبی (MRC) استفاده می‌شود [۳۱-۳۳]. در این مدل، لایه‌های مختلفی برای تصویر متنی تعریف و از یکدیگر جدا شده و هرکدام از این لایه‌ها، جداگانه کدگذاری می‌شوند. در بیشتر روش‌های مبتنی بر این مدل، سه لایه‌ی پس‌زمینه، پیش‌زمینه، و ماسک (پوشش) تعریف می‌شوند. لایه‌ی پیش‌زمینه شامل رنگ یا شدت روشنایی نواحی متنی است. لایه‌ی ماسک یک تصویر دودویی بوده و نواحی متنی را از نواحی غیرمتنی جدا می‌کند. لایه‌ی پس‌زمینه نیز شامل اطلاعات نواحی غیرمتنی مانند گرافیک، شکل، و رنگ پس‌زمینه‌ی تصویر است.

عمده روش‌های مبتنی بر این مدل روی تصاویر متنی چاپی کار کرده و تعداد بسیار کمی تصاویر دستنویس را در نظر گرفته‌اند. برای نمونه، در **[۳۴]** روشی برای فشرده‌سازی تصاویر متنی دستنویس هندی مبتنی بر جداکردن لایه‌های پیش زمینه و پس‌زمینه، کدگذاری دوره تداوم، و استفاده از استاندارد JPEG پیشنهاد شده است.

معمولاً برای فشرده‌سازی لایه‌های پیش‌زمینه و پس‌زمینه از روش‌های استاندارد مانند JPEG و JPEG2000 و برای فشرده‌سازی لایه‌ی ماسک از روش‌های استانداردی مانند JBIG2 استفاده می‌شود. گرچه از روش‌های غیراستاندارد و جدید نیز برای فشرده‌سازی این لایه‌ها استفاده می‌شود.

با استفاده از این مدل می‌توان به میزان فشرده‌سازی بالا و در عین حال کیفیت مناسبی برای تصاویر فشرده شده دست یافت **[۳۷-۳۵]**. این مدل جزو استاندارد ITU (به نام T.44) درآمده و حتی در استاندارد JPEG2000 در بخش تصاویر متنی مرکب نیز استفاده شده است. یکی دیگر از محاسن این مدل، برخورداری از سادگی و قانون‌مندی است؛ اما در مقابل، با برخی مشکلات از جمله مشکل سربار کدگذاری نیز مواجه است **[۳۶، ۳۵]**.

به این ترتیب، انتظار می‌رود که این مدل چند ویژگی مطلوب داشته باشد: اول، کدگذارهای مورد استفاده در این مدل مستقل از هم هستند. دوم، امکان استفاده از استانداردهای نسبتاً عمومیِ موجود فراهم می‌شود. سوم، بین نواحی متنی و غیرمتنی، تمایز در نظر گرفته شود؛ چهارم، می‌توان هر لایه را، بسته به میزان اهمیت آن، با نرخ بیت متفاوت و مشخصی کدگذاری نمود؛ و پنجم، برای کدگذاری هر لایه می‌توان با استفاده از فرضیات معقول و مختص به همان لایه، به کارایی فشرده‌سازی بالاتری در مقایسه با روش‌های عمومی فشرده‌سازی تصاویر دست یافت. برای مثال، برخی روش‌ها برای کدگذاری لایه‌ی پس‌زمینه از رنگ نماینده برای کدگذاری قسمت قابل توجهی از لایه‌ی پس‌زمینه استفاده می‌کنند **[۳۸]**. در مورد لایه‌ی پیش‌زمینه نیز تا حدی به طور مشابه می‌توان عمل کرد. در مورد لایه‌ی ماسک نیز از ایده‌هایی مانند انطباق

الگو [۳۹] و استفاده از ویژگی‌های رسم‌الخط [۴۱-۴۰] می‌توان استفاده کرد. در [۴۲] روشی مبتنی بر مدل محتوای ترکیبی ارائه شده که تصویر ورودی را به چهار لایه تجزیه کرده و هر لایه را به شیوه‌ی جداگانه‌ای کدگذاری می‌کند. تشکیل لایه‌ها وابسته به نوع کلمات در زبان عربی و جامد و مرکب بودن آن‌ها است.

روش‌های زیادی مبتنی بر این مدل تاکنون ارائه شده‌اند که از بین آن‌ها برخی مانند DjVu [۴۳]، DigiPaper [۴۴] و LuraDocument [۴۶-۴۵] تجاری‌سازی شده‌اند. از بین این روش‌ها می‌توان DjVu را یکی از بهترین و کارآمدترین روش‌های مبتنی بر مدل MRC دانست زیرا بسیاری از روش‌ها کارایی خود را با آن مقایسه کرده‌اند [۴۸-۴۷، ۳۵]. اما با این حال، کارایی این نوع از روش‌ها از چند جهت چندان مناسب نمی‌باشد: اول، لبه‌ها معمولاً متناظر با گذر نسبتاً آرام هستند نه ناگهانی، بنابراین مدل فوق نمی‌تواند کیفیت این گذر را حفظ کند. دوم، کارایی این مدل تا حد زیادی وابسته به روش مورد استفاده جهت جداسازی لایه‌ها یا همان قطعه‌بندی تصویر متنی است. سوم، حجم محاسباتی این روش‌ها در حالت کلی، نسبتاً بالا است. چهارم، نرخ بیت خروجی در این روش‌ها در حالت کلی، قابل کنترل نبوده و بنابراین مناسب برای انتقال با نرخ بیت ثابت روی شبکه‌های انتقال داده نمی‌باشند. پنجم، کارایی و عملکرد این روش‌ها در برخی موارد وابسته به چاپی و یا دستنویس بودن نوع تصویر متنی و حتی نوع زبان متن می‌باشد [۴۹، ۱]. ششم این که، در برخی روش‌های متعلق به این دسته، اثر کنگره‌ای یا ناهمواری مرزهای علائم و نویسه‌های متنی ایجاد می‌شود که موجب اُفت کیفیت دیداری تصویر بازسازی شده می‌شود. هفتم این که در برخی از روش‌ها از کدگذار/کدگشای مبتنی بر این مدل، از تبدیل موجک استفاده می‌شود که این امر موجب نشت رنگ بین دو ناحیه رنگی یکنواخت و مجاور می‌شود [۵۰]. هشتم این که دقت روش‌های قطعه‌بندی مورد استفاده در این دسته از روش‌ها گاهاً به حدی نیست که بتواند اشیاء متنی ظریف و کوچک را از اشیاء غیر متنی جدا کند. بنابراین، معمولاً برخی اجزاء متنی کوچک اما در عین حال مهم، مانند نقطه‌ها و سرکش حروف، به

درستی به عنوان شیء متنی شناسایی نشده و به عنوان یک شیء غیرمتنی با آن‌ها برخورد می‌شود. این امر موجب اُفت کیفیت و حتی اُفت خوانایی تصویر بازسازی شده می‌شود.

قطعه‌بندی یکی از مهم‌ترین مسائل مطرح در روش‌های مبتنی بر مدل MRC است. یک تفاوت عمده‌ی بین این روش‌ها در نوع روش قطعه‌بندی مورد استفاده است. قطعه‌بندی تصاویر متنی به نحوی معادل با دودویی کردن آن از طریق آستانه‌یابی است. برخی روش‌های آستانه‌یابی مورد استفاده در این کار عبارتند از الگوریتم k-means **[۴۳]**، ترکیب خوشه‌یابی و آستانه‌یابی وفقی **[۴۵-۴۴]**، آستانه‌یابی بلوکی مبتنی بر بهینه‌سازی نرخ بیت-اعوجاج **[۵۲-۵۱]**، طرح پایین به بالا به منظور کمینه کردن واریانس لایه‌های پیش زمینه و پس زمینه **[۵۳]**، و مدل استخراج نواحی چندگانه با رنگ ثابت یا MECCA **[۳۷]**.

یکی دیگر از وجوه تمایز بین روش‌های مبتنی بر مدل MRC نحوه‌ی کدگذاری لایه دودویی ماسک است. این لایه مانند یک تصویر متنی دودویی بوده و برای کدگذاری آن از روش‌های فشرده‌سازی تصاویر متنی دودویی استفاده می‌شود.

۲-۲ فرا تفکیک‌پذیری

فرا تفکیک‌پذیری را می‌توان براساس معیارهای مختلفی طبقه‌بندی کرد. این معیارها شامل حوزه کاربردی، تعداد تصاویر در دسترس با تفکیک پذیری پایین و روش بازسازی حقیقی است. اگر براساس معیار تعداد تصاویر با وضوح پایین روش‌ها را طبقه‌بندی کنیم، می‌توان دو گروه شامل روش‌های مبتنی بر چند تصویر و روش‌های مبتنی بر تک تصویر را داشت.

فراتفکیک‌پذیری تک تصویر در دو دهه اخیر توجه بسیاری از محققان را به خود جلب کرده است. الگوریتم‌های فرا تفکیکپذیری تک تصویر در حوزه‌ی تصاویر متنی را می‌توان به دو دسته تقسیم کرد: ۱) روش‌های مبتنی بر درون‌یابی ۲) روش‌های مبتنی بر یادگیری.

۲-۳ روش‌های مبتنی بر درون‌یابی

در این روش‌ها در واقع مقدار میانی به وسیله‌ی مقادیر همسایه‌ها تخمین زده می‌شود. در پردازش تصویر، درون‌یابی یک تصویر به این معنی است که یک سیگنال پیوسته به‌وسیله‌ی مجموعه‌ای از داده‌های تصاویر گسسته تخمین زده شود. روش‌های مبتنی بر درون‌یابی برای تخمین تصویر وضوح بالا از چندجمله-ای‌ها استفاده می‌کنند. در ابتدا تصویر ورودی را بزرگ می‌کنند سپس مقادیر نمونه‌های جدید را با درون‌یابی نمونه‌هایی که در همسایگی قرارگرفته‌اند به دست می‌آورند. این روش‌ها، روش‌های ساده و از لحاظ محاسباتی مؤثری هستند و در آن‌ها مدل تصویر برداری ساده در نظر گرفته می‌شود [۱۶,۵۴].

در واقع در این روش‌ها به نوعی ترکیبی از الگوریتم‌های درون‌یابی و بهبود تصاویر را داریم. یعنی یا تصویر را با روش‌های معمول درون‌یابی مانند دومکعبی بزرگ می‌کنند و سپس تصویر بزرگ شده را با تأکید بر لبه‌های متن بهبود می‌دهند و یا برعکس ابتدا تصویر را بهبود می‌دهند و سپس بزرگ می‌کنند. اخیراً از روش‌های درون‌یابی مبتنی بر ویژگی‌های آماری لبه به‌منظور فرا تفکیک‌پذیری مبتنی بر دورن‌یابی استفاده شده است. اما این روش‌ها، روش‌هایی زمان بر هستند [۵۵]. مرجع [۵۶] روش افزایش ابعاد ساده اما مؤثری برای بهبود خودکار وضوح تصویر/ویدئو، درحالی‌که اطلاعات ساختاری اساسی را حفظ می‌کند پیشنهاد داده است. مزیت اصلی روش پیشنهاد شده چارچوب بازخورد[1] کنترلی است که به‌خوبی اطلاعات تصویر وضوح بالا را بدون در نظر گرفتن هیچ آموزشی از داده ورودی بازیابی می‌کند. این باعث می‌شود که روش پیشنهادی هیچ وابستگی به کیفیت و تعداد نمونه‌های انتخاب شده که مسئله رایج در روش‌های مبتنی بر آموزش است، نداشته باشد. مزیت دیگر روش پیشنهاد شده سرعت بالای آن می‌باشد. در [۵۷] یک الگوریتم جدید زمان واقعی لایه‌بندی[2] مبتنی بر مدل خطی محلی به‌منظور تجزیه تصویر متنی ورودی به لایه‌های مختلف آن طراحی شده است. سپس برای افزایش ابعاد هر لایه یک روش بزرگ‌نمایی خاص در

[1] Feedback
[2] Matting

نظر گرفته شده است. در نهایت تصویر متنی با وضوح بالای نهایی حاصل از فرا تفکیک‌پذیری با ترکیب لایه‌های وضوح بالای ایجاد شده، بدست می‌آید.

۲-۴ روش‌های مبتنی بر یادگیری

در روش‌های مبتنی بر یادگیری، رابطه‌ی میان وصله‌های وضوح بالا و وصله‌های وضوح پایین با استفاده از یک پایگاه داده خارجی یاد گرفته می‌شود. سپس این رابطه به وصله‌های وضوح پایین تصویر ورودی اعمال می‌شود و یک تصویر وضوح بالا تولید می‌شود. این مسئله همانند حل یک دستگاه معادلات است به‌طوری‌که تعداد مجهولات از تعداد معادلات بیشتر باشد. در این حالت بینهایت جواب برای دستگاه معادله وجود خواهد داشت و همین موضوع مسئله فرا تفکیک‌پذیری را بدحالت می‌کند. برای حل این مشکل به یک سری فرضیات و اطلاعات پیشین نیاز است. روش‌های متفاوتی برای یادگیری رابطه‌ی میان وصله‌های وضوح بالا و پایین ارائه شده‌اند. موفقیت روش‌های فرا تفکیک‌پذیری مبتنی بر مثال به دو فاکتور مهم وابسته است: ۱) جمع‌آوری یک پایگاه داده مناسب شامل جفت وصله‌های وضوح بالا و پایین ۲) یادگرفتن رابطه بین وصله-های وضوح بالا و پایین.

الگوریتم‌های مبتنی بر یادگیری ابتدا در [۵۸] معرفی شدند. در این مقاله از شبکه عصبی برای بهبود وضوح تصاویر اثر انگشت استفاده شده است. این الگوریتم‌ها دارای یک فاز آموزش هستند که در آن‌ها رابطه‌ی میان تعدادی تصاویر وضوح بالا و وضوح پایین از یک کلاس ویژه شامل تصاویر چهره، اثر انگشت و غیره یاد گرفته می‌شود. سپس این رابطه به‌عنوان ترم پیشین برای بازسازی تصویر وضوح بالا به‌کار برده می‌شود. پایگاه آموزشی الگوریتم‌های فرا تفکیک‌پذیری مبتنی بر پایگاه داده باید قابلیت عمومیت‌گرایی خوبی داشته باشند [۵۹]. پایگاه داده خارجی بزرگ لزوماً نتایج بهتری تولید نمی‌کند در مقابل زیاد بودن نمونه‌های نامرتبط نه تنها بار محاسباتیِ پیدا کردن بهترین نگاشت را زیاد می‌کنند، بلکه باعث می‌شوند جواب خوبی نیز بدست نیاید. پس در این روش‌ها انتخاب پایگاه داده‌ای شامل تصاویر مرتبط با تصویر مورد

مطالعه موضوع مهمی است که باید در نظر گرفته شود. به‌طور مثال استفاده از پایگاه داده‌ای شامل تصاویر طبیعی به‌عنوان مجموعه‌ی آموزشی جواب خوبی برای بازسازی تصاویر متنی نخواهد داشت. موضوع مورد توجه دیگر، وضوح تصاویر استفاده شده در مجموعه‌ی آموزشی و هم‌چنین ویژگی‌های مختلف در نظر گرفته شده (به‌طور مثال در تصاویر متنی، اندازه کاراکترها، نوع فونت استفاده شده و ....) در این تصاویر می‌باشد. به‌عبارت دیگر مجموعه آموزشی باید شامل تصاویر وضوح بالا و پایین باشد تا آموزش رابطه بین وصله‌های وضوح بالا و پایین به خوبی یاد گرفته شود. از نظر ویژگی‌ها نیز باید ویژگی‌های مختلف را شامل شود تا بتواند برای تصاویر با ویژگی‌های متفاوت جواب خوبی داشته باشد (عمومیت گرایی بالایی داشته باشد). مشخص است که داشتن چنین مجموعه‌ای و آموزش آن فرآیندی زمان‌بر است. بخصوص اگر تصاویر مورد مطالعه تصاویری با درجه تفکیک مکانی بالا باشند. در ادامه مواردی از کارهای انجام شده در زمینه‌ی روش‌های فرا تفکیک‌پذیری مبتنی بر یادگیری در حوزه تصاویر متنی توضیح داده شده است.

مرجع [۶۰] یک روش فرا تفکیک‌پذیری مبتنی بر آموزش برای تصاویر متنی را معرفی می‌کند. این مقاله شامل دو فاز است: ۱) فاز آموزش ۲) فاز بازسازی. در فاز آموزش از چندین تصویر وضوح بالا که به‌عنوان نمونه‌های وضوح بالا در نظر گرفته می‌شوند برای بدست آوردن کاراکترهای با کیفیت بالا جهت آموزش استفاده می‌کند. سپس وصله‌های وضوح بالا را از هر تصویر انتخاب می‌کند. تنها وصله‌هایی انتخاب می‌شوند که در طول لبه‌های کاراکتر قرار دارند زیرا شکل لبه‌ها یک ویژگی بسیار مهم برای بیان کاراکترها است. وصله‌های وضوح پایین را با مات کردن و کاهش ابعاد وصله‌های وضوح بالا بدست می‌آورد. پایگاه داده شامل زوج وصله‌های وضوح بالا و پایین تولید شده است. پایگاه داده شامل الگوهای زیادی است که نسبت به شکل‌ها، اندازه، چرخش و موقعیت در وصله‌های تصویر متفاوت است. یک واژه‌نامه[1] وضوح بالای آموزش دیده از تمام این داده‌ها حتی اگر تعداد زیادی از اتم‌ها وجود داشته باشد، تضمین نمی‌کند که برای بازسازی

[1] Dictionary

تصویر وضوح بالا مناسب باشد. علاه‌بر این، پیچیدگی محاسباتی زیادی خواهد داشت. برای غلبه بر این مشکلات و بدست آوردن سود کامل از پایگاه داده آموزشی بزرگ، پیشنهاد می‌کنند که از آموزش چندین واژه‌نامه از پایگاه داده خوشه‌بندی شده استفاده شود. اما اطلاعاتی در رابطه با پایگاه داده به‌منظور راهنمای خوشه‌بندی مانند تعداد کلاس‌ها نداریم. بنابراین، از نسخه هوشمند K-میانگین[1] به نام iK-میانگین به‌منظور گردآوری الگوهای مشابه در خوشه‌های یکسان استفاده شده است. این روش خوشه‌بندی تعداد خوشه‌ها و مراکز اولیه برای روش K-میانگین را مشخص می‌کند. با دادن وصله‌های وضوح بالا و وصله‌های وضوح پایین متناسب با آن از هر خوشه، دو واژه‌نامه تزویج شده وضوح بالا و وضوح پایین با استفاده از روش Joint Sparse Coding آموزش دیده شده است. هدف از این روش یادگیری که براساس الگوریتم کدگذاری تنک[2] است، داشتن نمایش تنک یکسان برای هر جفت وصله وضوح بالا و پایین است. پس از آموزش واژه‌نامه نوبت به فاز بازسازی می‌رسد. نقطه شروع فاز بازسازی، تصویر وضوح پایین ورودی است. ابتدا تصویر وضوح پایین ورودی را با روش درون‌یابی دومکعبی ابعادش را افزایش می‌دهیم. سپس با استفاده از روش‌های کدگذاری تنک و با داشتن واژه‌نامه وضوح پایین آموزش دیده و وصله‌های وضوح پایین، ضرایب تنک را برای این وصله‌ها بدست می‌آورد. از آنجا که این ضرایب برای وصله‌های وضوح بالا برابر با وصله‌های وضوح پایین است، با ضرب این ضرایب بدست آمده در واژه‌نامه وضوح بالا، وصله وضوح بالا را بدست می‌آورد.

مرجع [۶۱] بیان می‌کند که برای جلوگیری از حجم زیاد محاسبات از تمام وصله‌ها در مرحله آموزش استفاده نکنیم بلکه یک معیار انتخاب برای وصله‌ها قرار دهیم. معیار انتخاب شده در این مقاله واریانس وصله‌ها می‌باشد. وصله‌هایی که واریانس آنها از آستانه از پیش مشخص شده بیشتر باشد، برای بازسازی در حوزه تنک انتخاب می‌شوند و وصله‌های باقی مانده با استفاده از روش سریع و راحت دومکعبی درون‌یابی می‌شوند. وصله‌های انتخاب شده، وصله‌هایی هستند که شامل ویژگی‌های لبه در تصاویر متنی هستند و در

[1] K-Means
[2] Sparse Coding

تصاویر متنی این ویژگی‌ها بسیار پر اهمیت هستند. وصله‌هایی که با درون‌یابی ساده بزرگ‌نمایی می‌شوند شامل نواحی هموار می‌باشند که ویژگی‌های مهم و قابل توجهی را در تصاویر متنی شامل نمی‌شوند.

روشی مبتنی بر کدگذاری تنک در [۶۲] معرفی می‌شود. روش‌های تنک بیان می‌کنند که وصله‌های تصویر می‌توانند به‌خوبی به‌صورت ترکیب خطی تنک از المان‌های مناسب واژه‌نامه آموزش داده شده نمایش داده شوند. یک روش فرا تفکیک‌پذیری قوی برای بهبود تصاویر متنی گرفته شده از دوربین با کیفیت پایین در [۶۳] ارائه شده است.

فصل سوم:

مبانی نظری

در این فصل به معرفی تئوری روش ارائه شده پرداخته می‌شود. از آنجاییکه از دو روش فرافکیک-پذیری مبتنی بر درون‌یابی و مبتنی بر آموزش استفاده شده‌است ابتدا مبانی نظری روش مبتنی بر درون‌یابی و سپس روش مبتنی بر آموزش بیان می‌شود.

قبل از اینکه مبانی نظری مربوط به روش‌های فرا تفکیک‌پذیری مطرح شوند ابتدا دو موضوع که در حوزه فرا تفکیک‌پذیری آشنایی با آنها مفید است معرفی می‌شوند.

۳-۱ وضوح تصویر

وضوح نوری[1] معیاری برای توانایی سیستم تصویربرداری یا جزئی از سیستم تصویربرداری است که جزئیات تصویر را نشان می‌دهد. به عبارت دیگر وضوح تصویر کوچک‌ترین جزئیاتی است که در یک تصویر قابل مشاهده است. وضوح یک موضوع بنیادی در قضاوت کیفیت سیستم پردازش یا وسیله به دست آوردن تصویر می‌باشد. در ساده‌ترین حالت وضوح تصویر به‌عنوان کوچک‌ترین جزئیات قابل اندازه‌گیری در نمایش بصری تعریف می‌شود. در حوزه پردازش تصاویر دیجیتال و ویدئو مفهوم وضوح به یکی از سه تعریف ذیل مربوط می‌شود [۶۴]:

وضوح فضایی[2]: هر تصویر از کوچک‌ترین عنصر به نام پیکسل ساخته شده است. وضوح فضایی به تعداد پیکسل‌ها در یک تصویر مربوط می‌شود و با معیار پیکسل بر واحد سطح اندازه‌گیری می‌شود. هرچه وضوح فضایی یک پیکسل بالاتر باشد تعداد پیکسل‌ها در تصویر بیشتر خواهد بود. وضوح فضایی بالاتر درک بالاتری از جزئیات در یک تصویر را به کاربر می‌دهد.

وضوح روشنایی[3]: به آن وضوح سطح خاکستری نیز گفته می‌شود. وضوح روشنایی به تعداد سطوح روشنایی یا سطوح خاکستری برای نمایش یک پیکسل گفته می‌شود. وضوح روشنایی با تعداد سطوح کوانتیزاسیون

[1] Optical-Resolution
[2] Spatial Resolution
[3] Intensity Resolution

افزایش پیدا می‌کند. یک تصویر خاکستری معمولاً به ۲۵۶ سطح کوانتیزه می‌شود و هر سطح با ۸ بیت بیان می‌شود. برای یک تصویر رنگی هر کانال رنگ با ۸ بیت نمایش داده می‌شود که ۲۴ بیت برای ارائه هر سطح رنگ نیاز است. این قابل بیان است که تعداد سطوح کوانتیزاسیون به نرخ نمونه‌برداری فضایی مربوط می‌شود. اگر حسگر دوربین سطوح کوانتیزاسیون کمتری داشته باشد باید نمونه‌برداری با نرخ بالاتری انجام شود تا بتواند روشنایی صحنه را تسخیر کند.

وضوح زمانی[1]: این وضوح نرخ فریم‌ها یا تعداد فریم‌های یک رشته ویدئویی را در ثانیه نشان می‌دهد. نرخ فریم‌های معمولی برای یک ویدیو در حدود ۲۵ فریم در ثانیه می‌باشد [۶۴].

در این پایان نامه واژه افزایش وضوح به وضوح فضایی مربوط می‌شود. همان‌طور که قبلاً نیز بیان شد وضوح فضایی به طور ویژه تعداد کل پیکسل‌ها در یک تصویر را نشان می‌دهد. به‌عنوان مثال برای یک تصویر دیجیتال ۳۰۰× ۳۰۰ تعداد کل پیکسل‌ها ۹۰۰۰۰ می‌باشد که در حدود ۰/۱ مگا پیکسل است. واضح است که هر چه وضوح فضایی بالاتر باشد جزئیات بیشتری در تصویر قابل نمایش است. این مسئله در رابطه با تصاویر متنی موضوعی مهم است. چرا که خوانایی انسانی و ماشینی تصاویر متنی، ارتباط مستقیمی با وضوح این تصاویر دارد. با توجه به اثرات روش‌های فشرده‌سازی و کاهش ابعاد روی تصاویر متنی که سبب می‌شوند وضوح تصویر پایین بیاید و خوانایی انسانی/ماشینی آن کم شود باید روشی انتخاب شود که بتواند اثرات مخرب گفته شده را بهبود دهد. یکی از روش‌های مهم این حوزه روش فرا تفکیک‌پذیری می‌باشد.

مسئله فرا تفکیک‌پذیری را نباید با فن‌های مشابه همانند درون‌یابی[2] و بازیابی[3] اشتباه گرفت. در مقایسه با روش‌های بهبود وضوح، فن فرا تفکیک‌پذیری نه‌تنها کیفیت تصویر را افزایش می‌دهد بلکه باعث افزایش وضوح فضایی یک تصویر نیز می‌شود. در روش‌های درون‌یابی جزئیات فرکانس بالا برخلاف روش‌های

[1] Temporal resolution
[2] Interpolation
[3] Restoration

فرا تفکیک‌پذیری بازیابی نمی‌شوند [۶۵]. در روش‌های بازیابی ابعاد تصویر ورودی و خروجی یکسان است و تنها کیفیت تصویر خروجی افزایش پیدا کرده است. در فرا تفکیک‌پذیری علاوه بر اینکه کیفیت تصویر خروجی افزایش پیدا می‌کند ابعاد آن نیز نسبت به تصویر ورودی افزایش پیدا می‌کند.

۳-۲ مدل تصویر

سیستم تصویر برداری دیجیتال دارای محدودیت‌های سخت‌افزاری برای گرفتن تصویر از یک صحنه می‌باشد. اما این سیستم‌ها تنها محدودیت سخت‌افزاری ندارند و عوامل دیگری نیز روی تصویر خروجی اثر می‌گذارند؛ عواملی مانند، حرکت میان حس‌گر دوربین و شیء، لنز و اپتیک دوربین، جو زمین، کم بودن نرخ نمونه برداری و غیره. رابطه (۳-۱) نحوه ایجاد یک تصویر وضوح پایین از روی تصویر وضوح بالا را نشان می‌دهد.

$$Y = DHFX + V \qquad (۳-۱)$$

در رابطه (۳-۱) Y تصویر وضوح پایین، F شامل اطلاعات حرکت[1]، H مدل ماتی[2] تصویر، D عمل‌گر کاهش اندازه[3]، X تصویر وضوح بالا و V نویز می‌باشد. شکل (۳-۱) مدل تصویر را نشان می‌دهد.

شکل ۳-۱ مدل تصویر [۶]

[1] Motion Information
[2] Blur
[3] Down sampling

در واقع در فراتفکیک‌پذیری سعی می‌شود با پیدا کردن مدل هر یک از عوامل مخرب تأثیر گذار بر تصویر و رفع اثر مخرب هر عامل روی تصویر، تا حد امکان به تصویر با وضوح بالای اولیه نزدیک شد.

۳-۳ روش‌های آستانه‌گذاری

در روش مبتنی بر درون‌یابی ارائه شده، به تصویر دودویی از تصویر ورودی نیاز داریم. روش‌های مختلفی جهت دودویی کردن تصاویر وجود دارد. دوتا از روش‌های معروف برای دودویی کردن تصاویر متنی روش‌های اتسو[1][۶۶] و میانگین متحرک[2][۶۷] می‌باشد.

۳-۳-۱ آستانه‌گذاری اتسو

در این الگوریتم، هدف، بیشینه نمودن واریانس بین کلاسی است. این روش یک معیار رایج برای آنالیز جداسازی آماری است. ایده پایه این است که کلاس‌هایی که به خوبی آستانه‌گذاری شده‌اند باید از بقیه‌ی کلاس‌ها جدا شوند و بالعکس یک آستانه بهترین جداسازی بین کلاس‌ها را براساس شدت روشنایی آنها انجام دهد. علاوه بر بهینه بودن، یک خاصیت مهم دیگر این روش این است که از نظر محاسباتی تماماً روی هیستوگرام تصویر انجام می‌شود.

فرض کنید $\{0,1,2,...,L-1\}$ تعداد L سطح روشنایی متمایز در یک تصویر دیجیتال با ابعاد $M \times N$ را مشخص کند و n_i تعداد پیکسل‌هایی باشد که شدت روشنایی i دارند. برای کل پیکسل‌های تصویر داریم؛ $M \times N = n_0 + n_1 + n_2 + ... + n_{L-1}$. هیستوگرام نرمالیزه شده در رابطه (۳-۲) نشان داده شده است:

$$\begin{cases} P_i = \dfrac{n_i}{MN} & ,i = 0,1,...,L-1 \\ \sum_{i=0}^{L-1} P_i = 1 & ,P_i \geq 0. \end{cases} \qquad (۳-۲)$$

[1] Otsu
[2] Moving Average

با انتخاب آستانه $0<K<L-1$ می‌توان تصویر ورودی را به دو کلاس C_1 و C_2 تقسیم کرد. C_1 شامل تمام پیکسل‌های تصویر با شدت روشنایی در بازه $[0,K]$ و C_2 شامل تمام پیکسل‌های تصویر در بازه $[K+1,L-1]$ می‌باشد. با استفاده از این آستانه، احتمال $P_1(k)$ (احتمال تعلق یک پیکسل به کلاس C_1) به صورت رابطه (۳-۲) بیان می‌شود.

$$P_1(k)=\sum_{i=0}^{k}P_i \qquad (۳-۳)$$

رابطه (۳-۳) در واقع احتمال رخ دادن کلاس C_1 است. به همین ترتیب برای کلاس C_2 نیز رابطه احتمال به صورت رابطه (۳-۴) می‌باشد.

$$P_2(k)=\sum_{i=k+1}^{L-1}P_i=1-P_1(k) \qquad (۳-۴)$$

متوسط شدت روشنایی پیکسل های متعلق به هر کلاس نیز از روابط (۳-۵) و (۳-۶) محاسبه می‌شوند.

$$m_1(k)=\frac{1}{P_1(k)}\sum_{i=0}^{k}iP_i \qquad (۳-۵)$$

$$m_2(k)=\frac{1}{P_2(k)}\sum_{i=k+1}^{L-1}iP_i \qquad (۳-۶)$$

شدت روشنایی متوسط تصویر ورودی نیز از رابطه (۳-۷) محاسبه می‌شود.

$$m_G=\sum_{i=0}^{L-1}iP_i \qquad (۳-۷)$$

به منظور ارزیابی مقدار آستانه در سطح k از مقدار نرمال شده‌ی آن، از کمیت بدون بعد در رابطه (۳-۸) استفاده می‌کنیم.

$$\eta=\frac{\sigma_B^2}{\sigma_G^2} \qquad (۳-۸)$$

در رابطه (۳-۸) σ_G^2 واریانس سراری است که از رابطه (۳-۹) محاسبه می‌شود.

$$\sigma_G^2 = \sum_{i=0}^{L-1} (i - m_G)^2 P_i \qquad (۳-۹)$$

در رابطه (۳-۸) σ_B^2 واریانس بین کلاسی است که از رابطه (۳-۱۰) محاسبه می‌شود.

$$\sigma_B^2 = P_1 (m_1 - m_G)^2 + P_2 (m_2 - m_G)^2 \qquad (۳-۱۰)$$

رابطه (۳-۱۰) می‌تواند به صورت رابطه (۳-۱۱) بازنویسی شود.

$$\begin{aligned} \sigma_B^2 &= P_1 P_2 (m_1 - m_2)^2 \\ &= \frac{(m_G P_1 - m)^2}{P_1 (1 - P_1)} \end{aligned} \qquad (۳-۱۱)$$

با وارد کردن k در روابط و بازنویسی روابط می‌توان (۳-۱۲) و (۳-۱۳) را بدست آورد.

$$\eta(k) = \frac{\sigma_B^2(k)}{\sigma_G^2} \qquad (۳-۱۲)$$

$$\sigma_B^2(k) = \frac{[m_G P_1(k) - m(k)]^2}{P_1(k)[1 - P_1(k)]} \qquad (۳-۱۳)$$

آستانه بهینه مقدار k^*ای است که به ازای آن عبارت $\sigma_B^2(k)$ را بیشینه کند.

$$\sigma_B^2(k^*) = \max_{0 \le k \le L-1} \sigma_B^2(k) \qquad (۳-۱۴)$$

به‌عبارت دیگر برای یافتن k^* رابطه‌ی (۳-۱۴) را به ازای تمام مقادیر k ارزیابی کرده و مقداری که بیشینه‌ی $\sigma_B^2(k)$ را نتیجه دهد، انتخاب می‌کنیم.

در نهایت آستانه‌گذاری را براساس معیار معرفی شده در رابطه (۳-۱۵) انجام می‌دهیم:

$$g(x,y)=\begin{cases}1, & f(x,y)>k^{*}\\ 0, & f(x,y)\le k^{*}\end{cases} \qquad (۱۵-۳)$$

۳-۳-۲ آستانه‌گذاری میانگین متحرک

این آستانه‌گذاری در طول خط روبش تصویر است. این روش روشی مؤثر برای آستانه‌گذاری تصاویر متنی است. اگر z_{k+1} شدت روشنایی نقطه‌ای باشد که روبشگر در گام $k+1$ با آن مواجه شده است، میانگین شدت روشنایی در آن نقطه از رابطه (۳-۱۶) محاسبه می‌شود:

$$\begin{aligned} m(k+1) &= \frac{1}{n}\sum_{i=k+2-n}^{k+1} z_i \\ &= m(k)+\frac{1}{n}(z_{k+1}-z_{k-n}) \end{aligned} \qquad (۱۶-۳)$$

در رابطه (۳-۱۶) n تعداد نقاطی است که برای محاسبه متوسط شدت روشنایی استفاده شده است و برای اولین پیکسل داریم، $m(1)=z_1/n$. در نهایت آستانه‌گذاری با استفاده از رابطه (۳-۱۷) و جای‌گذاری آن در رابطه (۳-۱۵) بدست می‌آید.

$$T_{xy}=bm_{xy} \qquad (۱۷-۳)$$

۳-۴ فیلتر هدایت

در روش مبتنی بر درون‌یابی ارائه شده نیاز به روشی است که بتواند اطلاعات مربوط به لبه‌ی متن تصویر ورودی را بدهد. برای این منظور از فیلتر هدایت[1] استفاده شده‌است.

کلید اصلی فیلتر هدایت [۶۸] مدل خطی محلی بین تصویر راهنما I و تصویر خروجی q است. فرض می‌شود q تبدیل خطی از I در پنجره w_k به مرکز k باشد:

$$q_i=a_k I_i+b_k, \quad \forall i \in w_k \qquad (۱۸-۳)$$

[1] Guided Filter

در رابطه (۳-۱۸) a_k و b_k ضرایب خطی ثابت در w_k هستند. پنجره w_k به صورت مربعی و به ضلع r در نظر گرفته شده است. مدل خطی محلی اطمینان می‌دهد که خروجی فیلتر یعنی q دارای لبه می‌باشد تنها اگر I لبه داشته باشد، زیرا $\nabla q = a\nabla I$.

برای تعیین ضرایب خطی a_k و b_k به ورودی p فیلتر نیاز داریم. خروجی q بصورت تفریق المان‌های ناخواسته nمانند نویز از ورودی pمدل می‌شود.

$$q_i = p_i - n_i \qquad (۳-۱۹)$$

راه‌حلی که در نظر گرفته شده است، روشی است که در رابطه (۳-۱۹) اختلاف بین p و q را کمینه کند در حالیکه مدل خطی (۳-۱۸) را شامل شود. به طور خاص، تابع هزینه ارائه شده در رابطه (۳-۲۰) در پنجره w_k کمینه شده است.

$$E(a_k, b_k) = \sum_{i \in w_k} ((a_k I_i + b_k - p_i)^2 + \varepsilon a_k^2) \qquad (۳-۲۰)$$

ε پارامتر تنظیم جهت a_kهای بزرگ است.

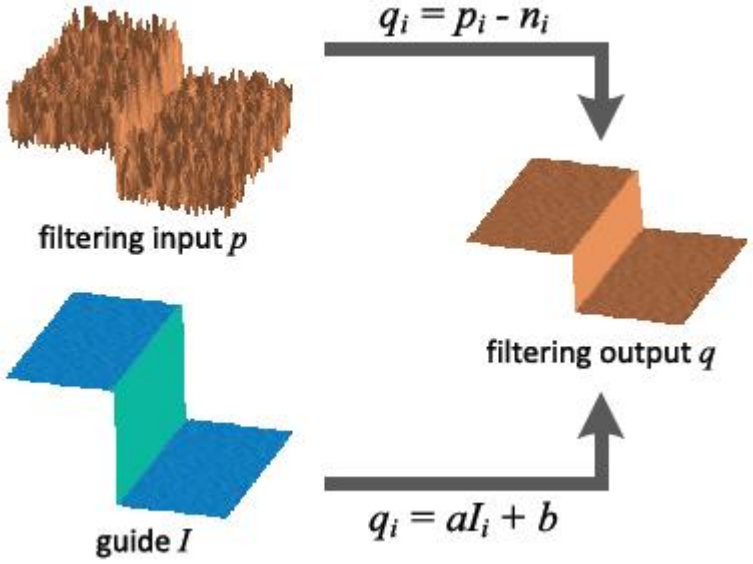
رابطه (۳-۲۱) رگرسیون لبه خطی[1] است که در [۷۰-۶۹] مدل شده است و راه‌حل آن بصورت زیر است:

$$a_k = \frac{\frac{1}{|w|}\sum_{i \in w_k} I_i p_i - \mu_k \overline{p_k}}{\sigma_k^2 + \varepsilon} \qquad (۳-۲۱)$$

$$b_k = \overline{p_k} - a_k \mu_k \qquad (۳-۲۲)$$

[1] Linear ridge regression

در روابط (۳-۲۱) و (۳-۲۲)، μ_k و σ_k^2 به ترتیب میانگین و واریانس I در w_k هستند. $|w|$ تعداد پیکسل‌ها در w_k و $\overline{p_k} = \frac{1}{|w|}\sum_{i \in w_k} p_i$ میانگین P در w_k است. با بدست آوردن ضرایب خطی a_k و b_k ما می‌توانیم خروجی q_i فیلتر را با استفاده از رابطه (۳-۱۸) محاسبه کنیم. شکل ۳-۲ فرایند فیلتر هدایت را نشان می‌دهد.



شکل ۳-۲ فرایند فیلتر هدایت [۶۸]

مشکلی که برای خروجی فیلتر هدایت وجود دارد این است که تمام پنجره‌های w_k که i را پوشش می‌دهند و با هم هم‌پوشانی دارند سبب می‌شوند خروجی q_i در رابطه (۳-۱۸) مقادیر متفاوتی داشته باشد. یک راه ساده برای حل این مشکل میانگین‌گیری از تمام مقادیر q_i ممکن می‌باشد. بنابراین پس از محاسبه a_k و b_k برای تمام پنجره‌های w_k در تصویر، خروجی فیلتر به صورت زیر محاسبه می‌شود:

$$q_i = \frac{1}{|w|}\sum_{k|i \in w_k} (a_k I_i + b_k). \qquad (۳-۲۳)$$

دقت شود که با داشتن $\sum_{k|i\in w_k} a_k = \sum_{k\in w_k} a_k$ ناشی از تقارن پنجره، می‌توان رابطه (۳-۲۳) را به صورت زیر بازنویسی کرد:

$$q_i = \overline{a_i} I_i + \overline{b_i} \qquad (۳-۲۴)$$

در رابطه (۳-۲۴) $\overline{a_i} = \frac{1}{|w|}\sum_{k\in w_i} a_k$ و $\overline{b_i} = \frac{1}{|w|}\sum_{k\in w_i} b_k$ میانگین ضرایب تمام پنجره‌های دارای هم‌پوشانی روی i هستند. استراتژی میانگین‌گیری از پنجره‌های هم‌پوشان در حذف نویز تصاویر مورد استفاده قرار می‌گیرد [۶۷].

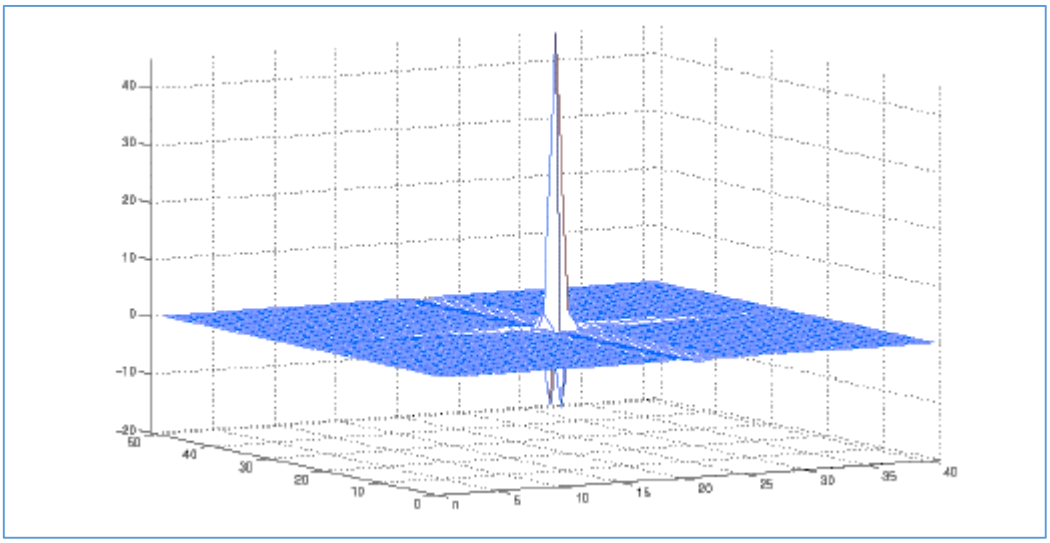
۳-۵ فیلتر تیگر

فیلتر تیگر[1] [۷۱]، فیلتری با تأکید بر تقویت لبه‌های تصویر متنی و از بین بردن نویز است. این فیلتر به منظور تقویت لبه‌های متن در روش ارائه شده مورد استفاده قرار گرفته شده است. فیلتر تیگر برای یک پیکسل تصویر y به مختصات (i, j) در رابطه (۳-۲۵) معرفی شده است:

$$\begin{aligned} y^T(i,j) = {} & 3y^2(i,j) - \frac{1}{2}y(i+1,j+1)y(i-1,j-1) \\ & - \frac{1}{2}y(i+1,j-1)y(i-1,j+1) \\ & - y(i+1,j)y(i-1,j) - y(i,j+1)y(i,j-1) \end{aligned} \qquad (۳-۲۵)$$

ساختار فیلتر تیگر در شکل ۳-۳ نشان داده شده است.

[1] Taeger Filter



شکل ۳-۳ شکل فیلتر تیگر [۷۱]

۳-۶ داده‌های آموزشی

همان‌طور که در فصل ۲ در رابطه با اهمیت پایگاه داده و تأثیر آنها در روش‌های فرا تفکیک‌پذیری مبتنی بر آموزش توضیح داده شد، در این روش‌ها وجود پایگاه داده‌ای از تصویر مرتبط با نوع تصویر مورد مطالعه لازم و ضروری است. بنابراین اولین گام در فراتفکیک‌پذیری تک تصویر مبتنی بر پایگاه داده خارجی تهیه مجموعه‌ای از تصاویر مرتبط با زمینه کاری می‌باشد. در انجام این پایان‌نامه پایگاه داده‌ای از تصاویر در درجه‌ی تفکیک مکانی ۶۰۰dpi و ۳۰۰dpi تهیه شد. تعداد ۱۰۱ تصویر برای هر درجه تفکیک مکانی ایجاد شد. در شکل‌های ۳-۴، ۳-۵، ۳-۶، ۳-۷، ۳-۸ و ۳-۹ تصاویری از پایگاه داده نشان داده شده‌است. در شکل‌های ۳-۴ تا ۳-۶ تصاویر با درجه‌ی تفکیک ۳۰۰dpi و در شکل‌های ۳-۷ تا ۳-۹ تصاویر با درجه‌ی تفکیک ۶۰۰dpi روبش[1] شده‌اند. در مشاهده اولیه به نظر می‌رسد تصاویر نشان داده شده در این شکل‌ها، تفاوتی را با هم ندارند. برای نشان دادن اختلاف بین این تصاویر، قسمتی از این تصاویر در شکل‌های ۳-۱۰ و ۳-۱۱ بزرگ‌نمایی شده‌اند.

[1] Scan

- k-means cannot guarantee convergence to the global minimum of $J(\theta, U)$ (which, hopefully, corresponds to the best possible clustering). In other words, it returns clusterings corresponding to local minima of $J(\theta, U)$. Consequently, different initializations of the algorithm may lead to different final clusterings. Care must be taken in the initialization of θ_j (see the practical hints that follow). If the initial values for, say, m_1, of θ_j lie away from the region where the data vectors lie, they may never be updated. As a consequence, k-means algorithm will proceed as if there were only $m - m_1$ θ_j's.
- Accurate estimation of the number of clusters (representatives) is crucial for the algorithm, since a poor estimate will prevent it from unraveling the clustering structure of X. More specifically, if a larger number of representatives is used, it is likely that at least one "physical" cluster will be split into two or more. On the other hand, if a smaller number of representatives is used, two or more physical clusters are likely to be represented by a single representative, which in general will lie in a sparse region (with respect to the number of data points) between those clusters.
- The algorithm is sensitive to the presence of outliers (that is, points that lie away from almost all data vectors in X) and "noisy" data vectors. Such points are the result of a noisy process unrelated to the clustering structure of X. Since both outliers and noisy points are necessarily assigned to a cluster, they influence the respective mean representatives.
- k-means is suitable for real-valued data and, in principle, should *not* be used with discrete-valued data.

Practical Hints

- Assuming that m is fixed, and to increase our chances of obtaining a reliable clustering, we may run k-means several times, each time using different initial values for the representatives, and select the best possible clustering (according to J). Three simple methods for choosing initial values for θ_j's are (a) random initialization, (b) random selection of m data vectors from X as the initial estimates of θ_j's, and (c) utilization of the clustering output of a simpler (e.g., sequential) algorithm as input.
- Two simple ways to estimate m are:
  - Use the methodology described for the BSAS algorithm.
  - For each value of m, in a suitably chosen range $[m_{min}, m_{max}]$, run the k-means algorithm n_{run} times (each time using different initial values) and determine the clustering (among the n_{run} produced) that minimizes the cost function J. Let J_m be the value of J for the latter clustering. Plot J_m versus m and search for a significant local change (it appears as a significant "knee"). If such a knee occurs, its position indicates the desired number of clusters. Otherwise, it is an indication that there is no clustering structure (that contains compact clusters) in the data set.
- Two simple ways to deal with outliers are (a) to determine the points that lie at "large" distances from most of the data vectors in X and discard them, or (b) to run k-means and identify the clusters with very few elements. An alternative is to use algorithms that are less sensitive to outliers (this is the case with the PAM algorithm, discussed later).

Example 7.5.1. Generate and plot a data set, X_3, that consists of $N = 400$ 2-dimensional points. These points form four equally sized groups. Each group contains vectors that stem from Gaussian distributions

شکل ۳-۴ تصویر ۳۰۰dpi از پایگاه داده ایجاد شده

k-Means, or Isodata, Algorithm

This is the most widely known clustering algorithm, and its rationale is very simple. In this case, the parameter vectors θ_j (also called *cluster representatives* or simply *representatives*) correspond to points in the l-dimensional space, where the vectors of data set X live. k-means assumes that the number of clusters underlying X, m, is known. Its aim is to move the points $\theta_j, j = 1,\ldots,m$, into regions that are dense in points of X (clusters).

The k-means algorithm is of iterative nature. It starts with some initial estimates $\theta_1(0),\ldots,\theta_m(0)$, for the parameter vectors $\theta_1,\ldots,\theta_m$. At each iteration t,

- the vectors x_i that lie close to each $\theta_j(t-1)$ are identified and then
- the new (updated) value of θ_j, $\theta_j(t)$, is computed as the mean of the data vectors that lie closer to $\theta_j(t-1)$.

The algorithm terminates when no changes occur in θ_j's between two successive iterations. To run the k-means algorithm, type

$$[theta, bel, J] = k\_means(X, theta\_ini)$$

where

X is an $l \times N$ matrix whose columns contain the data vectors,

theta_ini is an $l \times m$ matrix whose columns are the initial estimates of θ_j (the number of clusters, m, is implicitly defined by the size of *theta_ini*),

theta is a matrix of the same size as *theta_ini*, containing the final estimates for the θ_j's,

bel is an N-dimensional vector whose ith element contains the cluster label for the ith data vector,

J is the value of the cost function given in Eq. (7.1) (see below) for the resulting clustering.

Remarks

- k-means is suitable for unraveling compact clusters.
- The k-means algorithm is a fast iterative algorithm because (a) in practice it requires only a few iterations to converge and (b) the computations required at each iteration are not complicated. Thus, it poses as a candidate for processing large data sets.
- It can be shown that the k-means algorithm minimizes the cost function

$$J(\theta, U) = \sum_{i=1}^{N} \sum_{j=1}^{m} u_{ij} ||x_i - \theta_j||^2 \tag{7.1}$$

where $\theta = [\theta_1^T,\ldots,\theta_m^T]^T$, $||.||$ stands for the Euclidean distance, and $u_{ij} = 1$ if x_i lies closest to θ_j; 0 otherwise. In words, *k-means minimizes the sum of the squared Euclidean distances of each data vector from its closest parameter vector*. When the data vectors of X form m compact clusters (with no significant difference in size), it is expected that J is minimized when each θ_j is placed (approximately) in the center of each cluster, provided that m is known. This is not necessarily the case when (a) the data vectors do not form compact clusters, or (b) their sizes differ significantly, or (c) the number of clusters, m, has not been estimated correctly.

شکل ۳-۵ تصویر ۳۰۰dpi از پایگاه داده ایجاد شده

Remarks

- In its original form, BSAS is suitable for unraveling compact clusters (i.e., clusters whose points are aggregated around a specific point in the data space).
- The algorithm is sensitive to the order of data presentation and the choice of the parameter Θ. If the data are presented in a different order and/or the parameter Θ is given a different value, a different clustering may result.
- BSAS is *fast*, because it requires a single pass on the data set, X; thus, it is a good candidate for processing large data sets. However, in several cases the resulting clustering may be of low quality. Improvements can be achieved in a refinement step, as discussed in the following subsection.
- Sometimes, the clustering returned by BSAS is used as a starting point for other more sophisticated clustering algorithms, to generate clusterings of improved quality.
- BSAS results in a rough estimate of the number of clusters underlying the data set at hand.

To more accurately estimate the number of clusters in X, BSAS may be run for different values of Θ in a range $[\Theta_{min}, \Theta_{max}]$. For each value, r runs are performed using different orders of data presentation; then the number of clusters m_Θ most frequently met is identified and a plot of m_Θ versus Θ is performed. "Flat" areas in the plot is an indication of the existence of clusters; the flattest area is likely to correspond to the number of clusters underlying X (sometimes it is useful to also consider the second flattest area, provided that it is "significantly" large).

7.4.2 Clustering Refinement

Reassignment Procedure

This procedure is applied on a clustering that has already been obtained. It performs a single pass over the data set. The closest cluster for each vector, x, is determined. After all vectors have been considered, each cluster is redefined using the vectors identified as closest to it. If cluster representatives are used, they are re-estimated accordingly (a usual representative is the mean of all the vectors in a cluster).

To apply the reassignment procedure, type

$$[bel, new\_repre] = reassign(X, repre, order)$$

where *new_repre* contains in its columns the re-estimated values of the mean vectors of the clusters; all other parameters are defined as for function *BSAS*.

Merging Procedure

This procedure is also applied on a clustering $\Re$ of a given data set. Its aim is to merge clusters in $\Re$ that exhibit high "similarity" (low "dissimilarity"). Specifically, the cluster pair that, according to a preselected dissimilarity measure between sets (clusters), exhibits the lowest dissimilarity is determined. If this is greater than a (user-defined) cluster-dissimilarity threshold, M_1, the procedure is terminated. Otherwise, the two clusters are merged and the procedure is repeated on the resulting clustering.

Merging is sensitive to the value of parameter M_1, the choice of which mostly depends on the problem at hand. Loosely speaking, M_1 is chosen so that clusters lying "close" to each other and "away" from most of the rest are merged. This procedure should be used only in cases where the number of clusters is "large". In all cases, it is desirable to have the merging of clusters approved by an expert in the field of application.

شکل ۳-۶ تصویر ۳۰۰dpi از پایگاه داده ایجاد شده

- k-means cannot guarantee convergence to the global minimum of $J(\theta, U)$ (which, hopefully, corresponds to the best possible clustering). In other words, it returns clusterings corresponding to local minima of $J(\theta, U)$. Consequently, different initializations of the algorithm may lead to different final clusterings. Care must be taken in the initialization of θ_j (see the practical hints that follow). If the initial values for, say, m_1, of θ_j lie away from the region where the data vectors lie, they may never be updated. As a consequence, k-means algorithm will proceed as if there were only $m - m_1$ θ_j's.
- Accurate estimation of the number of clusters (representatives) is crucial for the algorithm, since a poor estimate will prevent it from unraveling the clustering structure of X. More specifically, if a larger number of representatives is used, it is likely that at least one "physical" cluster will be split into two or more. On the other hand, if a smaller number of representatives is used, two or more physical clusters are likely to be represented by a single representative, which in general will lie in a sparse region (with respect to the number of data points) between those clusters.
- The algorithm is sensitive to the presence of outliers (that is, points that lie away from almost all data vectors in X) and "noisy" data vectors. Such points are the result of a noisy process unrelated to the clustering structure of X. Since both outliers and noisy points are necessarily assigned to a cluster, they influence the respective mean representatives.
- k-means is suitable for real-valued data and, in principle, should *not* be used with discrete-valued data.

Practical Hints

- Assuming that m is fixed, and to increase our chances of obtaining a reliable clustering, we may run k-means several times, each time using different initial values for the representatives, and select the best possible clustering (according to J). Three simple methods for choosing initial values for θ_j's are (a) random initialization, (b) random selection of m data vectors from X as the initial estimates of θ_j's, and (c) utilization of the clustering output of a simpler (e.g., sequential) algorithm as input.
- Two simple ways to estimate m are:
  - Use the methodology described for the BSAS algorithm.
  - For each value of m, in a suitably chosen range $[m_{min}, m_{max}]$, run the k-means algorithm n_{run} times (each time using different initial values) and determine the clustering (among the n_{run} produced) that minimizes the cost function J. Let J_m be the value of J for the latter clustering. Plot J_m versus m and search for a significant local change (it appears as a significant "knee"). If such a knee occurs, its position indicates the desired number of clusters. Otherwise, it is an indication that there is no clustering structure (that contains compact clusters) in the data set.
- Two simple ways to deal with outliers are (a) to determine the points that lie at "large" distances from most of the data vectors in X and discard them, or (b) to run k-means and identify the clusters with very few elements. An alternative is to use algorithms that are less sensitive to outliers (this is the case with the PAM algorithm, discussed later).

Example 7.5.1. Generate and plot a data set, X_3, that consists of $N = 400$ 2-dimensional points. These points form four equally sized groups. Each group contains vectors that stem from Gaussian distributions

شکل ۳-۷ تصویر ۶۰۰dpi از پایگاه داده ایجاد شده

k-Means, or Isodata, Algorithm

This is the most widely known clustering algorithm, and its rationale is very simple. In this case, the parameter vectors θ_j (also called *cluster representatives* or simply *representatives*) correspond to points in the l-dimensional space, where the vectors of data set X live. k-means assumes that the number of clusters underlying X, m, is known. Its aim is to move the points $\theta_j, j = 1,\ldots,m$, into regions that are dense in points of X (clusters).

The k-means algorithm is of iterative nature. It starts with some initial estimates $\theta_1(0),\ldots,\theta_m(0)$, for the parameter vectors $\theta_1,\ldots,\theta_m$. At each iteration t,

- the vectors x_i that lie close to each $\theta_j(t-1)$ are identified and then
- the new (updated) value of θ_j, $\theta_j(t)$, is computed as the mean of the data vectors that lie closer to $\theta_j(t-1)$.

The algorithm terminates when no changes occur in θ_j's between two successive iterations. To run the k-means algorithm, type

$$[theta, bel, J] = k\_means(X, theta\_ini)$$

where

X is an $l \times N$ matrix whose columns contain the data vectors,

theta_ini is an $l \times m$ matrix whose columns are the initial estimates of θ_j (the number of clusters, m, is implicitly defined by the size of *theta_ini*),

theta is a matrix of the same size as *theta_ini*, containing the final estimates for the θ_j's,

bel is an N-dimensional vector whose ith element contains the cluster label for the ith data vector,

J is the value of the cost function given in Eq. (7.1) (see below) for the resulting clustering.

Remarks

- k-means is suitable for unraveling compact clusters.
- The k-means algorithm is a fast iterative algorithm because (a) in practice it requires only a few iterations to converge and (b) the computations required at each iteration are not complicated. Thus, it poses as a candidate for processing large data sets.
- It can be shown that the k-means algorithm minimizes the cost function

$$J(\theta, U) = \sum_{i=1}^{N} \sum_{j=1}^{m} u_{ij} ||x_i - \theta_j||^2 \tag{7.1}$$

where $\theta = [\theta_1^T,\ldots,\theta_m^T]^T$, $||.||$ stands for the Euclidean distance, and $u_{ij} = 1$ if x_i lies closest to θ_j; 0 otherwise. In words, *k-means minimizes the sum of the squared Euclidean distances of each data vector from its closest parameter vector*. When the data vectors of X form m compact clusters (with no significant difference in size), it is expected that J is minimized when each θ_j is placed (approximately) in the center of each cluster, provided that m is known. This is not necessarily the case when (a) the data vectors do not form compact clusters, or (b) their sizes differ significantly, or (c) the number of clusters, m, has not been estimated correctly.

شکل ۳-۸ تصویر ۶۰۰dpi از پایگاه داده ایجاد شده

Remarks

- In its original form, BSAS is suitable for unraveling compact clusters (i.e., clusters whose points are aggregated around a specific point in the data space).
- The algorithm is sensitive to the order of data presentation and the choice of the parameter Θ. If the data are presented in a different order and/or the parameter Θ is given a different value, a different clustering may result.
- BSAS is *fast*, because it requires a single pass on the data set, X; thus, it is a good candidate for processing large data sets. However, in several cases the resulting clustering may be of low quality. Improvements can be achieved in a refinement step, as discussed in the following subsection.
- Sometimes, the clustering returned by BSAS is used as a starting point for other more sophisticated clustering algorithms, to generate clusterings of improved quality.
- BSAS results in a rough estimate of the number of clusters underlying the data set at hand.

To more accurately estimate the number of clusters in X, BSAS may be run for different values of Θ in a range $[\Theta_{min}, \Theta_{max}]$. For each value, r runs are performed using different orders of data presentation; then the number of clusters m_Θ most frequently met is identified and a plot of m_Θ versus Θ is performed. "Flat" areas in the plot is an indication of the existence of clusters; the flattest area is likely to correspond to the number of clusters underlying X (sometimes it is useful to also consider the second flattest area, provided that it is "significantly" large).

7.4.2 Clustering Refinement

Reassignment Procedure

This procedure is applied on a clustering that has already been obtained. It performs a single pass over the data set. The closest cluster for each vector, x, is determined. After all vectors have been considered, each cluster is redefined using the vectors identified as closest to it. If cluster representatives are used, they are re-estimated accordingly (a usual representative is the mean of all the vectors in a cluster).

To apply the reassignment procedure, type

$$[bel, new\_repre] = reassign(X, repre, order)$$

where *new_repre* contains in its columns the re-estimated values of the mean vectors of the clusters; all other parameters are defined as for function *BSAS*.

Merging Procedure

This procedure is also applied on a clustering $\Re$ of a given data set. Its aim is to merge clusters in $\Re$ that exhibit high "similarity" (low "dissimilarity"). Specifically, the cluster pair that, according to a preselected dissimilarity measure between sets (clusters), exhibits the lowest dissimilarity is determined. If this is greater than a (user-defined) cluster-dissimilarity threshold, M_1, the procedure is terminated. Otherwise, the two clusters are merged and the procedure is repeated on the resulting clustering.

Merging is sensitive to the value of parameter M_1, the choice of which mostly depends on the problem at hand. Loosely speaking, M_1 is chosen so that clusters lying "close" to each other and "away" from most of the rest are merged. This procedure should be used only in cases where the number of clusters is "large". In all cases, it is desirable to have the merging of clusters approved by an expert in the field of application.

شکل ۳-۹ تصویر ۶۰۰dpi از پایگاه داده ایجاد شده

شکل ۳-۱۰ تصویر ۳۰۰dpi بزرگنمایی شده از پایگاه داده

شکل ۳-۱۱ تصویر ۶۰۰dpi بزرگنمایی شده از پایگاه داده

در تصاویر ۳-۱۰ و ۳-۱۱ مشاهده می‌شود تصویری که با درجه‌ی تفکیک ۶۰۰dpi روبش شده نسبت به تصویر با درجه‌ی تفکیک ۳۰۰dpi از نظر دیداری بخصوص در لبه‌های تصویر متنی کیفیت بهتری دارد.

۳-۷ واژه‌نامه

در چند دهه اخیر استفاده از واژه‌نامه در فرا تفکیک‌پذیری نظر بسیاری از محققان را به خود جلب کرده است. ایده‌ی این روش بر این اساس استوار است که می‌توان یک وصله از تصویر را بصورت ترکیب خطی از وصله‌های دیگر بیان کرد. بنابراین باید ضرایب بازسازی هر وصله که در ترکیب خطی شرکت دارد محاسبه شود. برای محاسبه‌ی ضرایب بازسازی هر وصله به یک واژه‌نامه نیاز است. برای ساخت واژه‌نامه روش‌های زیادی ارائه شده است. یکی از الگوریتم‌هایی که برای این منظور استفاده می‌شود، الگوریتم K-SVD [۷۲] می‌باشد. هزینه محاسباتی این الگوریتم زیاد است و هر چه تعداد اتم‌ها بیشتر باشد زمان ساخت واژه‌نامه نیز بالاتر می‌رود. راه دیگر این است که مستقیم از وصله‌های تصویر استفاده شود. به‌عبارتی دیگر هر وصله برداری می‌شود و بعنوان یک اتم در داخل واژه‌نامه قرار می‌گیرد. واژه‌نامه‌ای که بر این اساس

ساخته می‌شود، واژه‌نامه‌ی مبتنی بر وصله نامیده می‌شود. روش دیگر برای ساخت واژه‌نامه استفاده از ویژگی است. در [۷۳] با استفاده از ۴ فیلتر از تصویر ورودی ویژگی استخراج شده است. سپس بردار ویژگی‌های بدست آمده بعنوان اتم در داخل واژه‌نامه قرار گرفته‌اند. اگر هر وصله بصورت ترکیب خطی از اتم‌ها (وصله-هایی که به بردار تبدیل شده‌اند) بیان شود و n تعداد اتم‌ها در واژه‌نامه باشد، خواهیم داشت:

$$p_i = \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + ... = \sum_{j=1}^{n} \alpha_j p_j \qquad (۲۶-۳)$$

$$p_i = D_l q \qquad (۲۷-۳)$$

در روابط (۳-۲۶) و (۳-۲۷) p_i وصله‌ی وضوح پایین i-ام و $q = (\alpha_1 \ \ \alpha_2 \ \ ... \ \ \alpha_n)^T$ بردار ضرایب است که اهمیت هر اتم را برای ساخت نسخه‌ی وضوح بالای هر وصله‌ی وضوح پایین تعیین می‌کند. $D_l = (p_1, p_2, ..., p_n)$ واژه‌نامه‌ی وضوح پایین می‌باشد. وصله‌ی بازسازی شده باید تا حد ممکن به وصله مورد نظر نزدیک باشد. بردار ضرایب q با حل معادله بهینه‌سازی (۳-۲۸) به‌دست می‌آید:

$$q = \arg\min_q \| D_l q - p_i \| \qquad (۲۸-۳)$$

۳-۸ بهینه‌سازی

در حالت کلی هدف از فرا تفکیک‌پذیری تولید تصویر وضوح بالا (y_h) از تصویر ورودی (z_l) است. طبق رابطه‌ی (۳-۲۹) تصویر وضوح بالا و پایین به‌هم مرتبط می‌شوند:

$$z_l = SHy_h \qquad (۲۹-۳)$$

در رابطه (۳-۲۹) S و H به ترتیب عملگرهای کاهش نرخ نمونه‌ها و مات‌کنندگی می‌باشد. واژه-نامه‌های وضوح بالا و پایین از تصویر ورودی و نسخه‌ی کاهش مقیاس یافته‌ی آن تولید می‌شوند. از آنجایی که اتم‌ها در واژه‌نامه‌ی وضوح پایین D_l و واژه‌نامه‌ی وضوح بالا D_h به هم مرتبط هستند و از مکان‌های متناظر هم استخراج می‌شوند، انتظار می‌رود برای ضرایب q برای نسخه وضوح بالا و پایین یک وصله تغییر

نکند. بنابراین نسخه‌ی وضوح بالای یک وصله‌ی وضوح پایین با استفاده از واژه‌نامه‌ی وضوح بالا با استفاده از رابطه (۳-۳۰) تولید می‌شود:

$$x_i = D_h q \qquad (۳۰-۳)$$

در این رابطه x_i نسخه‌ی وضوح بالای وصله‌ی p_i است. از آنجایی‌که در رابطه‌ی (۳-۲۷) ماتریس D یک ماتریس مربعی نیست، این موضوع مسئله را بد حالت می‌کند و بی‌نهایت جواب خواهد داشت. برای حل این مشکل عبارتی را به مسئله اضافه می‌کنند. این عبارت را تنظیم کننده می‌نامند. یکی از تنظیم‌کننده‌های پرکاربرد تنظیم‌کننده تیخونوف [۷۴] است.

$$\vec{q} = \arg\min_q \|p - D_l q\|_2^2 + \lambda_{tik} \|q\|_2^2 \qquad (۳۱-۳)$$

به‌طوری که λ_{tik} پارامتر تنظیم‌کننده می‌باشد. حل رابطه (۳-۳۱) به‌صورت رابطه (۳-۳۲) به‌دست می‌آید:

$$\vec{q} = ((D_l)^t D_l + \lambda_{tik} I)^{-1} (D_l)^t p \qquad (۳۲-۳)$$

از ضرایب q برای ساخت x_i مطابق رابطه (۳-۳۳) استفاده می‌شود:

$$x_i = D_h q \qquad (۳۳-۳)$$

به‌طوری که ماتریس تصویر[1] از رابطه (۳-۳۴) به‌دست می‌آید:

$$P_G = ((D_l)^t D_l + \lambda_{tik} I)^{-1} (D_l)^t \qquad (۳۴-۳)$$

ماتریس P_G فقط یک‌بار محاسبه می‌شود، یعنی در طول اجرای الگوریتم فرا تفکیک‌پذیری فقط نیاز است که ماتریس تصویر در ورودی وضوح پایین ضرب شود.

در رابطه (۳-۳۱) به همه اتم‌ها وزن یکسانی داده می‌شود، به این معنی که وصله‌ی ورودی نسبت به تمام اتم‌های دیکشنری با وزن یکسان سنجیده می‌شود. این روش خیلی دقیق نیست و فقط سرعت محاسبه ضرایب q را بالا می‌برد. روش دوم برای حل مسئله (۳-۲۷) استفاده از ماتریس Γ است. Γ ماتریسی

[1] Projection

است که برحسب فاصله‌ی وصله ورودی به اتم‌های واژه‌نامه، به اتم‌ها وزن‌های متفاوتی می‌دهد. بنابراین رابطه (۳-۲۷) را می‌توان به‌صورت رابطه (۳-۳۵) نوشت:

$$\vec{q} = \arg\min_{q} \|D_l q - \tilde{p}_l\|_2^2 + \lambda_{tik} \|\Gamma q\|_2^2 \qquad (۳۵-۳)$$

به‌طوری که λ_{tik} پارامتر تنظیم‌کننده است. همان‌طور که در [۷۵] پیشنهاد شده است ماتریس Γ به‌صورت رابطه (۳-۳۶) تعریف می‌شود:

$$\Gamma_{jj} = \left\| p - d_j^L \right\|_2^2 \qquad (۳۶-۳)$$

در این رابطه d_j^L ستون j-ام واژه‌نامه D_l می‌باشد. در نتیجه رابطه محاسبه ضرایب q به‌صورت رابطه (۳-۳۷) بازنویسی می‌شود.

$$\vec{q} = ((D_l)^t D_l + \lambda_{tik} \Gamma\Gamma^t)^{-1} (D_l)^t \tilde{p}_l \qquad (۳۷-۳)$$

۳-۹ روش بازسازی سراسری

باید ذکر شود که رابطه (۳-۲۷) به‌دلیل وجود نویز و این فرض که هر وصله براساس وصله‌های همسایه قابل بیان است، جواب واقعی برای وصله‌ی وضوح بالای p_h وجود نخواهد داشت. اگر تصویر به‌دست آمده با X_0 نشان داده شود، این اختلاف به‌وسیله‌ی تصویر X_0 روی فضای $z_l = SHy_h$ با استفاده از رابطه (۳-۳۸) محاسبه می‌شود:

$$y_h^* = \arg\min_{yh} \|SHy_h - z_l\|_2^2 + c\|y_h - X_0\|_2^2 \qquad (۳۸-۳)$$

به‌طوری که در رابطه (۳-۳۸) S و H به‌ترتیب عملگرهای کاهش نرخ نمونه‌ها و ماتی می‌باشند. y_h تصویر وضوح بالا، z_l تصویر وضوح پایین و C عددی ثابت است. برای حل رابطه (۳-۳۸) از روش گرادیان نزولی استفاده شده است. بنابراین طبق رابطه (۳-۳۹) خواهیم داشت:

$$y_h(t+1) = y_h(t) + \beta[H^T S^T (z_l - SHy_h) + c(y_h - X_0)] \qquad (۳۹-۳)$$

در رابطه‌ی (۳-۳۹) $y_h(t)$ تخمین تصویر وضوح بالا در تکرار t-ام و β گام گرادیان نزولی است. y_h^* تخمین نهایی تصویر وضوح بالا می‌باشد. این تصویر تا حد ممکن به تصویر تخمین اولیه (X_0) نزدیک است. به‌منظور افزایش دقت بازسازی سراسری و از بین بردن نویز به رابطه (۳-۳۹) عبارت پیشین نیز اضافه می‌شود.

$$y_h^* = \arg\min_{yh} \|SHy_h - z_l\|_2^2 + c\|y_h - (X_0)\|_2^2 + \tau\rho(y_h) \qquad (۴۰-۳)$$

در رابطه (۳-۴۰)، $\rho(y_h)$ تابع جریمه‌ای است که عبارت پیشینی را در ارتباط با تصویر y_h به‌وجود می‌آورد. این تابع می‌تواند Huber، MRF، [۷۶] TV[1] و [۷۷] BTV[2] باشد.

[1] Total Variation
[2] Bilateral Toral Variation

فصل چهارم:

روش پیشنهادی

مشاهده شد که یکی از ویژگی‌های مهم تصاویر متنی درجه‌ی تفکیک مکانی بالای این تصاویر می‌باشد. این ویژگی به خوانایی انسانی و ماشینی تصاویر متنی کمک می‌کند. مشکل این ویژگی زیاد شدن حجم فایل تصویر متنی است. بنابراین نیاز به فشرده‌سازی تصاویر متنی امری ضروری به نظر می‌رسد. از طرف دیگر پس از فشرده‌سازی تصاویر متنی با توجه به اثرات مخرب ناشی از فشرده‌ساز و کاهش ابعاد تصویر ورودی برای رسیدن به حجم فایل‌های کم، خوانایی تصویر متنی کم می‌شود.

حال که به اهمیت فشرده‌سازی در تصاویر متنی پی بردیم، سوالی که وجود دارد این است که روش‌های فشرده‌سازی موجود تا چه میزان می‌توانند فشرده‌سازی را برای ما انجام دهند؟ آیا در نرخ بیت‌های پایین بدست آمده توسط این روش‌ها وضوح تصویر متنی و خوانایی آن مطلوب است؟

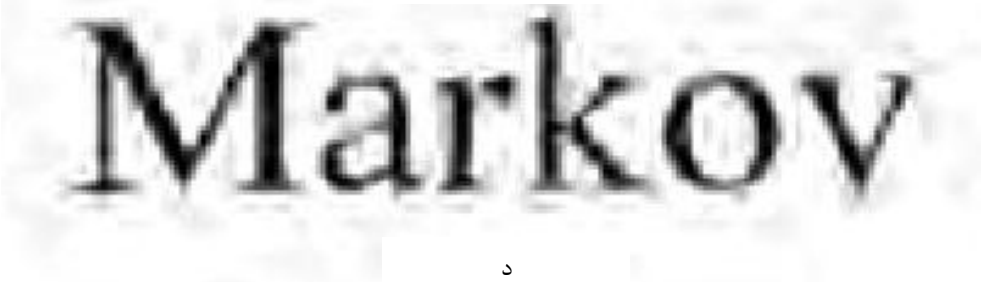
۴-۱ بررسی اثر روش‌های فشرده‌سازی روی تصاویر متنی

ابتدا اثر فشرده‌سازی حاصل از روش‌های مختلف روی تصاویر متنی نشان داده می‌شود . یک تصویر با وضوح بالا به فشرده‌سازهای JPEG، JPEG2000 و SPIHT داده‌شده و اثرات هر کدام از روش‌ها روی تصویر متنی فشرده شده بررسی شده‌است. نتایج حاصل برای نرخ بیت ۰۹/ در شکل ۴-۱ نشان داده شده است. همان‌طور که در شکل ۴-۱ دیده می‌شود، اثر فشرده‌ساز JPEG روی تصویر متنی به‌صورت اثر بلوکی و اثر فشرده‌سازهای JPEG2000 و SPIHT به‌صورت اثرات نشتی است. همه فشرده‌سازها برای نرخ بیت ۰۹/ عمل فشرده‌سازی را انجام داده‌اند. می‌توان از شکل ۴-۱ این‌طور نیز نتیجه گرفت که در یک نرخ بیت مشخص شدت اثر مخربِ روش‌های فشرده‌سازی مختلف روی تصویر متنی متفاوت است. به‌طوری‌که روش فشرده‌سازی JPEG اثر مخرب بیشتری روی تصویر نسبت به دو روش دیگر داشته است. به‌عبارت دیگر خوانایی تصویر متنی در دو روش دیگر نسبت به روش JPEG بهتر است.

الف

ب

ج



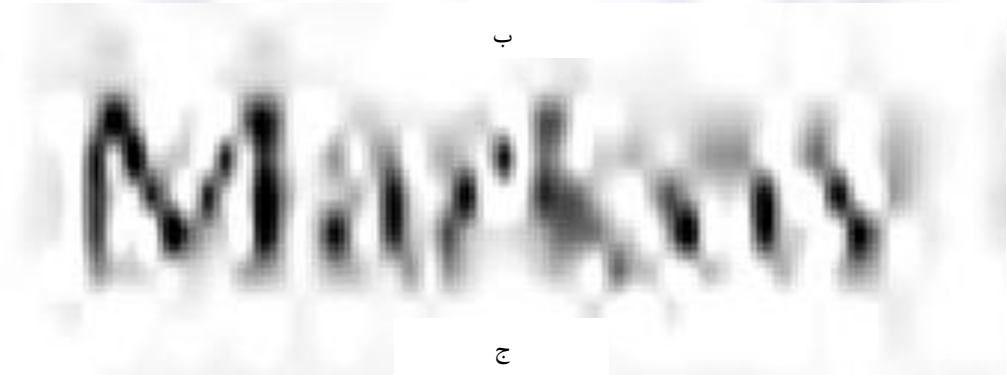
د

شکل ۴-۱ الف) تصویر وضوح بالای اصلی ب) اثر فشرده‌سازی JPEG ج) اثر فشرده‌سازی JPEG2000 د) اثر فشرده‌سازی SPIHT

در ادامه به‌منظور درک بهتری از اثر روش‌های JPEG2000 و SPIHT فشرده‌سازی تصویر متنی در نرخ بیت کمتری نسبت به حالت قبل بررسی شده است. در شکل ۴-۲ می‌توان نتایج حاصل از روش‌های فشرده‌سازی مذکور را برای نرخ بیت‌های ۰/۰۲ مشاهده کرد. اثر نشتی ناشی از این روش‌ها در این شکل به خوبی مشخص است.

الف

ب



ج

شکل ۴-۲ الف) تصویر وضوح بالای اصلی ب) اثر فشرده‌سازی با JPEG2000 ج) اثر فشرده‌سازی با SPIHT

در شکل ۴-۲ اثر ناشی از فشرده‌سازی تصویر متنی در نرخ بیت کم با استفاده از روش‌های JPEG2000 و SPIHT نشان می‌دهد که این اثرات تا چه اندازه می‌توانند مخرب باشند. در شکل ۴-۲ (ج) خوانایی انسانی نیز بسیار سخت می‌شود و اگر شخصی تصویر اصلی را ندیده باشد شاید در خواندن کلمه اشتباه کند. تا اینجای کار اثرات مخرب ناشی از روش‌های فشرده‌سازی مختلف روی تصویر متنی مشاهده شد. باید به این نکته نیز توجه شود که این روش‌ها تا چه نرخ بیتی می‌توانند فشرده‌سازی را انجام دهند.

در شکل ۴-۱ تصویر حاصل از فشرده‌سازی JPEG در کمترین نرخ فشرده‌سازی توسط این روش قرار دارد یعنی اینکه با استفاده از روش فشرده‌سازی JPEG برای این تصویر نمی‌توان به نرخ بیت کمتر از

۰/۰۹ رسید. این یک محدودیت در روش فشرده‌سازی JPEG ایجاد می‌کند. برای روش‌های JPEG2000 و کدگذاری SPIHT فشرده‌سازی می‌تواند در نرخ بیت کمتری انجام شود. در شکل ۴-۲ کمترین میزان فشرده‌سازی حاصل از فشرده‌ساز JPEG2000 روی تصویر متنی نشان داده شده است. یعنی فشرده‌ساز JPEG2000 روی این تصویر خاص نمی‌تواند نرخ بیت کمتر از ۰/۰۲ را داشته باشد. اما برای روش کدگذاری SPIHT محدودیتی در نرخ بیت وجود ندارد و می‌توان تا هر میزان دل‌خواهی تصویر را فشرده کرد اما محدودیت در این روش روی کاربرد می‌باشد. به‌عبارت دیگر در روش کدگذاری SPIHT محدودیت زمانی ایجاد می‌شود که خوانایی انسانی و ماشینی به حد نامطلوب خود برسد.

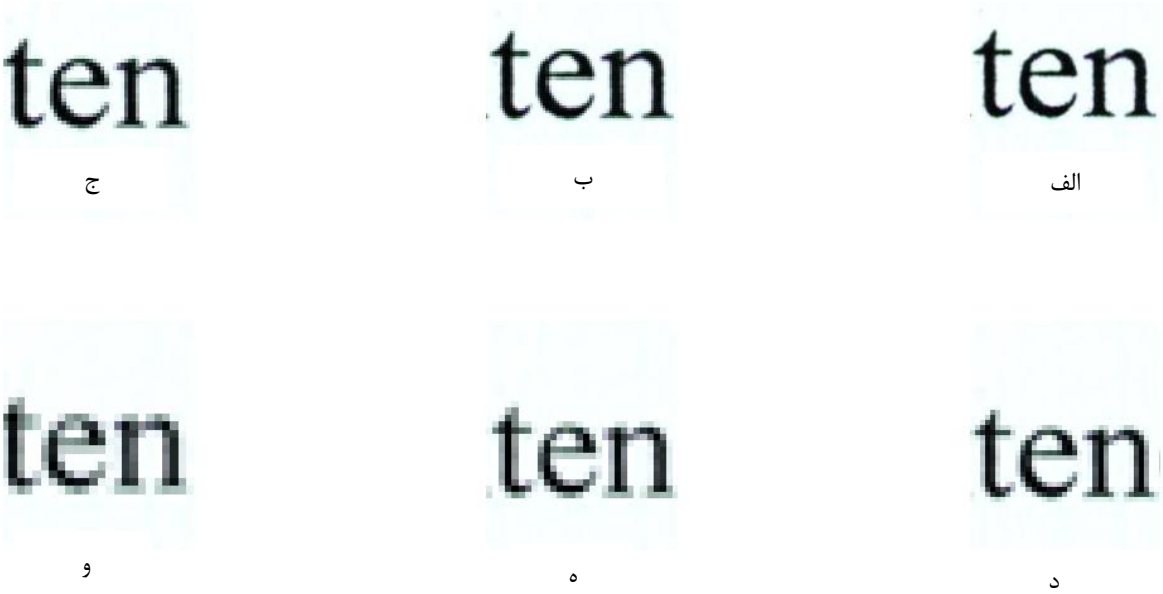
از بررسی روش‌های فشرده‌سازی روی تصاویر متنی می‌توان نتیجه گرفت که در روش‌های JPEG و JPEG2000 برای رسیدن به نرخ بیت‌های پایین محدودیت وجود دارد. روش پیشنهادی سعی در به حداقل رساندن این محدودیت دارد.

۴-۲ طرح پیشنهادی جهت رسیدن به نرخ بیت‌های کمتر

در فشرده‌سازی تصاویر متنی برای تصویری در یک ابعاد مشخص و ثابت فشرده‌ساز می‌تواند حجم فایل تصویر فشرده شده را تا میزان مشخصی کم کند. با توجه به اینکه در این پایان‌نامه روی تصاویر با تفکیک مکانی بالا کار شده‌است و این تصاویر حجم فایل زیادی برای ذخیرسازی اشغال می‌کنند، هدف این است تا جایی که امکان دارد حجم فایل تصویر فشرده‌شده جهت ذخیره‌سازی کمتر شود.

ایده‌ی پیشنهادی در این پایان‌نامه کاهش ابعاد تصویر متنیِ با درجه‌ی تفکیک مکانی بالا برای رسیدن به نرخ بیت‌های کمتر است. برای این منظور می‌توان از یکی از روش‌های کاهش ابعاد مانند روش دومکعبی ابتدا تصویر متنیِ با درجه‌ی تفکیک مکانی بالای ورودی را کاهش ابعاد داد و سپس آنرا با یکی از روش‌های فشرده‌سازی معرفی شده، فشرده کرد.

باید به این نکته توجه شود که عمل کاهش ابعاد خود می‌تواند روی وضوح تصویر متنی ورودی اثر مخربی داشته باشد. به این صورت که کاهش ابعاد می‌تواند سبب هموار کردن جزئیات فرکانس بالا شود. در تصاویر متنی جزئیات فرکانس بالا، لبه‌های متن می‌باشند. با توجه به اینکه در تصاویر متنی لبه‌ها مهم‌ترین بخش این تصاویر را تشکیل می‌دهند، اثر کاهش ابعاد روی این ویژگی‌های تصاویر متنی مطلوب نیست. در شکل ۴-۳ اثر کاهش ابعاد تصاویر متنی روی لبه‌های متن نشان داده شده است.



الف ب ج

د ه و

شکل ۴-۳ اثر کاهش ابعاد روی تصویر با درجه تفکیک مکانی بالا. الف) تصویر اصلی ب) کاهش ابعاد با ضریب ۲. ج) کاهش ابعاد با ضریب ۳. د) کاهش ابعاد با ضریب ۴. ه) کاهش ابعاد با ضریب ۵. و) کاهش ابعاد با ضریب ۶.

همان‌طور که از شکل ۴-۳ مشخص است هرچه ابعاد تصویر بیشتر کاهش داده می‌شود لبه‌های تصویر بیشتر هموار می‌شوند. این اثر می‌تواند روی خوانایی انسانی و ماشینی تصاویر متنی تأثیر گذار باشد.

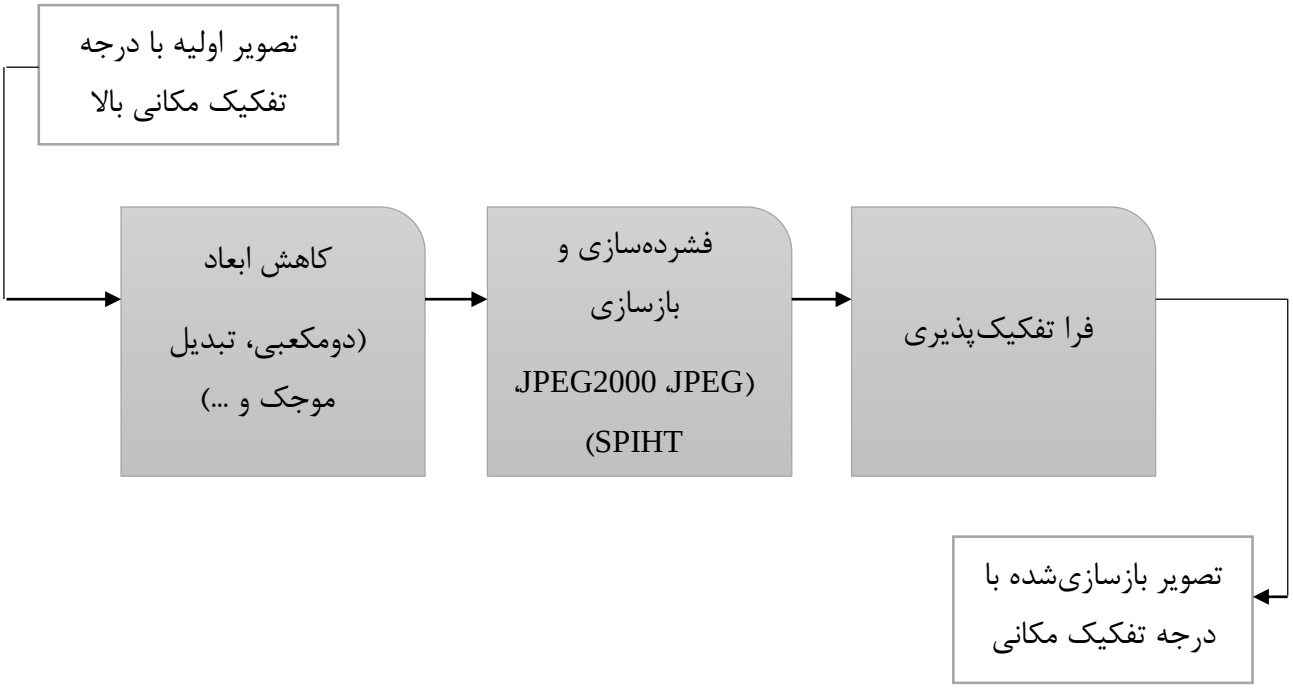
تا اینجا روشی جهت افزایش میزان فشرده‌سازی تصاویر متنی یا کاهش نرخ بیت این تصاویر با استفاده از روش‌های فشرده‌سازی موجود پیشنهاد شده‌است. اما روش پیشنهادی در ظاهر به نظر می‌رسد که با حرف‌های گفته‌شده در رابطه با ویژگی‌های تصاویر متنی در تناقض است. درجه‌ی تفکیک مکانی از ویژگی‌های مهم تصاویر متنی معرفی شد اما در روش پیشنهادی، ابعاد تصویر با درجه‌ی تفکیک مکانی بالای

ورودی کاهش داده شده‌است. برای حل این مشکل به روشی نیاز است که بتواند به‌طور همزمان در کنار بهبود اثرات مخرب ناشی از فشرده‌سازی و کاهش ابعاد تصاویر متنی، ابعاد تصاویر متنی را نیز افزایش دهد. برای این منظور از روش‌های فرا تفکیک‌پذیری در مرحله بازسازی استفاده شده‌است.

۴-۳ مرحله بازسازی

با توجه به مشکلات مطرح شده برای رسیدن به میزان فشرده‌سازی بالاتر، در مرحله بازسازی نیاز به روشی است که بتواند به‌طور همزمان هم اثرات مخرب ناشی از فشرده‌سازی و کاهش ابعاد را بهبود دهد و هم ابعاد تصویر را بزرگ کند. برای این منظور از روش‌های فرا تفکیک‌پذیری در مرحله بازسازی استفاده شده‌است که طبق دانسته‌ی ما تا کنون کاری در این زمینه یعنی ترکیب فرا تفکیک‌پذیری با روش‌های فشرده‌سازی تصاویر متنی برای رسیدن به میزان فشرده‌سازی بیشتر، انجام نشده است.

مراحل روش پیشنهادی در شکل ۴-۴ نشان داده شده است. روش پیشنهادی در شکل ۴-۴ این گونه بیان می‌کند که ابتدا ابعاد تصویر با درجه‌ی تفکیک مکانی بالای ورودی با استفاده از یکی از روش‌های کاهش ابعاد مانند دومکعبی کاهش داده‌شود. این کار برای رسیدن به میزان فشرده‌سازی بیشتر در روش‌های فشرده‌سازی مانند JPEG و JPEG2000 می‌باشد. پس از کاهش ابعاد تصویر ورودی، آن با یکی از روش‌های فشرده‌سازی، فشرده شود. اینکه تا چه اندازه بتوان میزان فشرده‌سازی را افزایش داد به توانایی روش در مرحله بازسازی بستگی دارد. چون در تصاویر متنی محدودیت روی خوانایی انسانی و ماشینی این تصاویر می‌باشد میزان فشرده‌سازی را تا جایی می‌توان افزایش داد که روش استفاده شده در مرحله بازسازی بتواند اثرات مخرب ایجاد شده‌ی ناشی از فشرده‌سازی و کاهش ابعاد را برطرف نماید و تصویری با وضوح بالا در اختیار قرار دهد. در روش پیشنهادی از فرا تفکیک‌پذیری مبتنی بر دورن‌یابی و آموزش استفاده شده‌است.



شکل ۴-۴ مراحل روش پیشنهادی

۴-۴ فرا تفکیک‌پذیری مبتنی بر درون یابی

جهت بازسازی تصاویر فشرده‌شده روش فرا تفکیک‌پذیری مبتنی بر درون‌یابی در نظر گرفته شده است که می‌تواند اثرات نامطلوب حاصل از فشرده‌سازی تصاویر متنی را تا حد ممکن بر طرف کند.

پس از اینکه تصاویر را فشرده کردیم باید توانایی روش در بازسازی تصویر را ارزیابی می‌کردیم. ایده‌ی اصلی روش استفاده شده در مرحله بازسازی با استفاده از فرا تفکیک‌پذیری مبتنی بر درون‌یابی از مقاله [۷۸] گرفته شده است.

در روش استفاده شده ابتدا تصویر ورودی را به سه زیر لایه شامل؛ ۱) لایه شامل اطلاعات لبه تصویر متن ۲) لایه پیش‌زمینه[1] و ۳) لایه پس‌زمینه تجزیه می‌کنیم. سپس با توجه به اهمیت هر لایه به روشی

[1] Fore Ground

مجزا آن لایه را بزرگ‌نمایی می‌کنیم. سپس لایه‌ها را با هم ترکیب می‌کنیم و به تصویر وضوح بالای نهایی خواهیم رسید. پس اولین مرحله مربوط به تجزیه تصویر ورودی به سه زیر لایه گفته شده می‌باشد.

۴-۴-۱ لایه‌بندی تصویر

لایه‌بندی تصویر بیان می‌کند که تصویر I می‌تواند به‌صورت مجموع وزن‌داری از تصویر پیش‌زمینه F و تصویر پس‌زمینه B با استفاده از لایه‌بند آلفا α به‌عنوان وزن مدل شود.

$$I = \alpha F + (1-\alpha)B \qquad (۱-۴)$$

در رابطه (۴-۱)، α می‌تواند هر مقداری در بازه [۰,۱] داشته باشد. اگر $\alpha_k = 0$ در این صورت پیکسل I_k پس‌زمینه و اگر $\alpha_k = 1$ پیکسل I_k را پیش‌زمینه می‌نامند. در غیر این صورت پیکسل ترکیبی از پیش‌زمینه و پس‌زمینه است. لایه‌بند تصویر با هدف بازیابی لایه آلفا α، پیش‌زمینه F و پس‌زمینه B برای تصویر ورودی I اعمال می‌شود.

شکل ۴-۵ مثالی از لایه‌بندی تصویر متنی را نشان می‌دهد. شکل ۴-۵(الف) تصویر رنگی ورودی را نشان می‌دهد. شکل ۴-۵(ب) لایه آلفا را بعد از لایه‌بندی تصویر نشان می‌دهد. شکل‌های ۴-۵(ج) و ۴-۵(د) به‌ترتیب لایه‌های پیش‌زمینه و پس‌زمینه را نشان می‌دهند. لایه آلفا شامل اطلاعات مهم مربوط به لبه تصویر متنی و لایه‌های پیش‌زمینه و پس‌زمینه نیز به‌ترتیب شامل اطلاعات مربوط به پیش‌زمینه و پس‌زمینه هستند. همان‌طور که از شکل ۴-۵ مشخص است لایه آلفا شامل لبه‌های هموار متن می‌باشد. لایه‌های پیش‌زمینه و پس‌زمینه تنها شامل اطلاعات رنگ متن و پس‌زمینه هستند. این نکته برای فرا تفکیک‌پذیری تصاویر متن بسیار مفید است و به جلوگیری از مصنوعات و ماتی با پردازش جداگانه روی لبه متن و اجزای رنگ کمک می‌کند.

شکل ۴-۵ مثالی از لایه‌بندی تصویر متنی. (الف) تصویر ورودی. (ب) لایه آلفا. (ج) لایه پیش‌زمینه. (د) لایه پس‌زمینه. [۷۸]

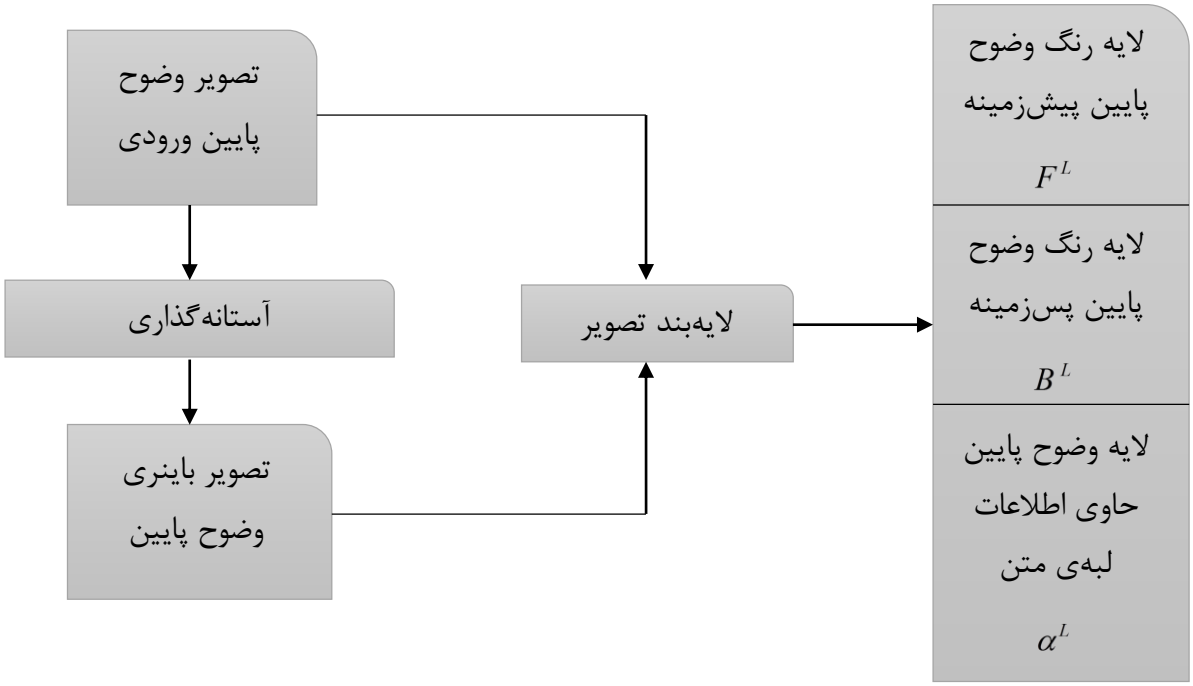
همان‌طور که گفته شد اولین مرحله از روش ارئه شده، تجزیه تصویر ورودی I^L به سه لایه متفاوت شامل لایه وضوح پایین رنگ پیش‌زمینه F^L، لایه وضوح پایین رنگ پس‌زمینه B^L و لایه وضوح پایین لبه متن α^L است. بازیابی α^L، F^L و B^L با داشتن تنها تصویر ورودی I^L یک مسأله نامعین است. برای حل این مسأله روش‌های لایه‌بندی متفاوتی ارائه شده است [۸۰-۷۹]. این روش‌ها به‌طور معمول زمان‌بر می-باشند. در اینجا ما از یک روش کارا به‌منظور لایه‌بندی استفاده می‌کنیم. مرحله اول روش انجام شده در شکل ۴-۶ نشان داده شده است. همان‌طور که از این شکل دیده می‌شود ابتدا تصویر با وضوح پایین ورودی در این مرحله به سه زیر لایه شامل لایه حاوی اطلاعات متن، پیش‌زمینه و پس‌زمینه تجزیه می‌شود.

مرحله‌ی اول : تجزیه تصویر به سه لایه

F^L : لایه رنگ وضوح پایین پیش‌زمینه

B^L : لایه رنگ وضوح پایین پس‌زمینه

α^L : لایه وضوح پایین حاوی اطلاعات لبه‌ی متن



شکل ۴-۶ مرحله اول روش فرا تفکیک‌پذیری مبتنی بر درون‌یابی

۴-۴-۲ لایه شامل اطلاعات لبه متن

ابتدا تصویر ورودی را به یک تصویر دودویی تبدیل می‌کنیم. تصویر دودویی A به‌دست آمده به‌منظور تصویر پوشش متن در نظر گرفته می‌شود. ما از روش آستانه‌گذاری متغیر و اتسو به منظور دودویی نمودن تصویر ورودی استفاده نموده‌ایم. به‌منظور استخراج اطلاعات لبه تصویر متن اصلی، از فیلتر حفظ لبه

یعنی فیلتر هدایت که در فصل ۳ معرفی شد؛ استفاده شده است. تصویر دودویی به‌دست آمده به‌عنوان تصویر اصلی و تصویر ورودی به‌عنوان تصویر راهنما به فیلتر داده می‌شود. فیلتر هدایت تا زمانی‌که براساس مدل خطی محلی است می‌تواند تصویر متن دودویی A را به‌عنوان تقریبی از لبه‌های تصویر متن اصلی در نظر بگیرد و اطمینان دهد که $\nabla A = \nabla I^L$. اگر G را معرف فیلتر هدایت در نظر بگیریم، لایه‌ی حاوی اطلاعات لبه متن α^L را می‌توان به‌صورت زیر محاسبه نمود:

$$\alpha^L = G(A, I^L; r; \varepsilon) \qquad (۲-۴)$$

در رابطه (۴-۲) r و ε دو پارامتر فیلتر هدایت می‌باشند. شکل ۴-۵(ب) لایه حاوی اطلاعات لبه متن α^L را نشان می‌دهد.

۴-۴-۳ لایه‌های پیش‌زمینه و پس‌زمینه

به‌طور خاص مدل خطی محلی بیان می‌کند که در پنجره‌های محلی کوچک، کانال α می‌تواند به‌صورت ترکیب خطی از تصویر I نمایش داده شود.

$$\alpha_k = aI_k^c + b \qquad k \in \omega_i \qquad (۳-۴)$$

در رابطه (۴-۳) ω_i پنجره کوچکی حول پیکسل i و c بیان‌گر کانال رنگ می‌باشد. a و b در پنجره محلی ثابت می‌باشند. برای سادگی در ادامه شاخص کانال c حذف شده است. براساس رابطه (۴-۱)، به‌راحتی به‌دست می‌آید که $a = \frac{1}{F - B}$ و $b = \frac{B}{B - F}$ است. با کمینه کردن خطای بازسازی و عبارت تنظیم، بازسازی پیش‌زمینه و پس‌زمینه می‌تواند با کمینه کردن تابع انرژی زیر حل شود [۷۸]:

$$\sum_{i \in w_k} (((F_K - B_K)\alpha_i + B_K - I_i)^2 + \gamma(F_K - B_K)^2) \qquad (۴-۴)$$

$\gamma(F_K - B_K)^2$ عبارت تنظیم می‌باشد و γ ثابتی است که معمولاً برابر ۰,۰۱ تنظیم می‌شود. تابع انرژی نمایش داده شده در رابطه (۴-۴) می‌تواند در شکل بسته حل شود و حل آن به صورت زیر می‌باشد [۸۱]:

$$F_K - B_K = \frac{\frac{1}{|w_K|}\sum_{i \in W_K} \alpha_i I_i - \mu_K \bar{I}_K}{\sigma_K^2 + \gamma} \qquad (۵-۴)$$

$$B_K = \bar{I}_K - (F_K - B_K)\mu_K \qquad (۶-۴)$$

در رابطه (۴-۵) W_K پنجره ۳*۳ می‌باشد زیرا استفاده از پنجره بزرگ‌تر ممکن است فرض خطی محلی را نقض کند. μ_K و σ_K^2 در روابط (۴-۵) و (۴-۶) به‌ترتیب میانگین و واریانس α در w_k می‌باشند و $|w_k|$ تعداد پیکسل‌ها در پنجره محلی است.

۴-۴-۴ بزرگ‌نمایی جداگانه لایه‌ها و ترکیب آن‌ها

این قسمت از روش فرا تفکیک‌پذیری شامل افزایش ابعاد لایه‌های به‌دست آمده از قسمت قبل می‌باشد. در شکل ۴-۷ این مرحله نشان داده شده است.

همان‌طور که در شکل ۴-۵(ج) و (د) دیده می‌شود، لایه‌های پیش‌زمینه و پس‌زمینه به‌طور معمول هموار هستند و اطلاعات مفید کمتری دارند درحالی‌که لایه‌ی حاوی اطلاعات لبه متن نشان داده شده در شکل ۴-۲(ب) شامل عمده اطلاعات مربوط به شکل متن می‌باشد. بنابراین برای افزایش ابعاد سه لایه مطابق شکل ۴-۷ روش‌های مختلفی در نظر گرفته می‌شود. با توجه به اهمیت لایه شامل اطلاعات لبه متن، این لایه را با تأکید بر تقویت لبه‌ها بزرگ می‌کنیم. از طرف دیگر به‌دلیل اینکه لایه‌های پیش‌زمینه و پس‌زمینه از اهمیت اطلاعاتی کمی در رابطه با ویژگی‌های متن برخوردار هستند، از روش درون‌یابی ساده دومکعبی برای بزرگ‌نمایی آنها استفاده می‌کنیم. نتایج حاصل از این روش بزرگ‌نمایی برای تصاویر پیش‌زمینه و

پس‌زمینه وضوح بالای F^H و B^H قابل قبول است. چون درون‌یابی دومکعبی لبه‌های تصویر متنی را مات می‌کند بنابراین لایه‌ی حاوی اطلاعات لبه متن را قبل از درون‌یابی ابتدا بهبود می‌دهیم. برای بهبود لبه‌های تصویر متنی از فیلتر تیگر[1] استفاده می‌کنیم. فیلتر تیگر یک فیلتر غیرخطی است که ما را قادر می‌سازد تا لبه‌های تصویر متن را تقویت و نویز را حذف کنیم. پس از بهبود لایه حاوی اطلاعات لبه متن، به‌منظور بزرگ‌نمایی این لایه از درون‌یابی ساده دومکعبی استفاده می‌کنیم نمونه‌های لایه بهبود داده شده را افزایش می‌دهیم. افزایش نمونه‌های لایه لبه متن می‌تواند به صورت زیر انجام شود:

$$\alpha^H = Bicubic(\alpha^L + \beta Teager(\alpha^L)) \qquad (۷-۴)$$

در رابطه (۴-۷) β ضریب وزن‌دهی برای کنترل تیزی لبه‌ها می‌باشد که معمولا برابر ۱ قرار داده می‌شود. $Teager()$ فیلتر تیگر و $Bicubic()$ درون‌یابی دومکعبی را نشان می‌دهد. پس از آنکه لایه‌های وضوح بالای α^H، F^H و B^H به‌دست آمدند، تصویر وضوح بالای نهایی با ترکیب این لایه‌ها و به‌صورت رابطه (۴-۸) بدست می‌آید [۷۸]:

$$I^H = \alpha^H F^H + (1-\alpha^H) B^H \qquad (۸-۴)$$

مرحله سوم از روش فرا تفکیک‌پذیری یعنی ترکیب لایه‌های وضوح بالای بدست آمده در شکل ۴-۸ نشان داده شده است.

نمونه‌ای از اجرای روش پیشنهادی روی تصویری با درجه‌ی تفکیک مکانی بالا برای روش‌های فشرده‌سازی مختلف در شکل‌های ۴-۹، ۴-۱۰، ۴-۱۱ و ۴-۱۲ نشان داده شده است. همان‌طور که از این شکل‌ها دیده می‌شود، روش پیشنهادی توانسته است اثرات مخرب ناشی از روش‌های فشرده‌سازی JPEG، JPEG2000 و کدگذار SPIHT را به خوبی حذف کند و تصویر را به وضوح اولیه نزدیک کند.

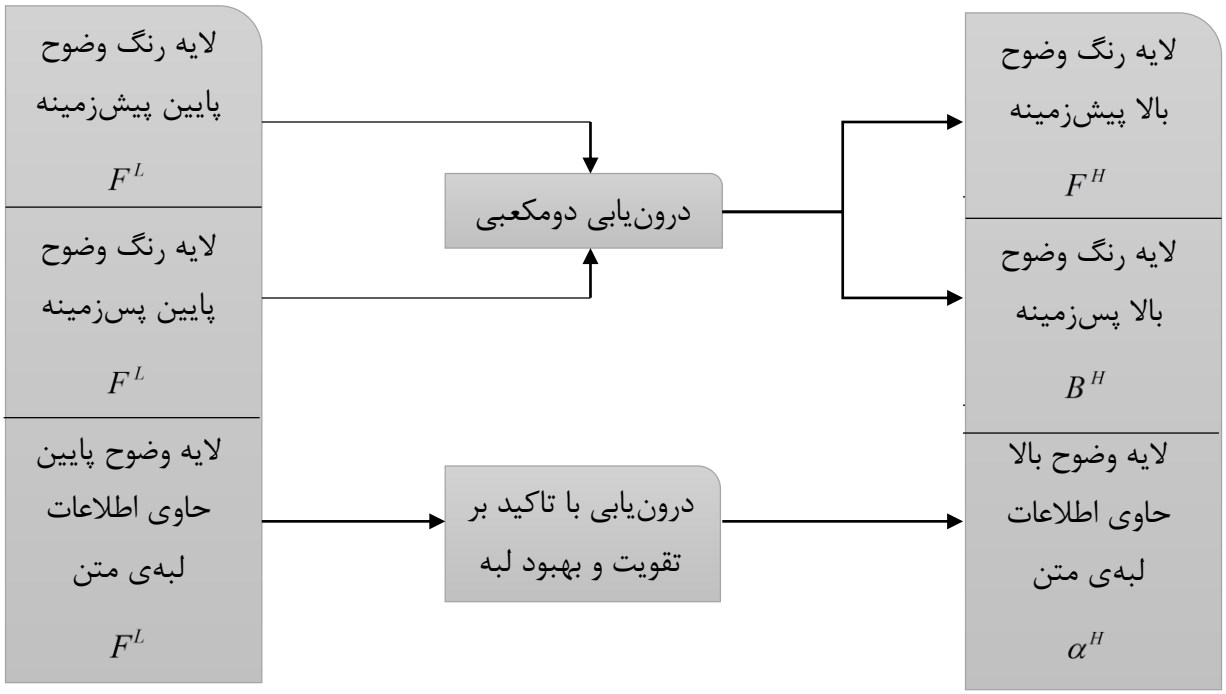
[1] Teager Filter

مرحله‌ی دوم : افزایش نمونه‌های سه لایه وضوح پایین بصورت مجزا

F^H : لایه رنگ وضوح بالا پیش‌زمینه

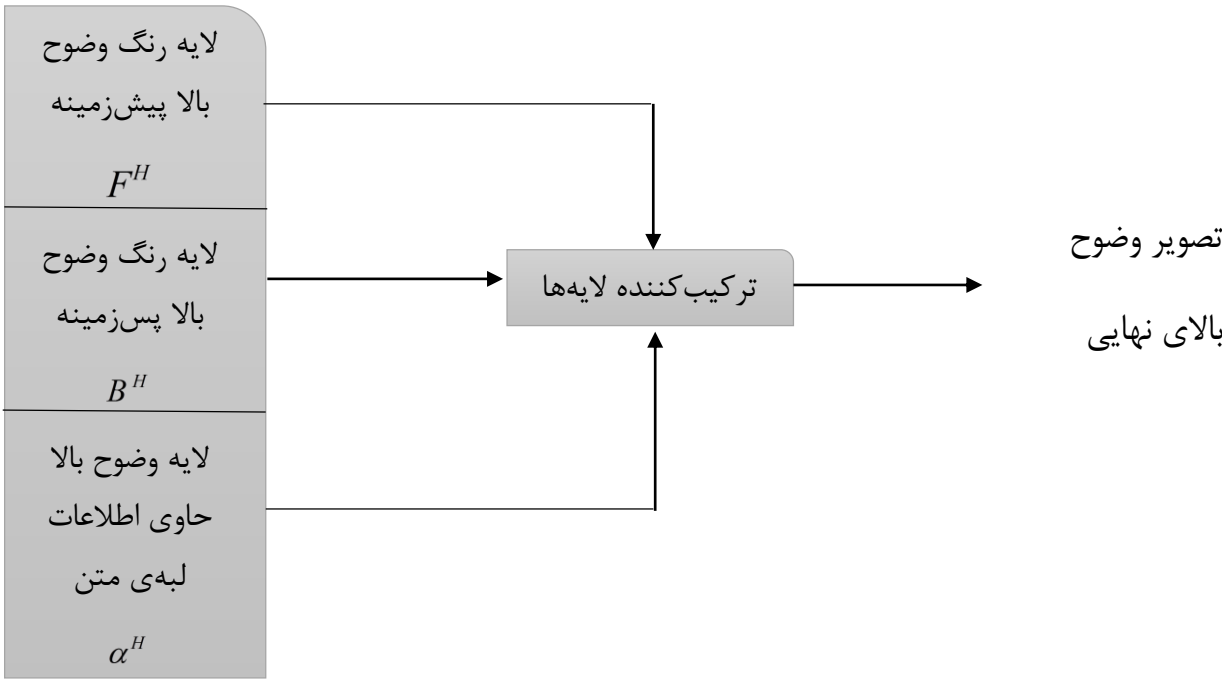
B^H : لایه رنگ وضوح بالا پس‌زمینه

α^H : لایه وضوح بالا حاوی اطلاعات لبه‌ی متن

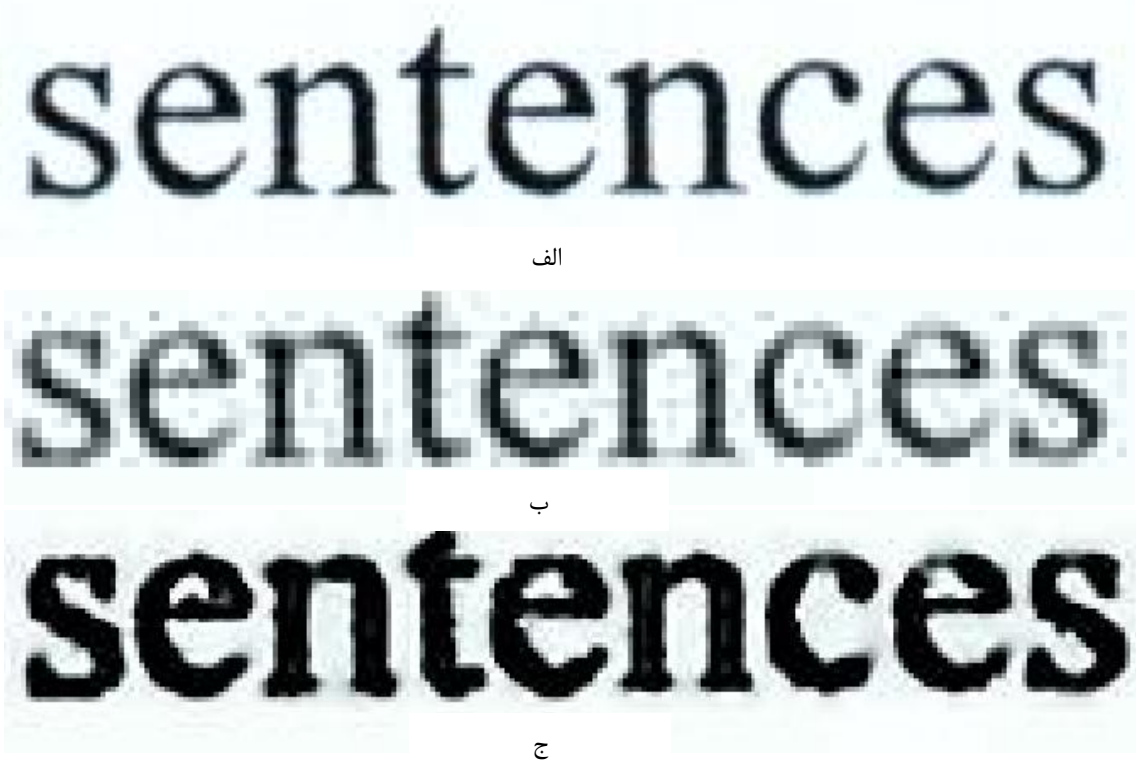


شکل ۴-۷ مرحله دوم روش فرا تفکیک‌پذیری مبتنی بر درون‌یابی

مرحله‌ی سوم : ترکیب لایه‌های وضوح بالا



شکل ۴-۸ مرحله سوم روش فرا تفکیک‌پذیری مبتنی بر درون‌یابی

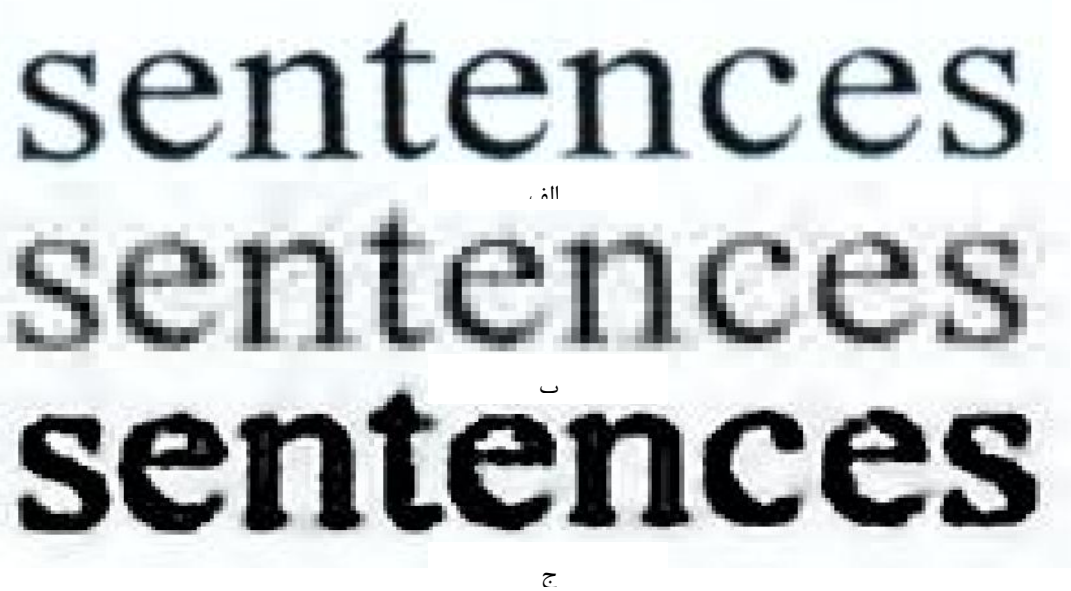


الف

ب

ج

شکل ۴-۹ ترکیب روش پیشنهادی با JPEG الف) تصویر اصلی ب) تصویر فشرده‌شده با نرخ بیت ۰/۰۹ ج) روش پیشنهادی

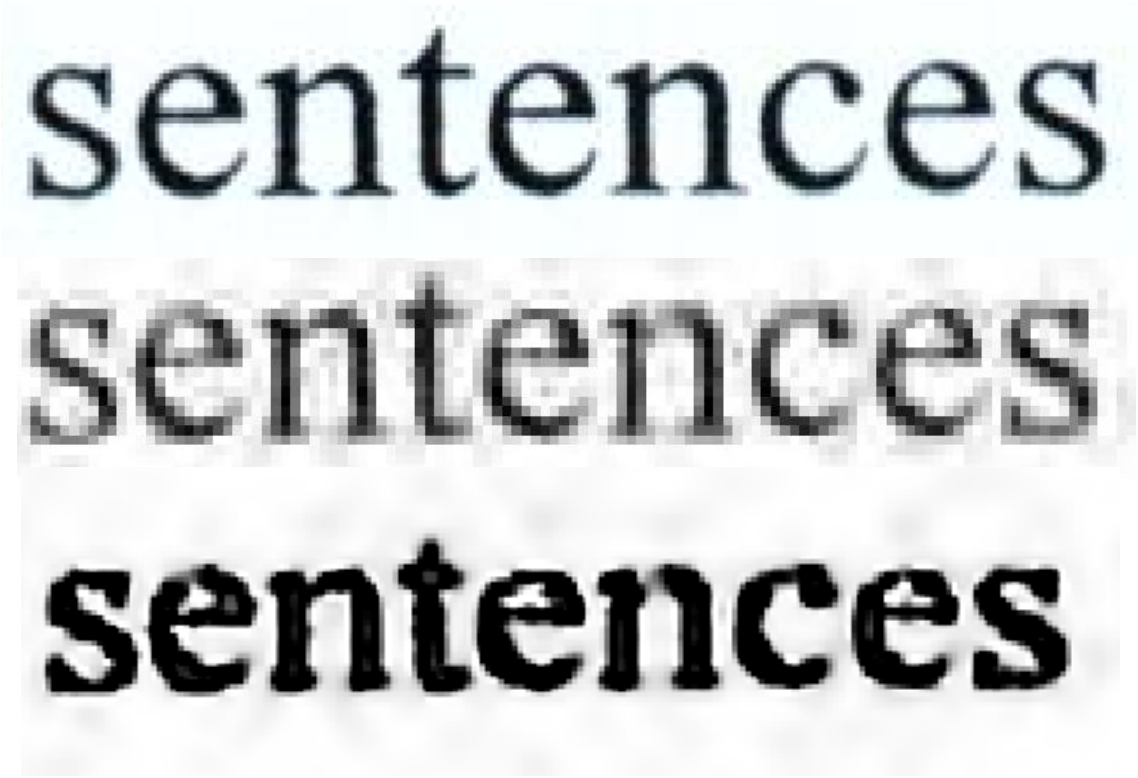


الف

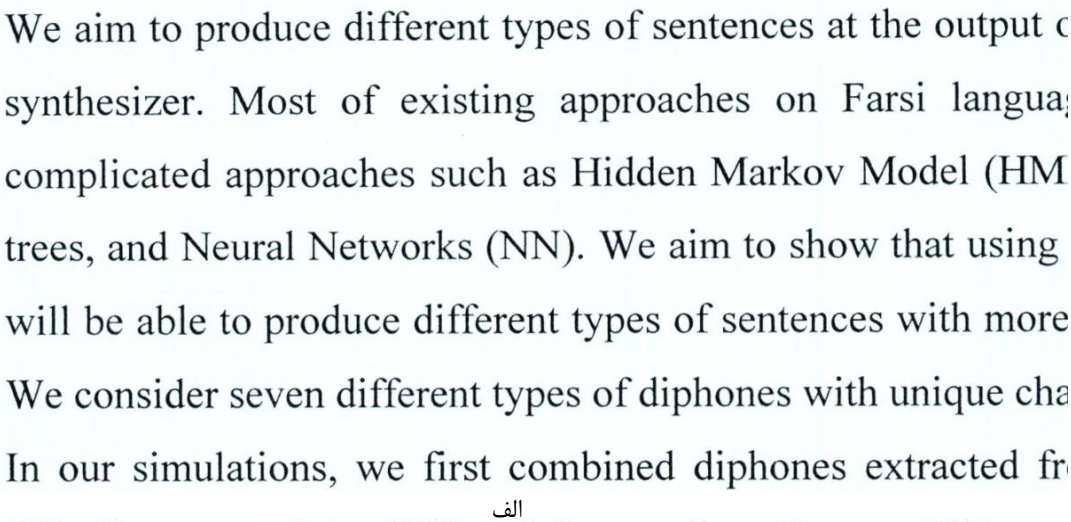
ب

ج

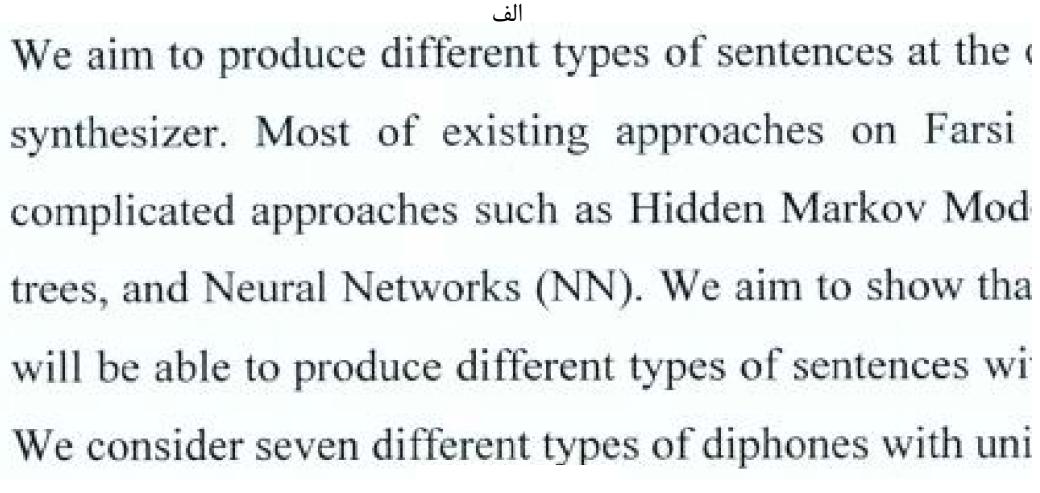
شکل ۴-۱۰ ترکیب روش پیشنهادی با JPEG2000 الف) تصویر اصلی ب) تصویر فشرده‌شده با نرخ بیت ۰/۰۹ ج) روش پیشنهادی



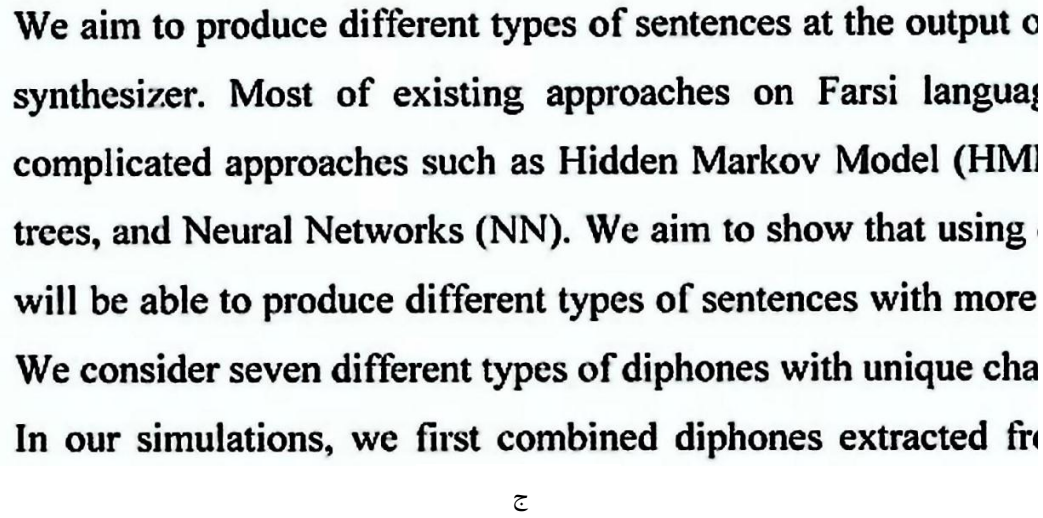
شکل ۴-۱۱ ترکیب روش پیشنهادی با SPIHT الف) تصویر اصلی ب) تصویر فشرده‌شده با نرخ بیت ۰/۰۹ ج) روش پیشنهادی



الف

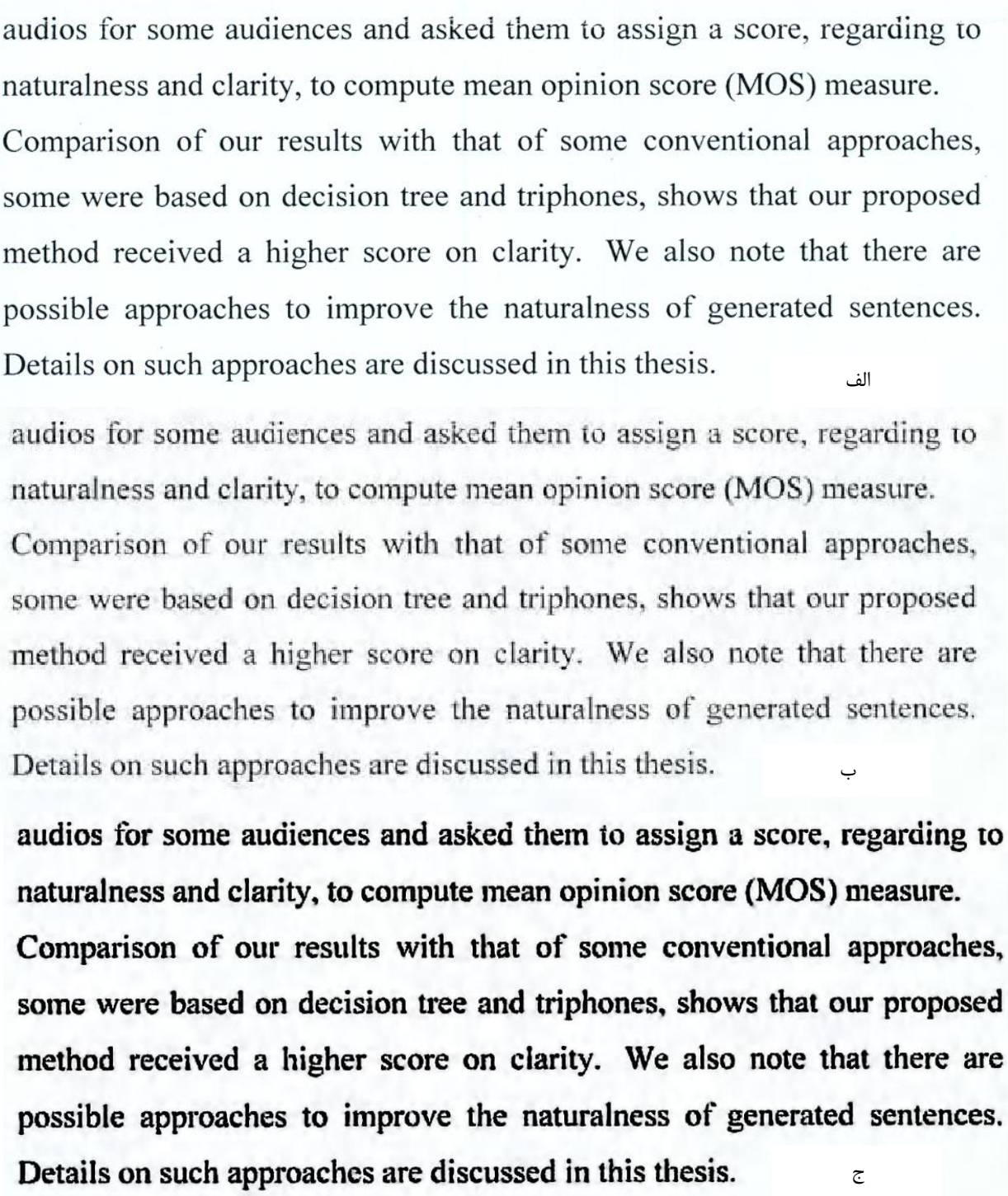


ب

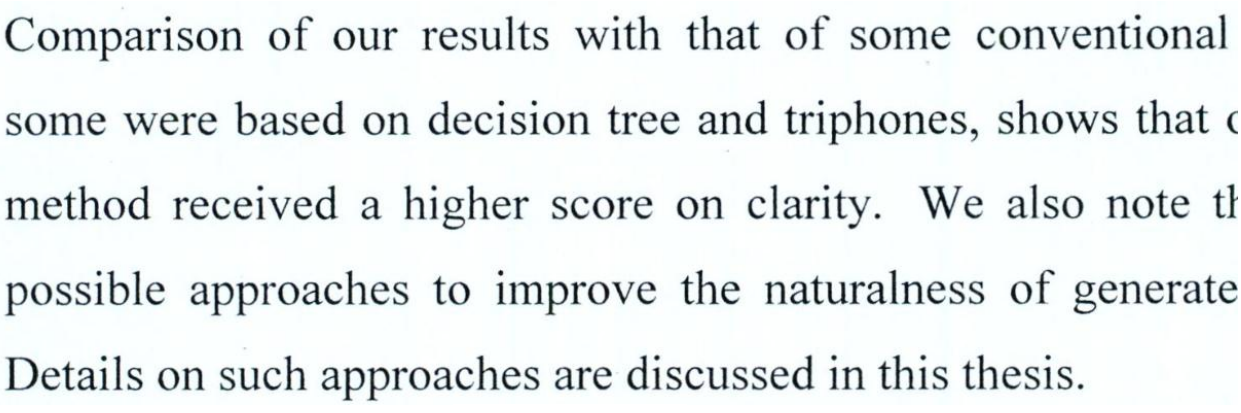


ج

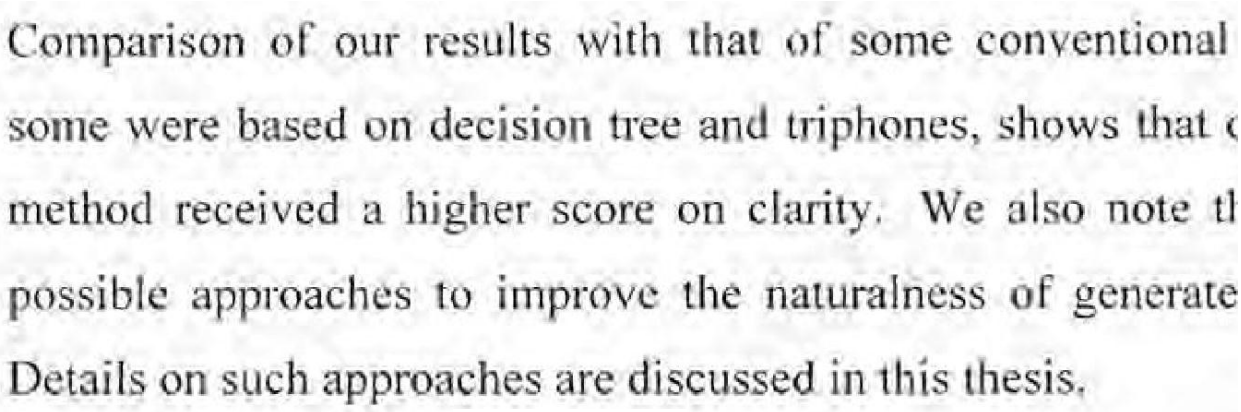
شکل ۴-۱۲ ترکیب روش پیشنهادی با JPEG الف) تصویر ۳۰۰dpi ب) تصویر فشرده‌شده با نرخ بیت ۰/۱ ب) روش پیشنهادی



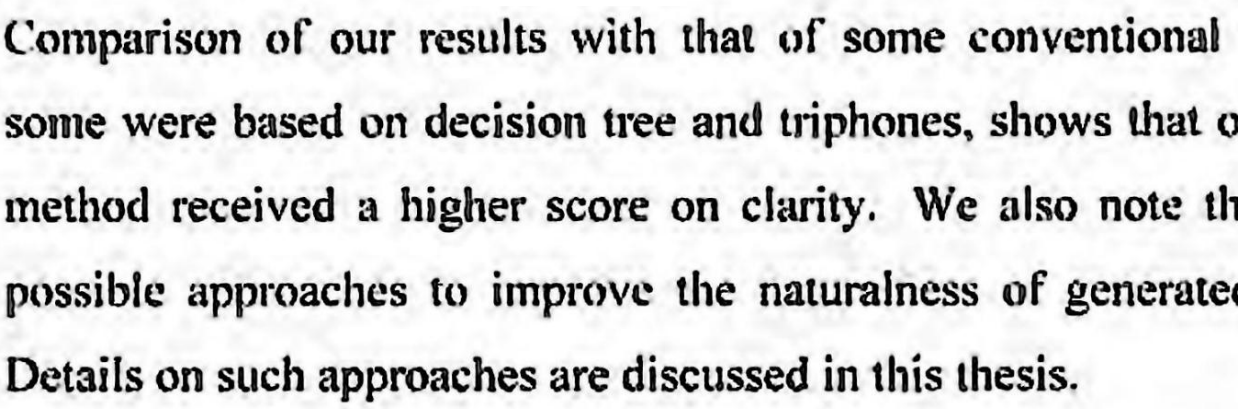
شکل ۴-۱۳ ترکیب روش پیشنهادی با JPEG2000 الف) تصویر dpi۶۰۰ ب) تصویر فشرده‌شده با نرخ بیت ۰/۰۱۵ ب) روش پیشنهادی



الف



ب



ج

شکل ۴-۱۴ ترکیب روش پیشنهادی با SPIHT الف) تصویر ۳۰۰dpi ب) تصویر فشرده‌شده با نرخ بیت ۰/۱ ب) روش پیشنهادی

۴-۵ فرا تفکیک‌پذیری مبتنی بر آموزش

در روش‌های مبتنی بر آموزش همان‌طور که در فصل‌های قبل گفته شد سعی بر آن است که رابطه بین وصله‌های وضوح بالا و پایین یاد گرفته شود. با یاد گرفتن این رابطه، می‌توان رابطه را به تصویر با وضوح پایین اعمال کرد و تصویر با وضوح بالا را داشت.

برای این روش ابتدا پایگاه داده آموزشی شامل داده‌های آموزشی وضوح بالا و وضوح پایین ایجاد شد. برای ایجاد داده‌های وضوح پایین، ابتدا تصویر با ضریب مشخصی کاهش ابعاد داده شد سپس به یک روش فشرده‌سازی داده شد و نتیجه حاصل از فشزده‌ساز بعنوان داده‌ی وضوح پایین در نظر گرفته شد. با ایجاد پایگاه داده نوبت به ساخت واژه‌نامه می‌رسد. برای ساخت واژه‌نامه از ویژگی استفاده شده است. ویژگی‌ها جزئیات فرکانس بالای تصاویر بودند که با استفاده از ۴ فیلتر بالاگذر بدست آمدند. این فیلترها در رابطه‌ی (۴-۹) نشان داده شده‌اند.

$$f_1 = [-1, 0, 1], \qquad f_2 = f_1^T$$
$$f_3 = [1, 0, -2, 0, 1], \qquad f_4 = f_3^T \tag{۴-۹}$$

با ساخت واژه‌نامه و آموزش آنها، ضرایب بازسازی برای وصله‌های وضوح بالا را بدست آورده و با ضرب در واژه‌نامه وضوح بالا، وصله‌های وضوح بالا ساخته می‌شود.

برای این روش از کدهای نوشته شده برای مقالات [۸۲-۸۱] استفاده شده‌است. پس از اعمال روی تصویر ورودی جواب خوبی بدست نیامد. با این روش اثرات ناشی از فشرده‌سازی بهبود داده نشد و تصویر حاصل بهبود خوبی را نداشت.

فصل پنجم:

نتایج تجربی

۵-۱ معرفی

در ابتدا پایگاه داده، نرم‌افزارهای استفاده شده و معیارهای ارزیابی معرفی شده‌اند. در این فصل نتایج حاصل از اجرای روش پیشنهادی روی دو گروه از تصاویر متنی واقعی با درجه‌ی تفکیک مکانی متفاوت نشان داده شده است. با روبش تصاویر متنی واقعی دو گروه از تصاویر شامل تصاویر روبش شده با درجه‌ی تفکیک ۶۰۰dpi و ۳۰۰dpi جمع‌آوری شده‌است. از روش دومکعبی برای کاهش ابعاد تصویر با درجه‌ی تفکیک مکانی بالا استفاده شده‌است. در ادامه به‌منظور فشرده‌سازی تصاویر از نرم‌افزارهای Advanced JPEG Compressor و JPEG2000 Compressor و کد نوشته شده در نرم‌افزار متلب[1] برای روش SPIHT استفاده شده‌است. پس از اعمال روش پیشنهادی باید معیاری وجود داشته باشد که میزان کارایی روش ارزیابی شود. برای این منظور معیارهای بازشناسی متن[2] (OCR)، متوسط امتیاز نظرسنجی[3] (MOS) و PSNR مورد استفاده قرار گرفته‌اند.

- معیار OCR

برای OCR تصویر بازسازی شده با روش پیشنهادی، از نرم‌افزار ABBY Fine Reader استفاده شده‌است. برای بدست آوردن صحت OCR، تعداد کاراکترهایی که نرم‌افزار معرفی شده به درستی تشخیص داده است به تعداد کل کاراکترها تقسیم شده‌اند.

- معیار MOS

برای معیار MOS از ۲۰ نفر خواسته شد که به تصاویری که به آنها نشان داده می‌شود از نظر خوانایی و کیفیت از ۱ تا ۵ نمره دهند. به این صورت که ۵ بهترین نمره‌ای است که می‌توان به یک تصویر با کیفیت عالی داد و ۱ ضعیف‌ترین نمره‌ای است که می‌توان به یک تصویر داد. در نهایت برای هر تصویر مجموع ۲۰

[1] Matlab
[2] Optical Character Recognition
[3] Mean Opinion Score

نمره‌ای که به آن اختصاص داده شده است را میانگین‌گیری کرده و نمره نهایی برای تصویر مورد نظر گزارش شده‌است.

- معیار PSNR

این معیار میزان نزدیک بودن تصویر بازسازی شده به تصویر اصلی را نشان می‌دهد. برای محاسبه PSNR از رابطه زیر استفاده شده‌است.

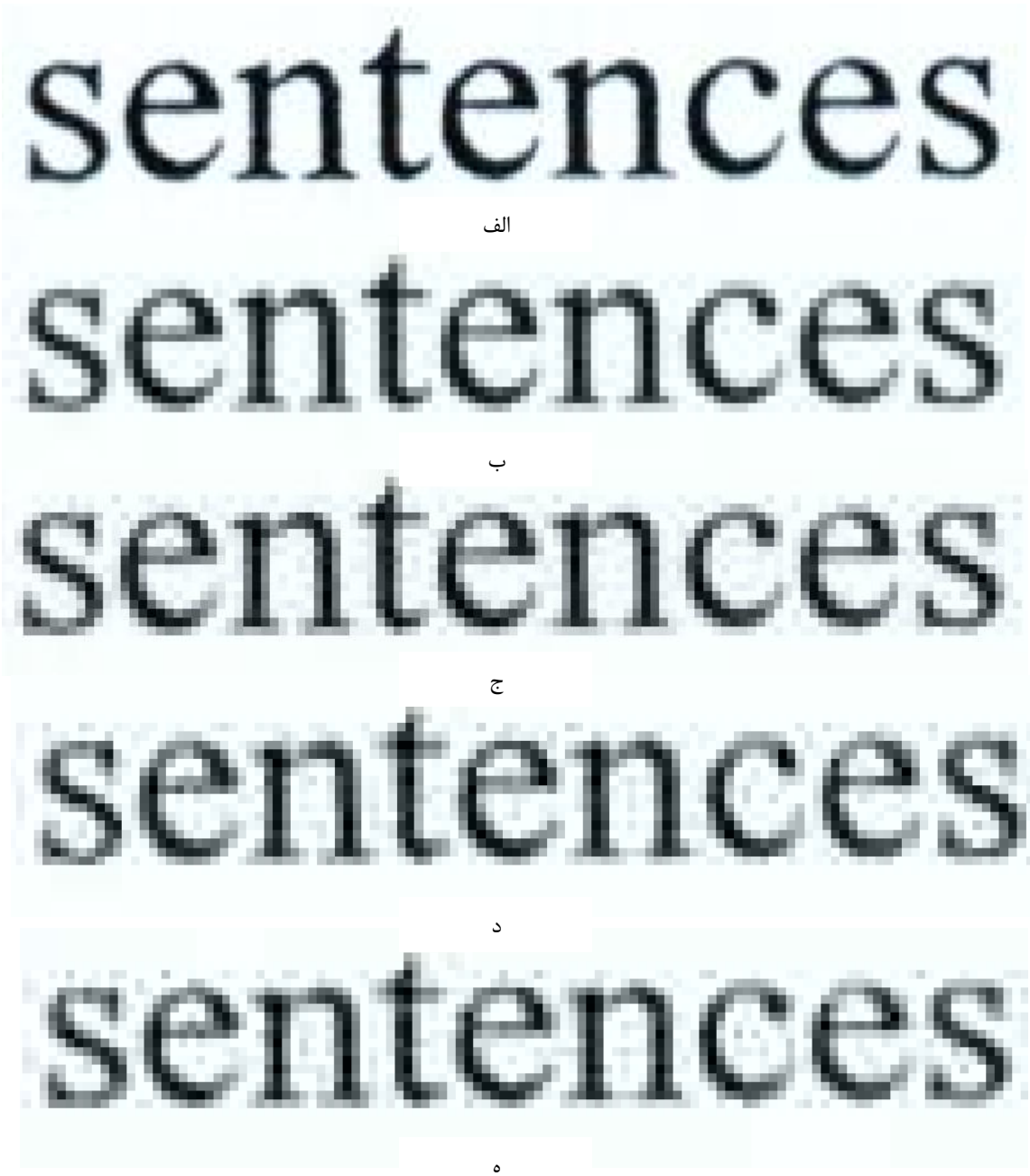
$$PSNR = 10\log_{10}\left(\frac{255^2 mn}{\left\|\hat{X} - I\right\|^2}\right) \qquad (۱-۵)$$

در رابطه (۵-۱) I و $\hat{X}$ به‌ترتیب تصویر اصلی و تصویر بازسازی شده و mn تعداد پیکسل‌های تصویر اصلی هستند.

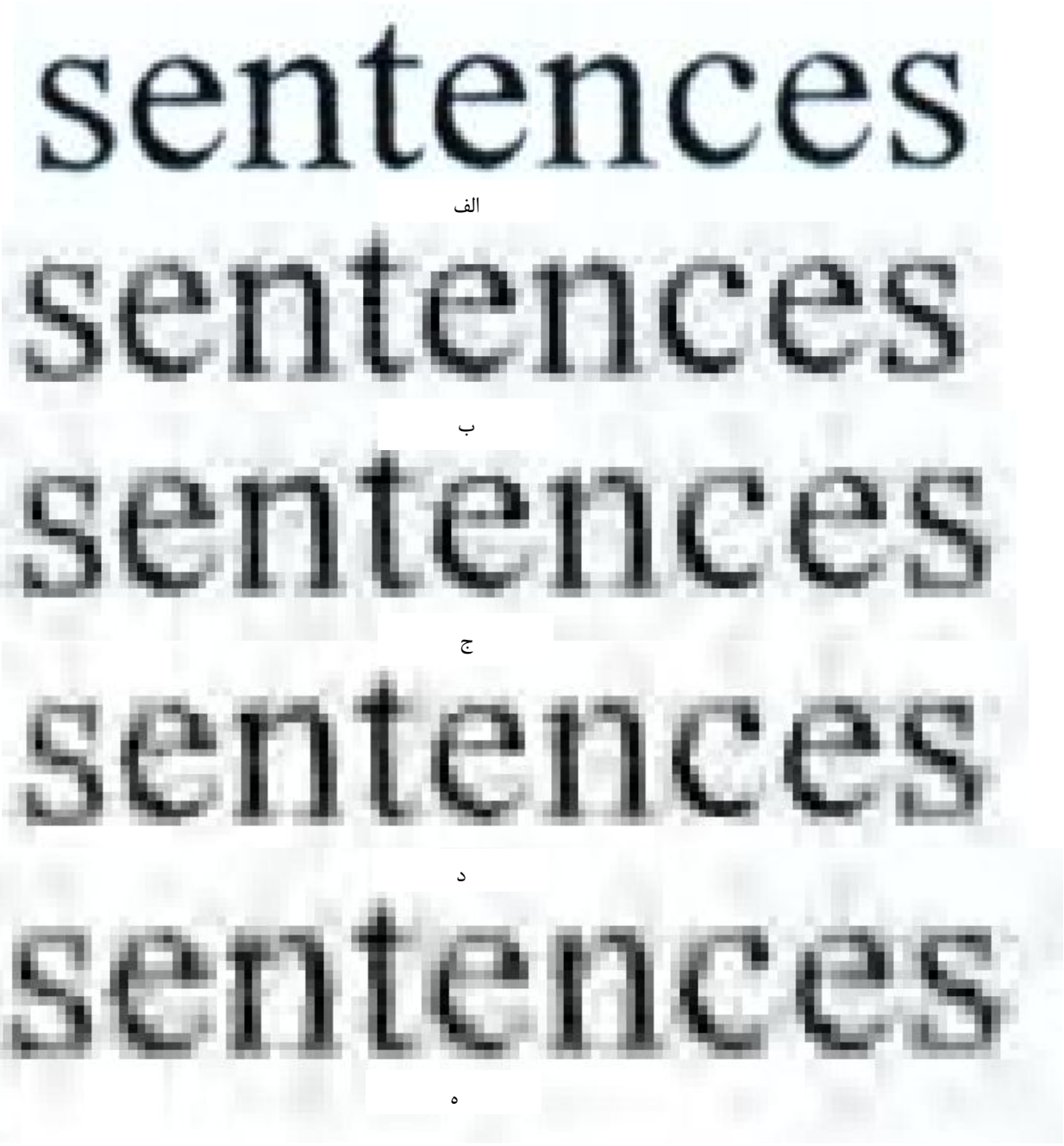
از آنجایی که تصاویر متنی یا توسط انسان مورد مطالعه قرار می‌گیرد و یا توسط دستگاه‌های تشخیص متن مورد بررسی قرار می‌گیرد ما سعی کردیم که روش بتواند در این دو زمینه جواب خوبی داشته باشد و معیار PSNR به اندازه‌ی این دو معیار در روش در نظر گرفته شده اهمیت ندارد.

در ادامه اثر روش‌های فشرده‌سازی در نرخ بیت‌های متفاوت روی یک تصویر با درجه‌ی تفکیک مکانی بالا در شکل‌های ۵-۱، ۵-۲ و ۵-۳ نشان داده شده است.

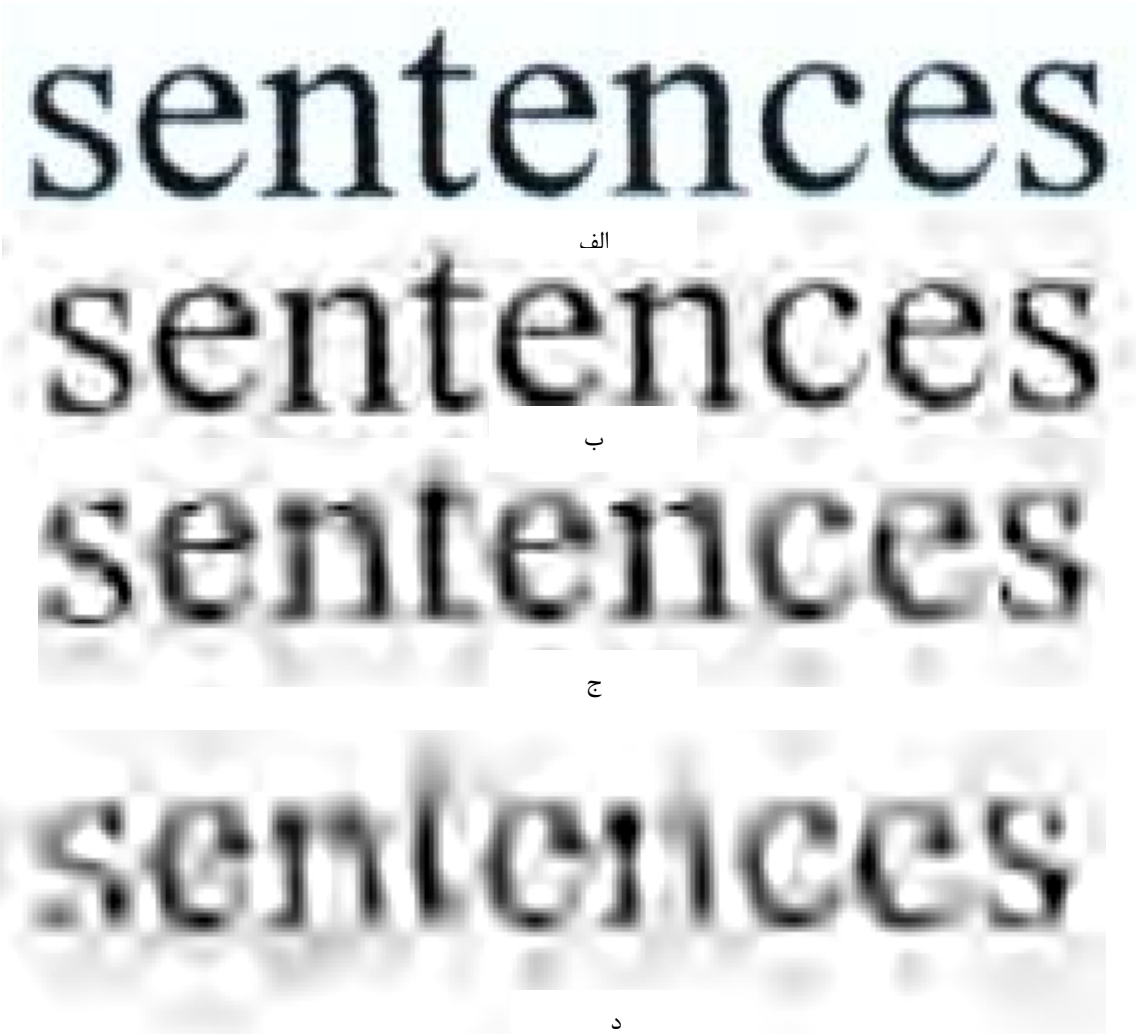
در شکل ۵-۱ اثر فشرده‌سازی با روش JPEG نشان داده شده است. مشخص است که با افزایش میزان فشرده‌سازی، کیفیت تصویر متنی خراب می‌شود. همان‌طور که از شکل مشخص است روش JPEG در نرخ بیت‌های کم در تصویر متنی اثرات بلوکی ایجاد می‌کند. در شکل ۵-۲ اثرات ناشی از روش فشرده-سازی JPEG2000 نشان داده شده است. همان‌طور که از نتایج این شکل مشخص است روش فشرده‌سازی JPEG2000 در نرخ بیت‌های پایین در تصویر اثر نشتی ایجاد می‌کند. در شکل ۵-۳ نیز اثرات مخرب ناشی از روش کدگذاری SPIHT نشان داده شده است. این روش نیز مانند روش JPEG2000 اثرات نشتی روی تصویر ایجاد می‌کند.



شکل ۵-۱ اثرات روش فشرده‌سازی JPEG الف) تصویر اصلی ب) فشرده‌سازی با نرخ بیت ۰/۲ ج) فشرده‌سازی با نرخ بیت ۰/۱۵ د) فشرده‌سازی با نرخ بیت ۰/۱ ه) فشرده‌سازی با نرخ بیت ۰/۰۹



شکل ۵-۲ اثرات روش فشرده‌سازی JPEG2000 الف) تصویر اصلی ب) فشرده‌سازی با نرخ بیت ۰/۰۷ ج) فشرده‌سازی با نرخ بیت ۰/۰۶ د) فشرده‌سازی با نرخ بیت ۰/۰۵ ه) فشرده‌سازی با نرخ بیت ۰/۰۴



الف

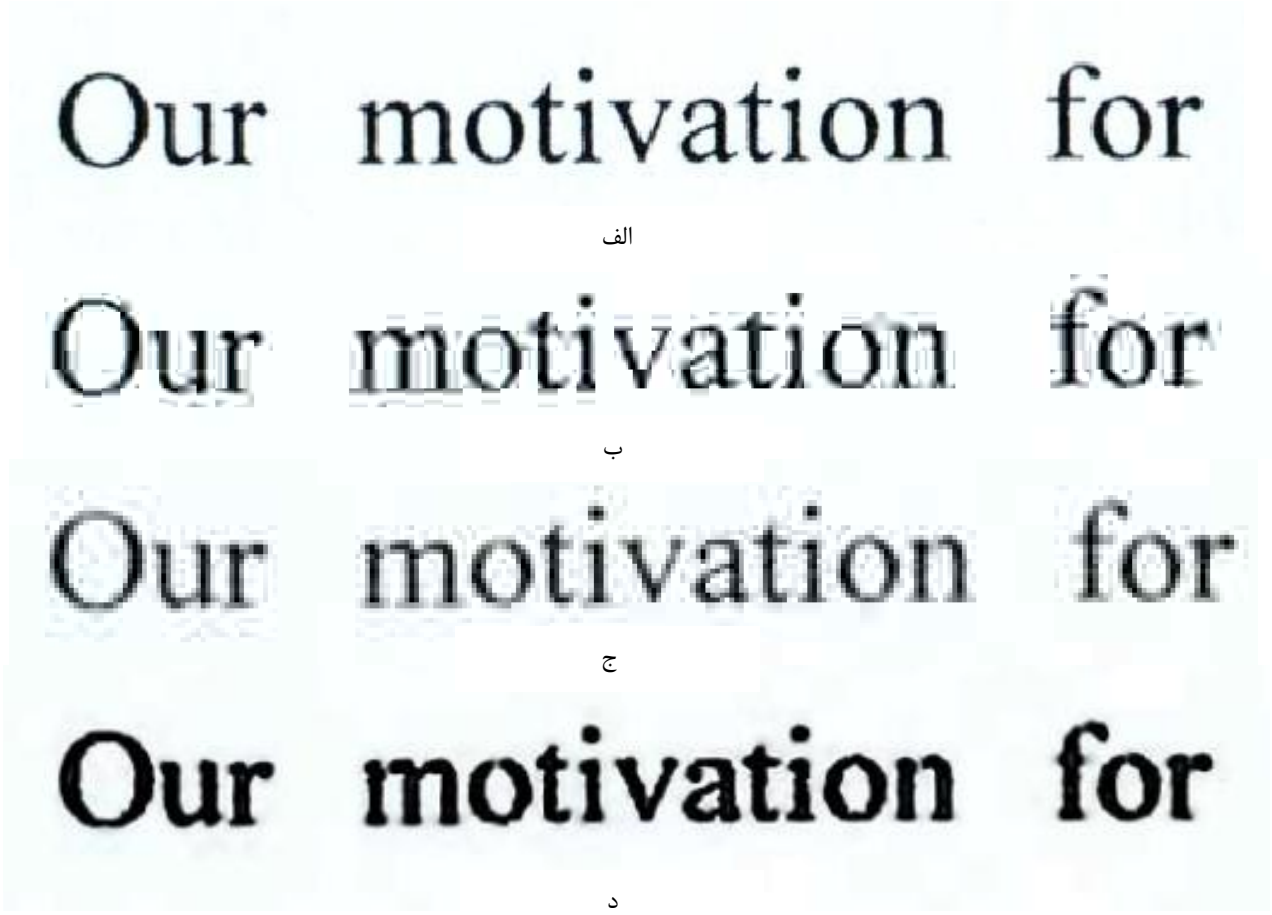
ب

ج

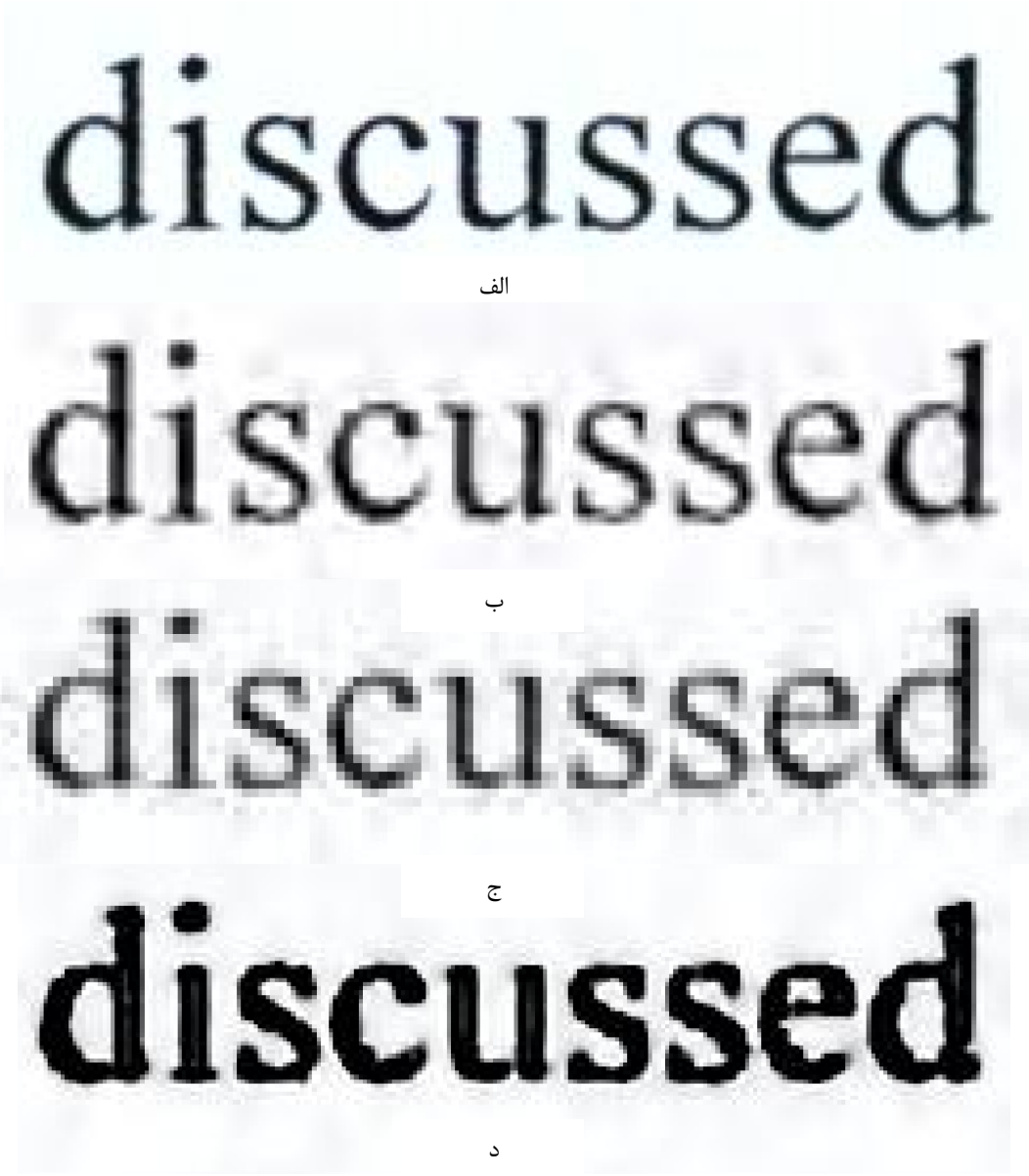
د

شکل ۵-۳ اثرات روش فشرده‌سازی SPIHT الف) تصویر اصلی ب) فشرده‌سازی با نرخ بیت ۰/۱ ج) فشرده‌سازی با نرخ بیت ۰/۰۸ د) فشرده‌سازی با نرخ بیت ۰/۰۴

در ادامه برای دیدن اثر کاهش ابعاد روی تصاویر فشرده‌شده چند تصویر که پس از کاهش ابعاد و بدون کاهش ابعاد فشرده‌سازی روی آنها انجام گرفته نشان داده می‌شود. شکل‌های ۵-۴، ۵-۵ و ۵-۶ این تصاویر را نشان می‌دهد.



شکل ۵-۴ مقایسه فشرده‌سازی قبل و بعد از کاهش ابعاد الف) تصویر اصلی ب) تصویر فشرده‌شده با JPEG بدون کاهش ابعاد ج) تصویر فشرده‌شده با کاهش ابعاد د) تصویر بازسازی شده



الف

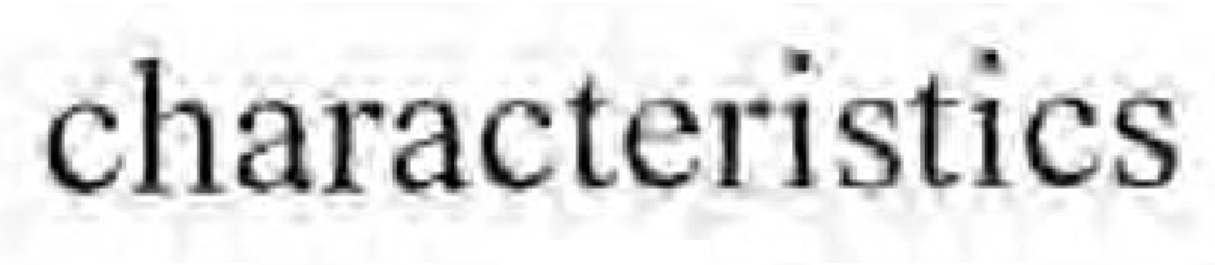
ب

ج

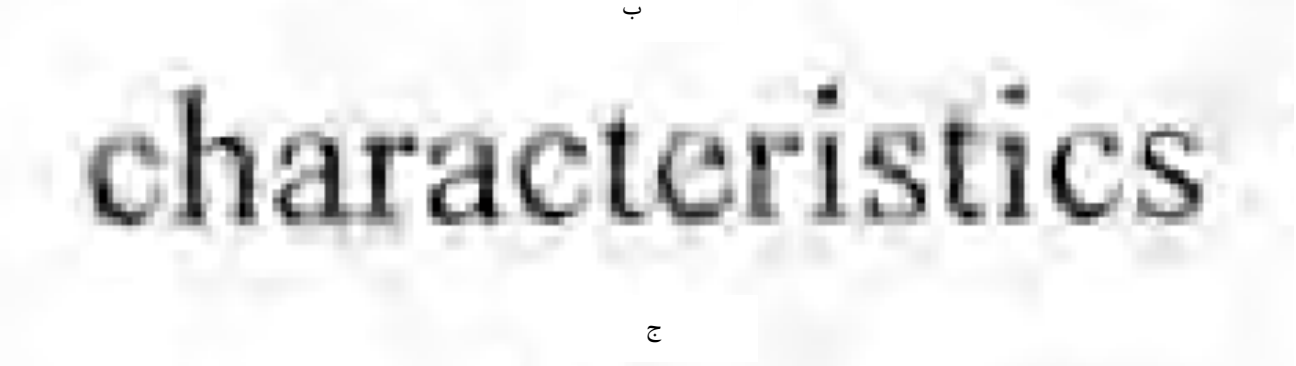
د

شکل ۵-۵ مقایسه فشرده‌سازی قبل و بعد از کاهش ابعاد الف) تصویر اصلی ب) تصویر فشرده‌شده با JPEG2000 بدون کاهش ابعاد ج) تصویر فشرده‌شده با کاهش ابعاد د) تصویر بازسازی شده

الف



ب



ج

د

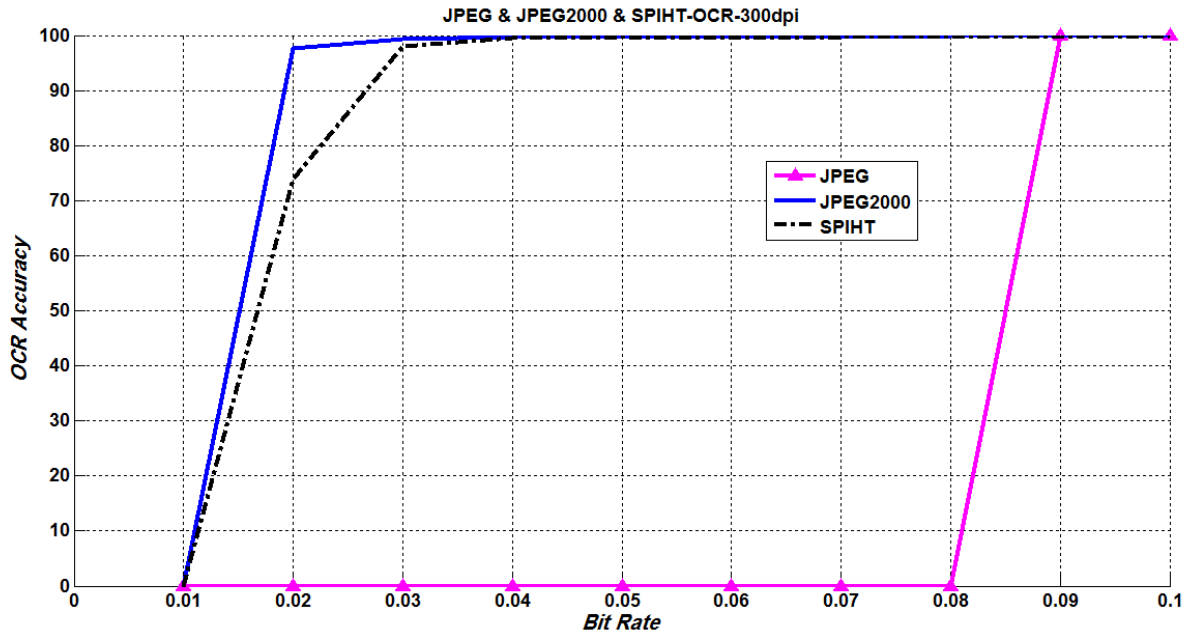
شکل ۵-۶ مقایسه فشرده‌سازی قبل و بعد از کاهش ابعاد الف) تصویر اصلی ب) تصویر فشرده‌شده با SPIHT بدون کاهش ابعاد ج) تصویر فشرده‌شده با کاهش ابعاد د) تصویر بازسازی شده

۵-۲ ارزیابی روش پیشنهادی براساس معیار OCR

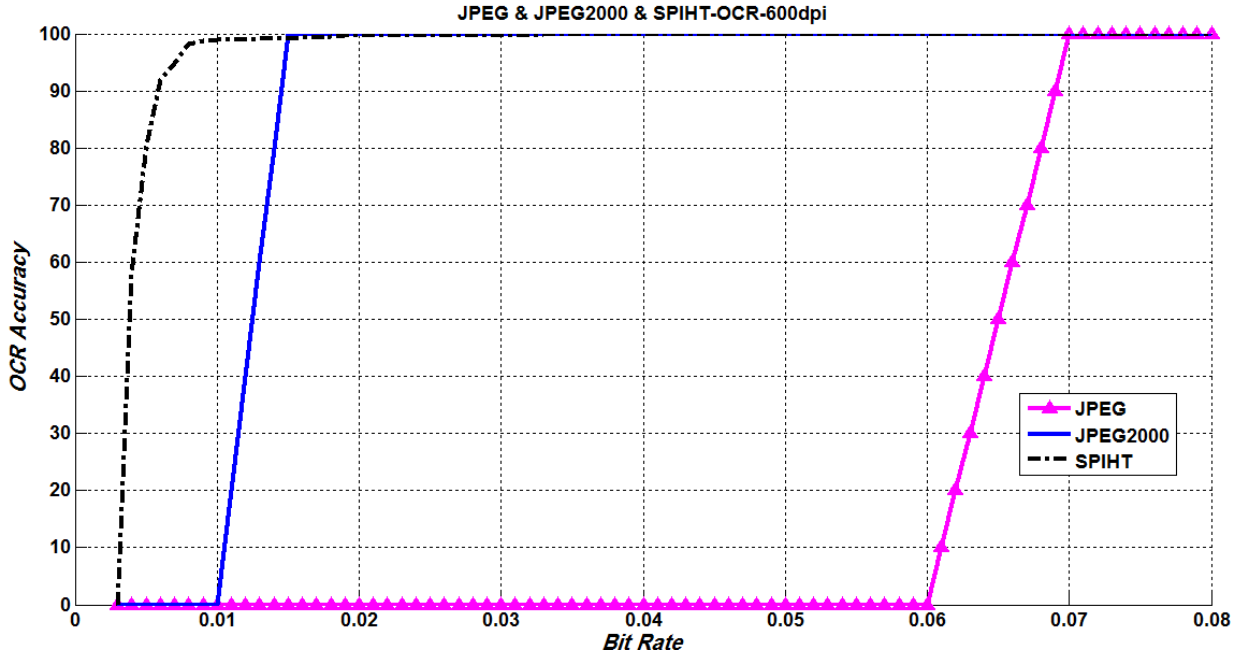
در ابتدا دو دسته تصاویر با درجه‌ی تفکیک مکانی متوسط و بالا (به‌ترتیب ۳۰۰ و ۶۰۰dpi) به‌طور مستقیم به روش‌های فشرده‌سازی JPEG، JPEG2000 و کدگذاری SPIHT داده شده‌اند و خروجی حاصل از هر روش براساس معیار OCR ارزیابی شده‌است. نتایج به‌دست آمده به‌ترتیب در شکل‌های ۵-۷ و ۵-۸ نشان داده شده‌است.

همان‌طور که در شکل ۵-۷ دیده می‌شود، روش فشرده‌سازی JPEG برای تصویر با درجه‌ی تفکیک مکانی متوسط تا نرخ بیت ۰/۰۹ جواب مطلوبی دارد. برای نرخ بیت‌های بالاتر از ۰/۰۹ نیز جواب ۱۰۰ درصد برای صحت OCR به‌دست می‌آید. اما با این روش فشرده‌سازی برای تصویر داده شده رسیدن به نرخ بیت کمتر از ۰/۰۹ امکان پذیر نبوده است. روش فشرده‌سازی JPEG2000 نیز تا نرخ بیت ۰۲/ جواب مطلوبی داشته است. و برای نرخ بیت‌های بالاتر از ۰/۰۲ نیز ۱۰۰ درصد صحت OCR را دارد. برای این روش رسیدن به نرخ بیت کمتر از ۰/۰۲ امکان پذیر نبوده است. اما روش SPIHT برخلاف دو روش قبل محدودیتی روی نرخ بیت آن وجود ندارد و تا هر نرخ بیتی که بخواهیم می‌توانیم تصویر ورودی را فشرده کنیم. در این روش برای نرخ بیت‌های کمتر از ۰/۰۲ از نظر معیار OCR جواب مطلوبی به‌دست نمی‌آید. به‌طور مثال برای نرخ بیت ۰/۰۱۵، صحت OCR، ۱۳ درصد بود که صفر در نظر گرفته شده است.

در رسم نمودارها برای نرخ بیت‌هایی که رسیدن به آنها امکان پذیر نبوده است داده‌های مصنوعی صفر ایجاد شده‌است. به‌طور مثال برای نرخ بیت‌های کمتر از ۰/۰۹ در روش فشرده‌سازی JPEG یا نرخ بیت‌های کمتر از ۰/۰۲ در روش فشرده‌سازی JPEG2000 داده‌های مصنوعی صفر برای جواب OCR استفاده شده است.



شکل ۵-۷ مقایسه روش‌های فشرده‌سازی از نظر معیار OCR برای تصویر با درجه‌ی تفکیک مکانی ۳۰۰dpi



شکل ۵-۸ مقایسه روش‌های فشرده‌سازی از نظر معیار OCR برای تصویر با درجه‌ی تفکیک مکانی ۶۰۰dpi

در شکل ۵-۸ نتایج حاصل از اعمال روش‌های فشرده‌سازی مختلف روی تصویری با درجه تفکیک مکانی بالا از نظر معیار OCR مقایسه شده است. همان‌طور که از این شکل مشخص است روش فشرده‌سازی JPEG برای تصویر ورودی توانسته است تا نرخ بیت ۰/۰۷ جواب مطلوبی را داشته باشد. برای نرخ بیت‌های

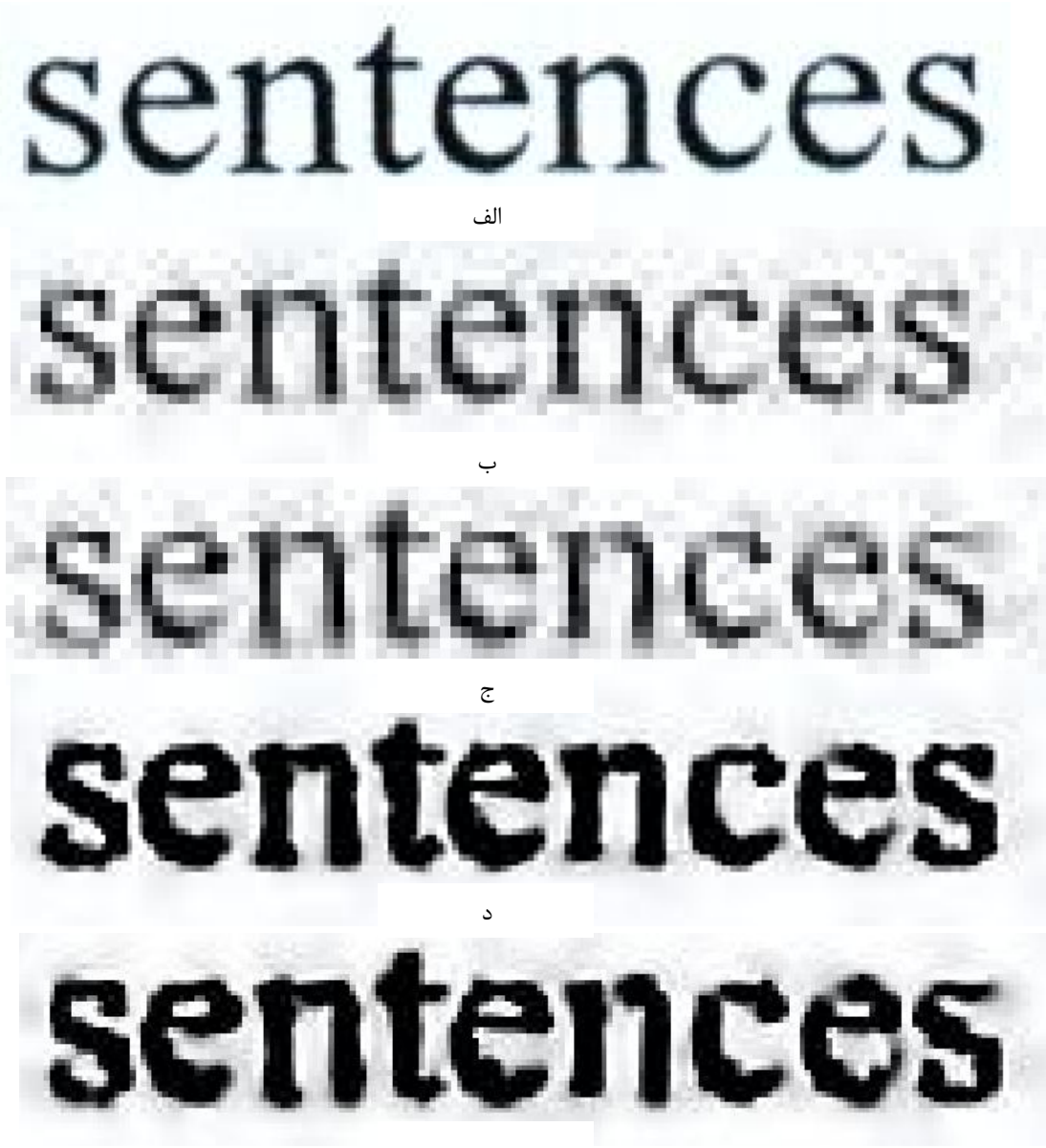
بیشتر نیز صحت OCR، ۱۰۰ درصد است. در این شکل نیز محدودیت روش JPEG برای رسیدن به نرخ بیت‌های پایین مشخص است. به‌طوریکه رسیدن به نرخ بیت‌های کمتر از ۰/۰۷ با این روش برای تصویر با درجه‌ی تفکیک مکانی بالای ورودی امکان پذیر نبوده است. روش فشرده‌سازی JPEG2000 تا نرخ بیت ۰/۰۱۵ جواب مطلوبی داشته است. برای نرخ بیت‌های بیشتر از ۰/۰۱۵ صحت OCR، ۱۰۰ درصد بوده است. در این روش نیز برای رسیدن به نرخ بیت‌های پایین محدودیت وجود داشته است به‌طوریکه رسیدن به نرخ بیت کمتر از ۰/۰۱۵ امکان پذیر نبوده است. در روش فشرده‌سازی SPIHT محدودیتی روی نرخ بیت وجود ندارد و با هر نرخ بیت دل‌خواه می‌توانسته‌ایم تصویر را فشرده کنیم. اما صحت معیار OCR برای نرخ بیت‌های کمتر از ۰/۰۰۶ خیلی مطلوب نیست. برای نرخ بیت‌های ۰/۰۰۵ و ۰/۰۰۶، صحت OCR به‌ترتیب ۸۰/۴۳ و ۹۱/۹۲ درصد به‌دست آمده است.

مشاهده شد که روش‌های JPEG و JPEG2000 برای فشرده‌سازی تصاویر متنی دارای محدودیت هستند. یعنی میزان فشرده‌سازی آنها روی یک تصویر محدوده مشخصی دارد و امکان فشرده‌سازی بیشر از یک میزان مشخص وجود ندارد. برای رفع این مشکل ایده‌ی کاهش ابعاد تصاویر متنی با درجه‌ی تفکیک مکانی بالا مطرح شده‌است. به‌این صورت که برای رسیدن به نرخ بیت‌های کمتر ابتدا ابعاد تصویر ورودی را کاهش داده سپس تصویر با یکی از روش‌های گفته شده، فشرده می‌شود. با استفاده از این روش می‌توان به نرخ بیت‌های کمتر برای ذخیره‌سازی تصویر با درجه‌ی تفکیک مکانی بالا رسید.

با ترکیب ایده‌ی کاهش ابعاد با روش‌های فشرده‌سازی توانستیم به نرخ بیت‌های کمتری برسیم. اما اثرات مخرب ناشی از فشرده‌سازی و کاهش ابعاد به همراه کاهش وضوح تصویر متنی سبب کم شدن خوانایی تصویر فشرده‌شده می‌شود. برای حل این مشکلات از روش فرا تفکیک‌پذیری در مرحله بازسازی استفاده شده‌است.

در شکل‌های ۵-۹ و ۵-۱۰ و ۵-۱۱ نتیجه ترکیب روش پیشنهادی به‌ترتیب با روش‌های فشرده-سازی JPEG، JPEG2000 و SPIHT نشان داده شده است. همان‌طور که از شکل ۵-۹ دیده می‌شود، روش پیشنهادی توانسته است اثرات بلوکی ناشی از روش فشرده‌سازی JPEG را تا حد امکان حذف کند. در شکل ۵-۱۰ نیز نتایج حاصل از ترکیب روش پیشنهادی با روش JPEG2000 نشان داده شده است. روش پیشنهادی در ترکیب با این روش فشرده‌سازی نیز نتایج قابل قبولی داشته است. شکل ۵-۱۱ نتایج حاصل از ترکیب روش پیشنهادی با روش فشرده‌سازی SPIHT را نشان می‌دهد. این روش برخلاف دو روش قبل محدودیتی در نرخ بیت ندارد اما در نرخ بیت‌های پایین خوانایی تصویر متنی را بشدت کم می‌کند. ترکیب روش پیشنهادی با SPIHT برای چند نرخ بیت مختلف در شکل ۵-۸ نشان داده شده است.

شکل ۵-۹ روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG الف) تصویر اصلی ب) فشرده‌سازی با نرخ بیت ۰/۰۹ ج) فشرده‌سازی با نرخ بیت ۰/۰۸ د) فشرده‌سازی با نرخ بیت ۰/۰۷ ه) روش پیشنهادی برای نرخ بیت ۰/۰۹ و) روش پیشنهادی برای نرخ بیت ۰/۰۸ ز) روش پیشنهادی برای نرخ بیت ۰/۰۷



الف

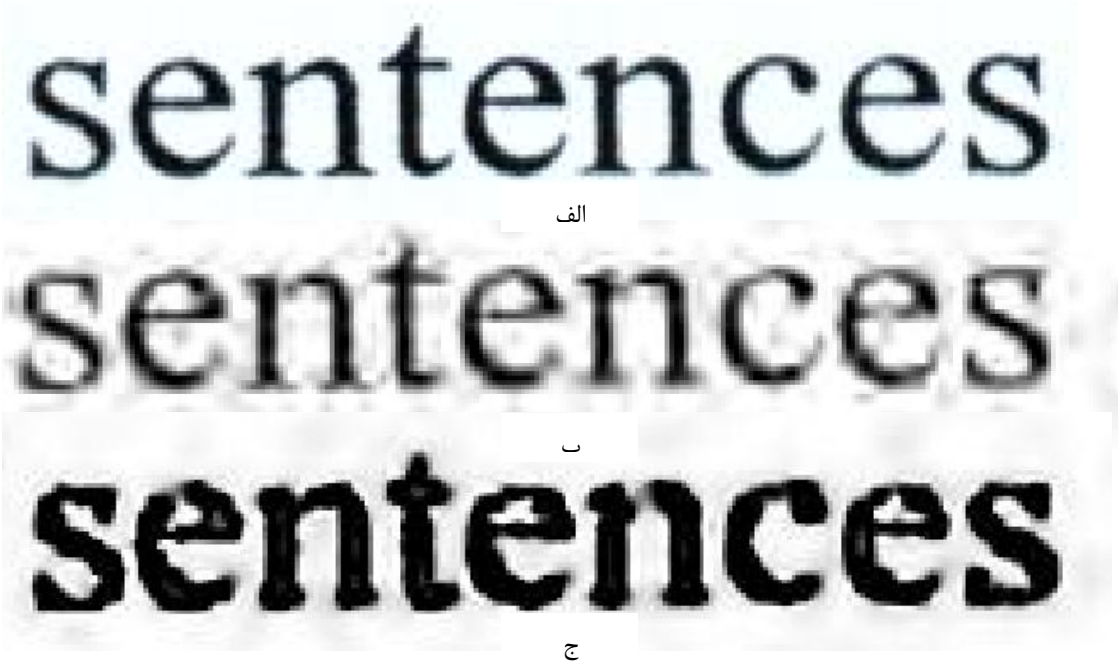
ب

ج

د

ه

شکل ۵-۱۰ روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG2000 الف) تصویر اصلی ب) فشرده‌سازی با نرخ بیت ۰/۰۱۵ ج) فشرده‌سازی با نرخ بیت ۰/۰۱ د) روش پیشنهادی برای نرخ بیت ۰/۰۱۵ ه) روش پیشنهادی برای نرخ بیت ۰/۰۱



الف

ب

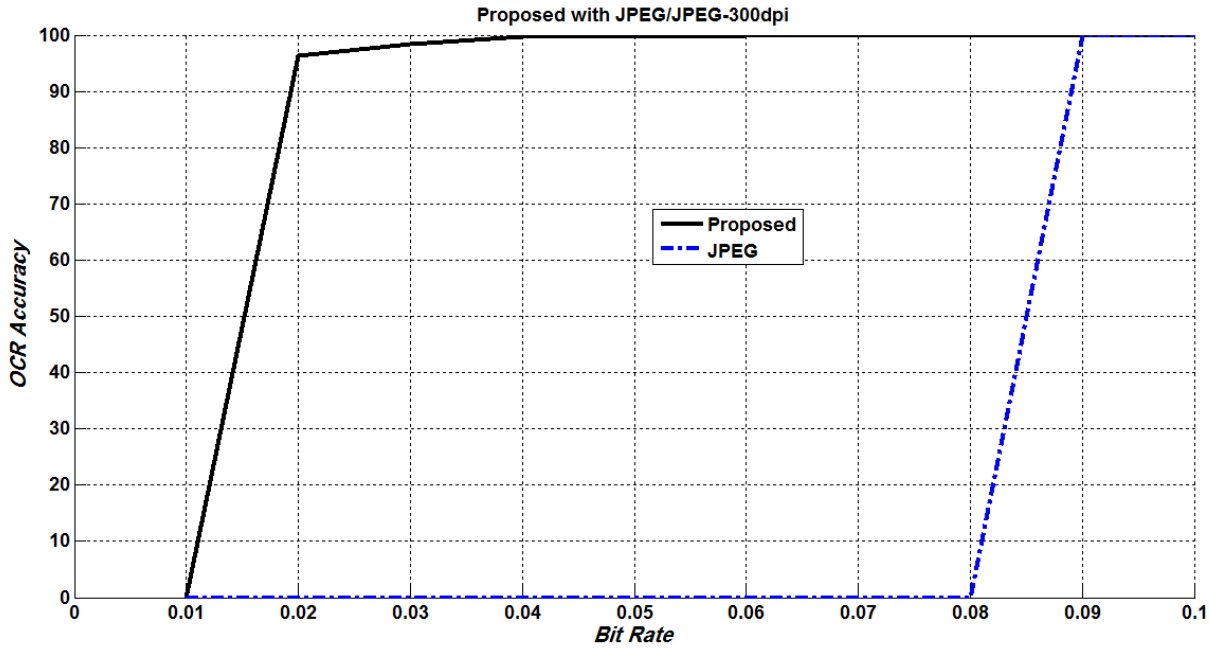
ج

شکل ۵-۱۱ روش پیشنهادی در ترکیب با روش فشرده‌سازی SPIHT الف) تصویر اصلی ب) فشرده‌سازی با نرخ بیت ۰/۰۹ ج) روش پیشنهادی برای نرخ بیت ۰/۰۹

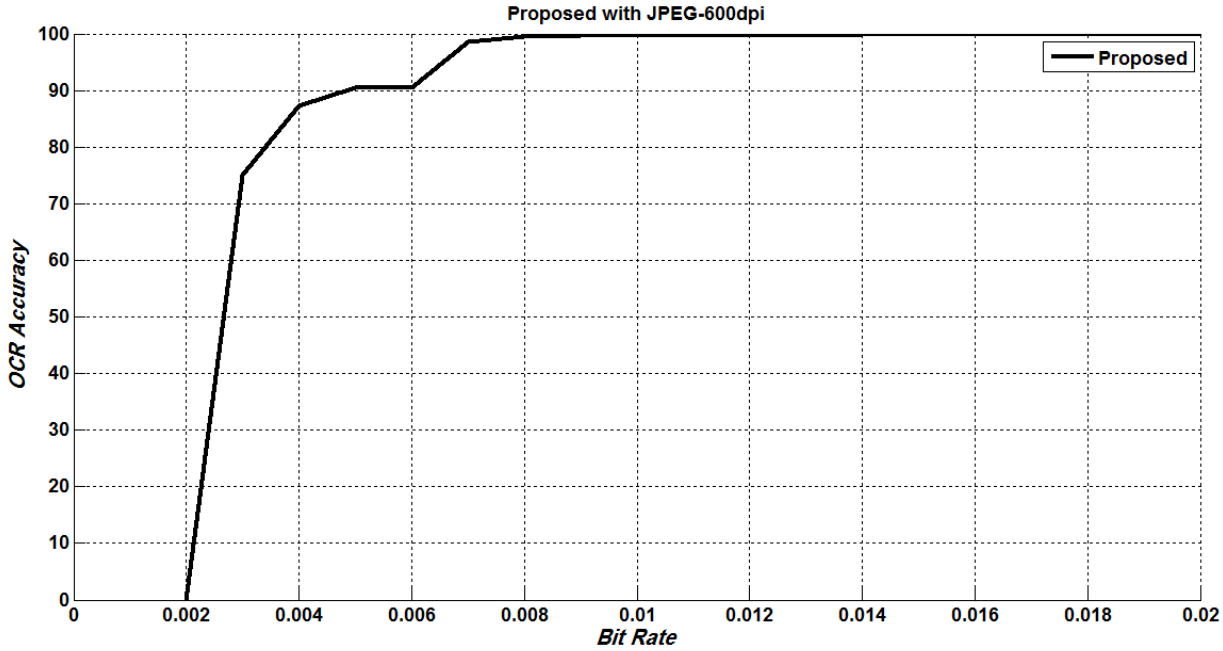
نتایج نشان داده شده در شکل‌های بالا کارایی روش پیشنهادی را در ترکیب با روش‌های فشرده‌سازی مختلف نشان می‌دهد. در ادامه برای نشان دادن این نکته که در ترکیب روش پیشنهادی با روش‌های فشرده‌سازی تا چه میزان می‌توانیم نرخ بیت را پایین بیاوریم براساس معیارهای معرفی شده، چندین نمودار می‌آوریم.

در شکل‌های ۵-۱۲ و ۵-۱۳ نتایج حاصل از OCR روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG روی دو تصویر متنی که به‌ترتیب با درجه‌ی تفکیک مکانی ۳۰۰ و ۶۰۰dpi اسکن شده‌اند، نشان داده شده است. همان‌طور که از شکل ۵-۱۲ مشخص است روش فشرده‌سازی JPEG به تنهایی تا ۰/۰۹ نرخ بیت فشرده‌سازی را انجام داده است اما روش پیشنهادی برای همان تصویر توانسته فشرده‌سازی را در نرخ بیت ۰/۰۲ انجام دهد. از مقایسه شکل ۵-۱۳ با شکل ۵-۸ می‌توان کارایی روش پیشنهادی در کم کردن نرخ بیت فشرده‌سازی را مشاهده کرد. با مشاهده شکل ۵-۸ می‌توان دید که روش فشرده‌سازی

JPEG به تنهایی برای تصویر با درجه‌ی تفکیک مکانی بالای ورودی توانسته است تا ۰/۰۷ نرخ بیت فشرده‌سازی را انجام دهد. اما با ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG توانسته‌ایم جهت فشرده‌سازی این تصویر نرخ بیت را تا ۰/۰۰۴ کاهش دهیم. معیارهای محدود کننده نرخ بیت، نتایج حاصل از OCR و MOS بوده است. یعنی فشرده‌سازی تا جایی ادامه داده‌شده که نتایج برای معیارهای گفته شده قابل قبول باشد.



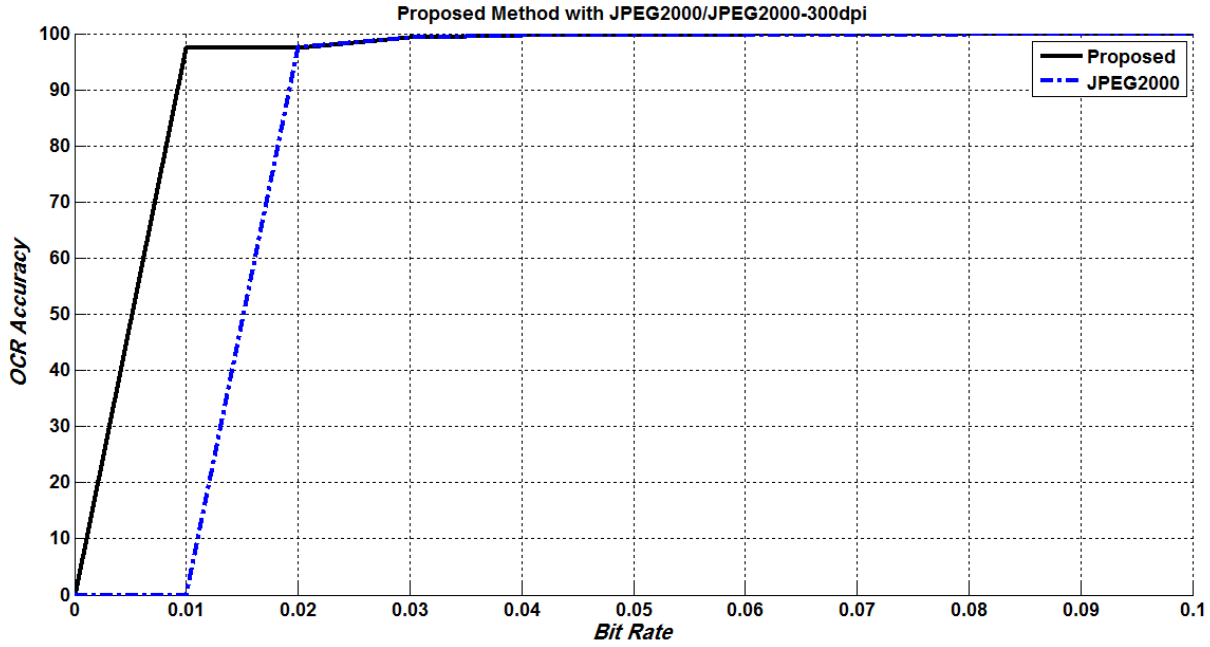
شکل ۵-۱۲ نتایج OCR روش پیشنهادی در ترکیب با فشرده‌سازی JPEG روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



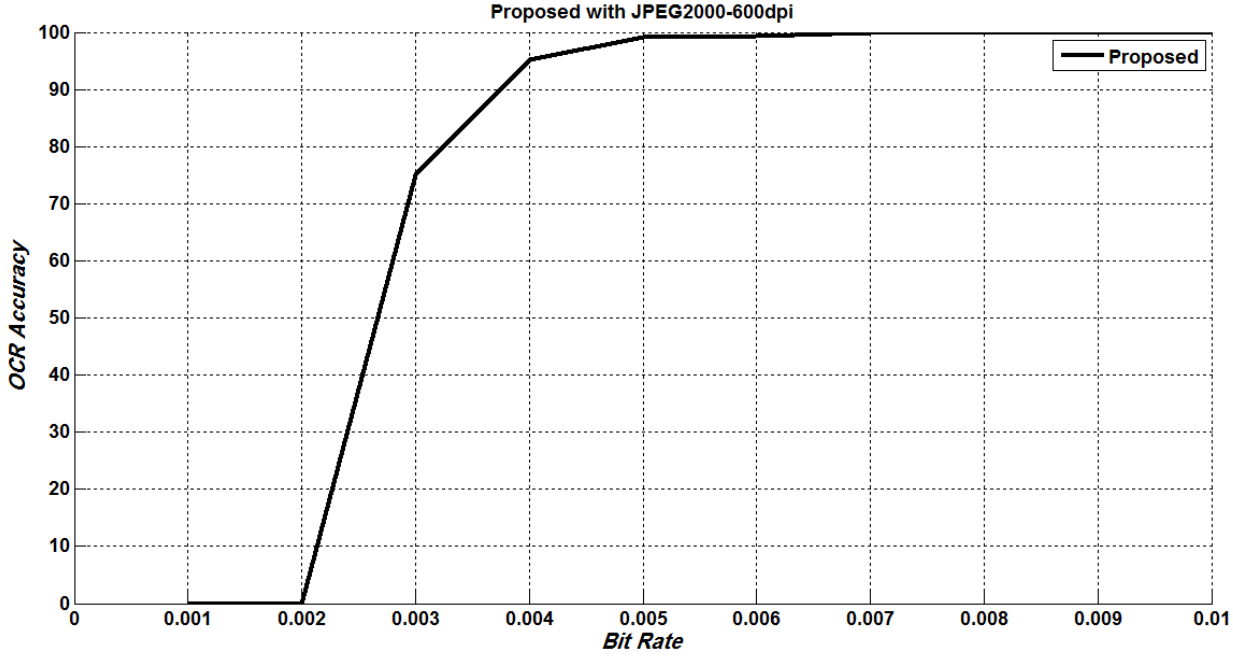
شکل ۵-۱۳ نتایج OCR روش پیشنهادی در ترکیب با فشرده‌سازی JPEG روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

در ادامه در شکل‌های ۵-۱۴ و ۵-۱۵ نتایج حاصل از ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG2000 روی تصاویر استفاده شده در ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG از نظر معیار OCR نشان داده شده است. همان‌طور که از شکل ۵-۱۴ مشخص است در ترکیب روش پیشنهادی با JPEG2000 می‌توان به نرخ بیت کمتری نسبت به اعمال روش JPEG2000 به تنهایی رسید.

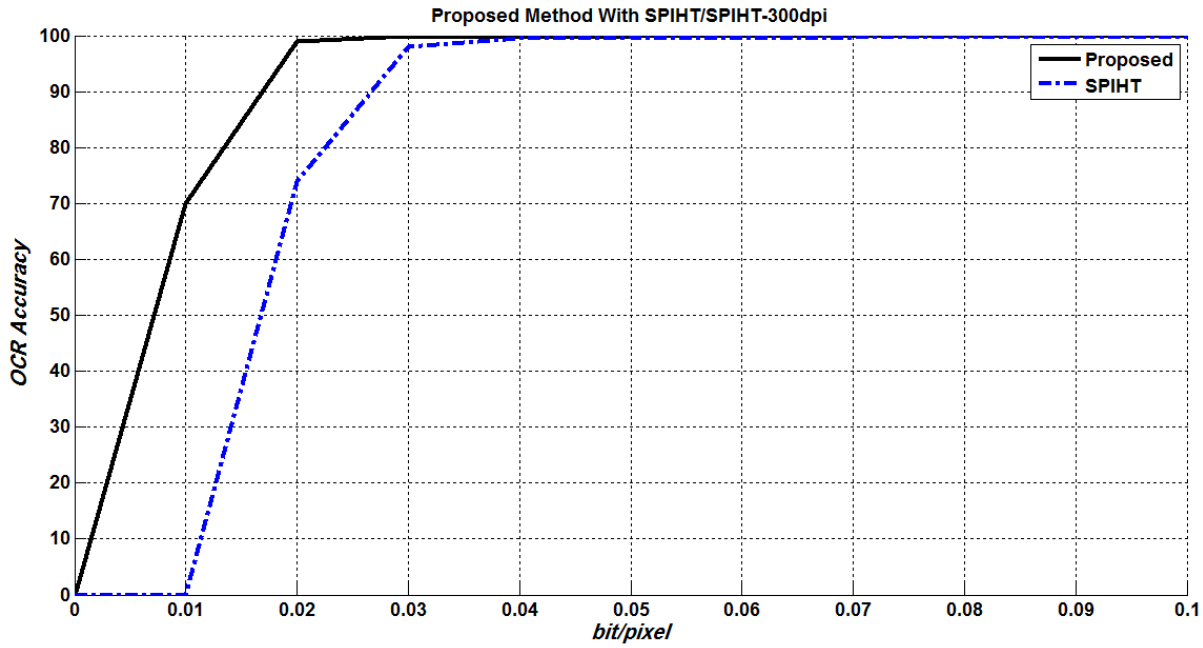
در شکل ۵-۱۵ نیز روش پیشنهادی در ترکیب با JPEG2000 روی تصویری با درجه تفکیک مکانی ۶۰۰dpi اعمال شده است. در مقایسه شکل ۵-۱۵ و ۵-۸ می‌توان دید که روش پیشنهادی توانسته نرخ بیت را از ۰/۰۱۵ به ۰/۰۰۴ برساند. در اینجا نیز فشرده‌سازی تا جایی ادامه داده‌شده که معیار OCR جواب مطلوبی داشته باشد.



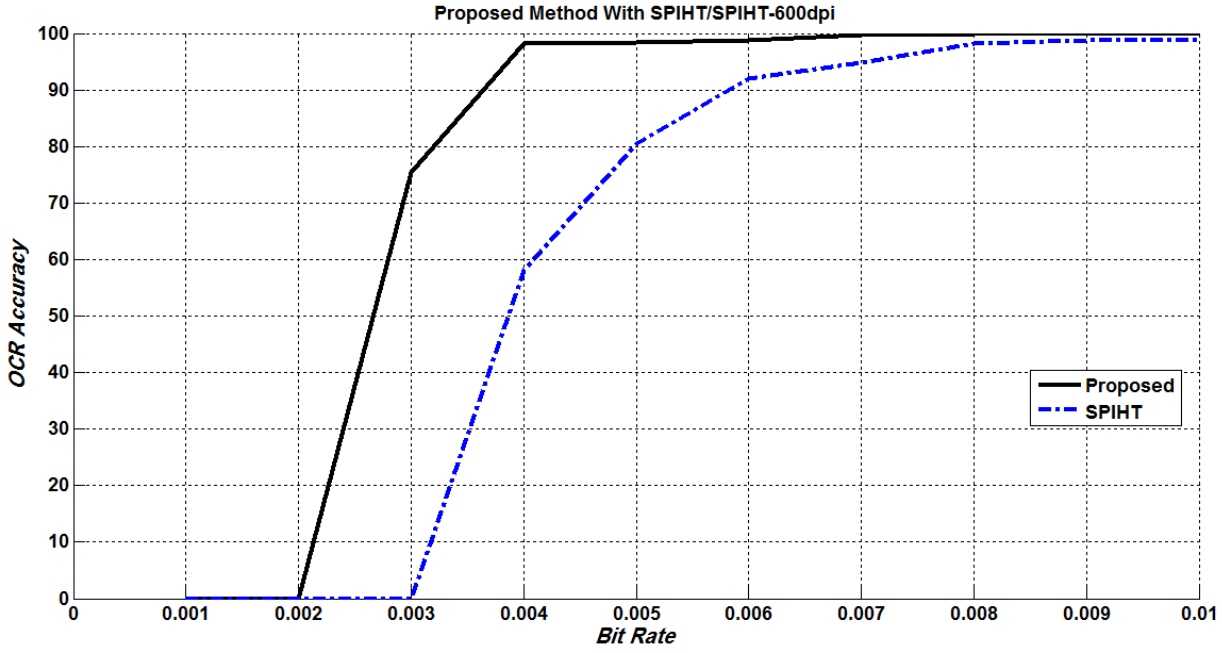
شکل ۵-۱۴ نتایج OCR روش پیشنهادی در ترکیب با فشرده‌سازی JPEG2000 روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



شکل ۵-۱۵ نتایج OCR روش پیشنهادی در ترکیب با فشرده‌سازی JPEG2000 روی تصویر با درجه تفکیک مکانی ۶۰۰dpi



شکل ۵-۱۶ نتایج OCR روش پیشنهادی در ترکیب با فشرده‌سازی SPIHT روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



شکل ۵-۱۷ نتایج OCR روش پیشنهادی در ترکیب با فشرده‌سازی SPIHT روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

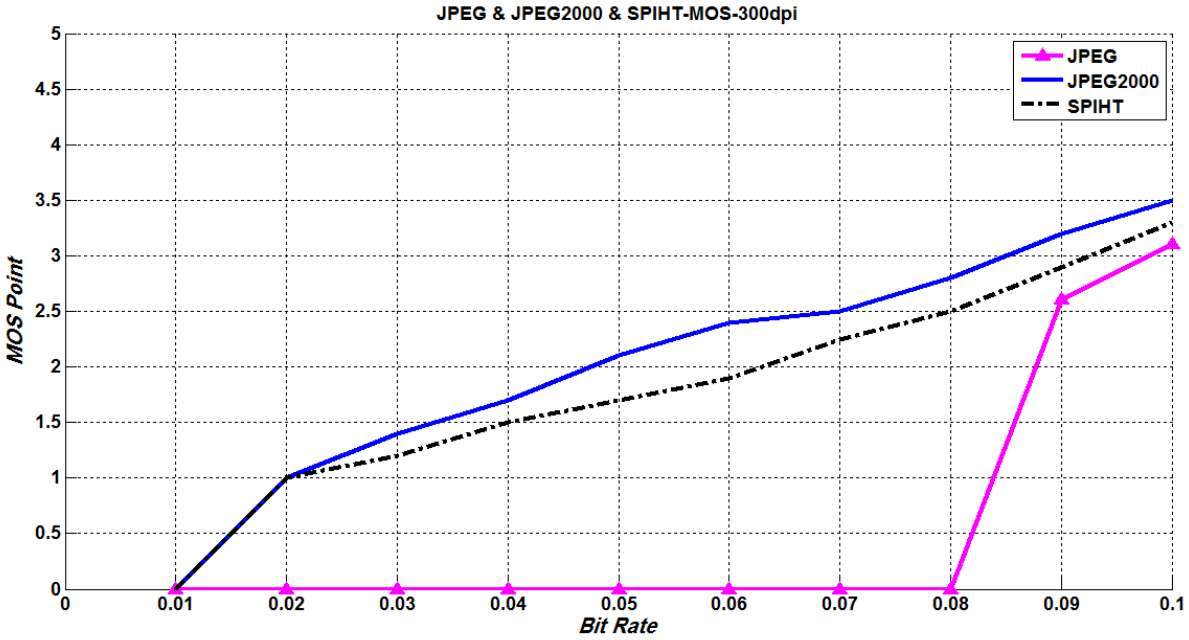
در شکل‌های ۵-۱۶ و ۵-۱۷ نتایج حاصل از ترکیب روش پیشنهادی با روش SPIHT نشان داده شده است. در شکل ۵-۱۶ نتیجه اعمال روش پیشنهادی روی تصویری با درجه تفکیک ۳۰۰dpi در نرخ بیت‌های

مختلف را نشان می‌دهد. شکل ۵-۱۷ نیز نتایج حاصل از اعمال روش پیشنهادی روی تصویری با درجه تفکیک ۶۰۰dpi را نشان می‌دهد. نرخ بیت نسبت به حالت بدون ترکیب روش پیشنهادی خیلی تغییری نکرده است اما صحت OCR بهتر شده است.

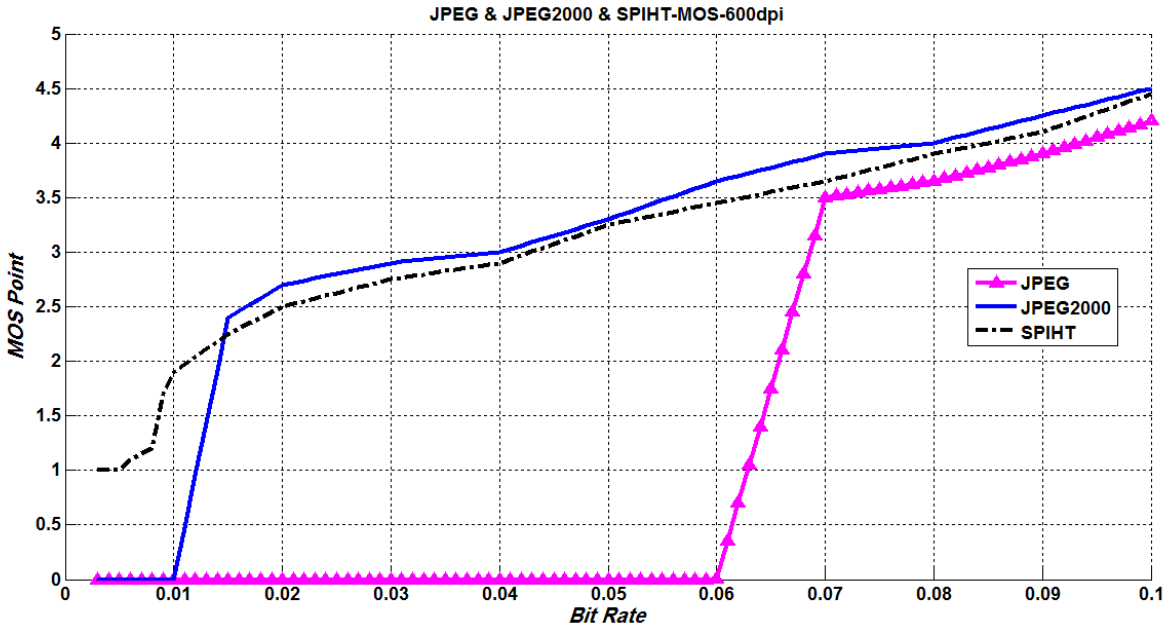
۵-۳ ارزیابی روش پیشنهادی براساس معیار MOS

در ادامه به‌منظور ارزیابی روش پیشنهادی از معیار MOS استفاده شده‌است. نتایج حاصل از فشرده-سازی با روش‌های JPEG، JPEG2000، SPIHT و روش پیشنهادی به ۲۰ نفر نشان داده‌شد و از آنها خواسته‌شد که از ۱ تا ۵ به تصاویر نمره بدهند. در شکل‌های ۵-۱۸ و ۵-۱۹ نتایج حاصل برای روش‌های JPEG، JPEG2000 و SPIHT روی دو تصویر به‌ترتیب با درجه‌ی تفکیک مکانی ۳۰۰dpi و ۶۰۰dpi نشان داده شده است. نتایج نشان داده شده میانگین ۲۰ نمره برای هر تصویر می‌باشد. نتایج حاصل از شکل‌های ۵-۱۸ و ۵-۱۹ نشان می‌دهند که افراد کیفیت روش JPEG2000 را در مقایسه با دو روش دیگر در نرخ بیت‌های قابل مقایسه، مطلوب‌تر می‌دانسته‌اند.

در این شکل‌ها نیز برای نرخ بیت‌هایی که رسیدن به آنها با روش‌های گفته شده امکان پذیر نبوده است، داده‌های مصنوعی صفر ایجاد شده‌اند.



شکل ۵-۱۸ مقایسه روش‌های فشرده‌سازی از نظر معیار MOS برای تصویر با درجه‌ی تفکیک مکانی ۳۰۰dpi

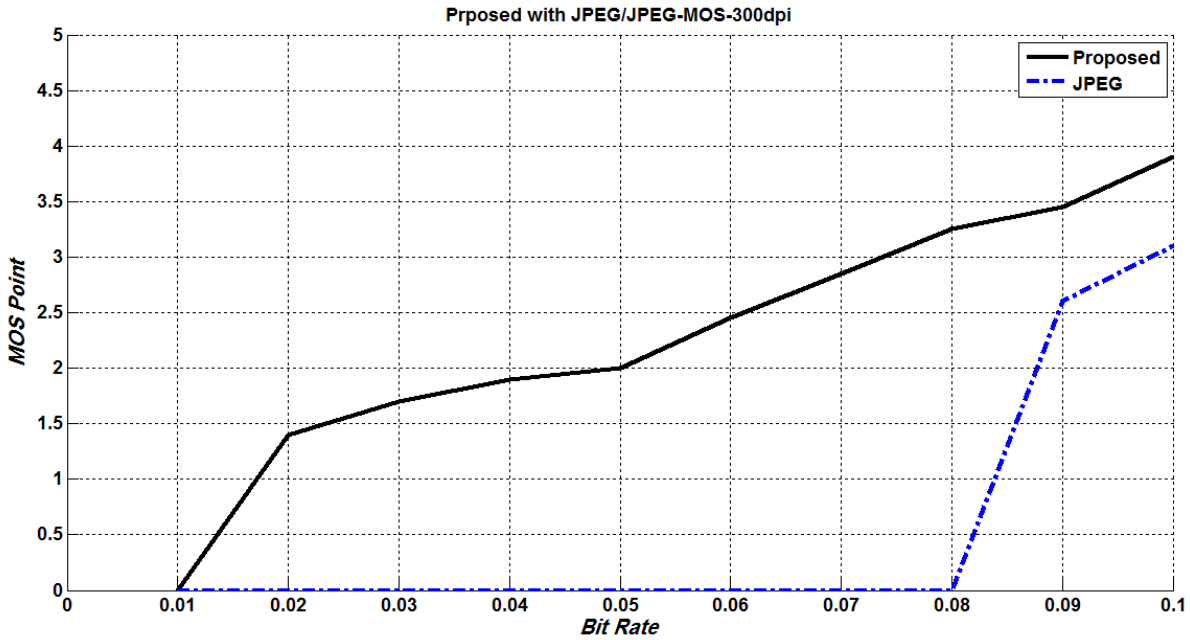


شکل ۵-۱۹ مقایسه روش‌های فشرده‌سازی از نظر معیار MOS برای تصویر با درجه‌ی تفکیک مکانی ۶۰۰dpi

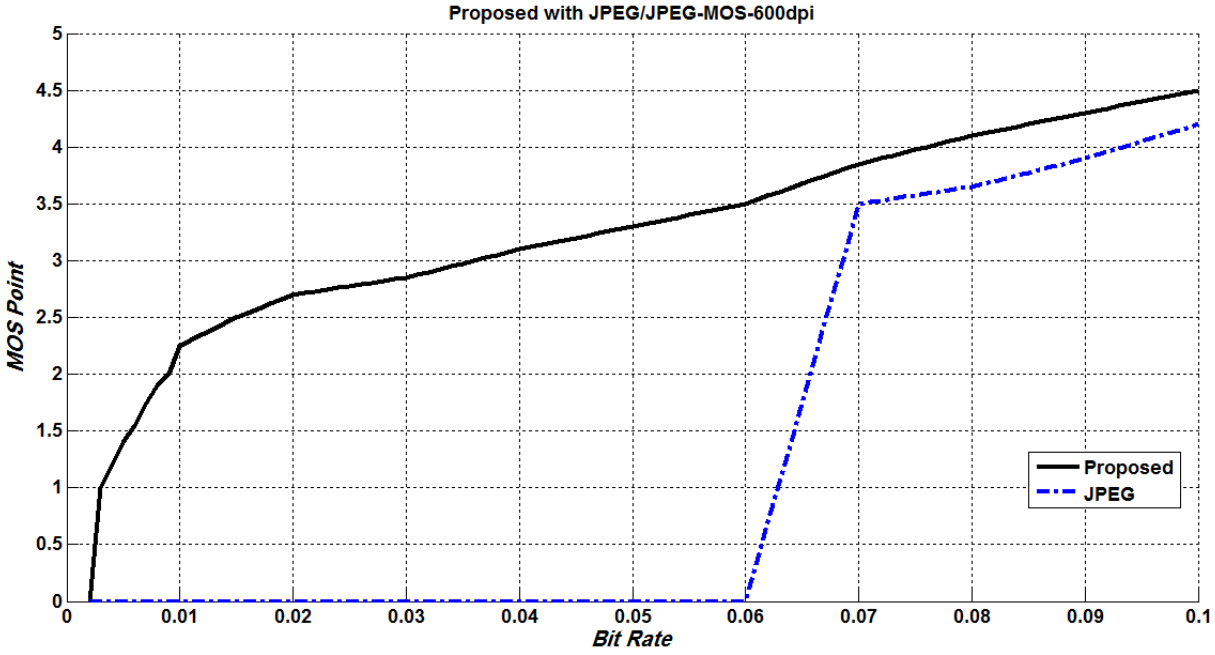
در ادامه نتایج MOS برای روش پیشنهادی در ترکیب با روش‌های فشرده‌سازی آورده شده است.

در شکل‌های ۵-۲۰ و ۵-۲۱ نتایج حاصل از ترکیب روش پیشنهادی با JPEG برای دو تصویر گفته شده

نشان داده شده است. در هر دو نمودار خط ممتد مربوط به روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG و خط-نقطه مربوط به روش فشرده‌سازی JPEG به تنهایی است.



شکل ۵-۲۰ نتایج MOS روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



شکل ۵-۲۱ نتایج MOS روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

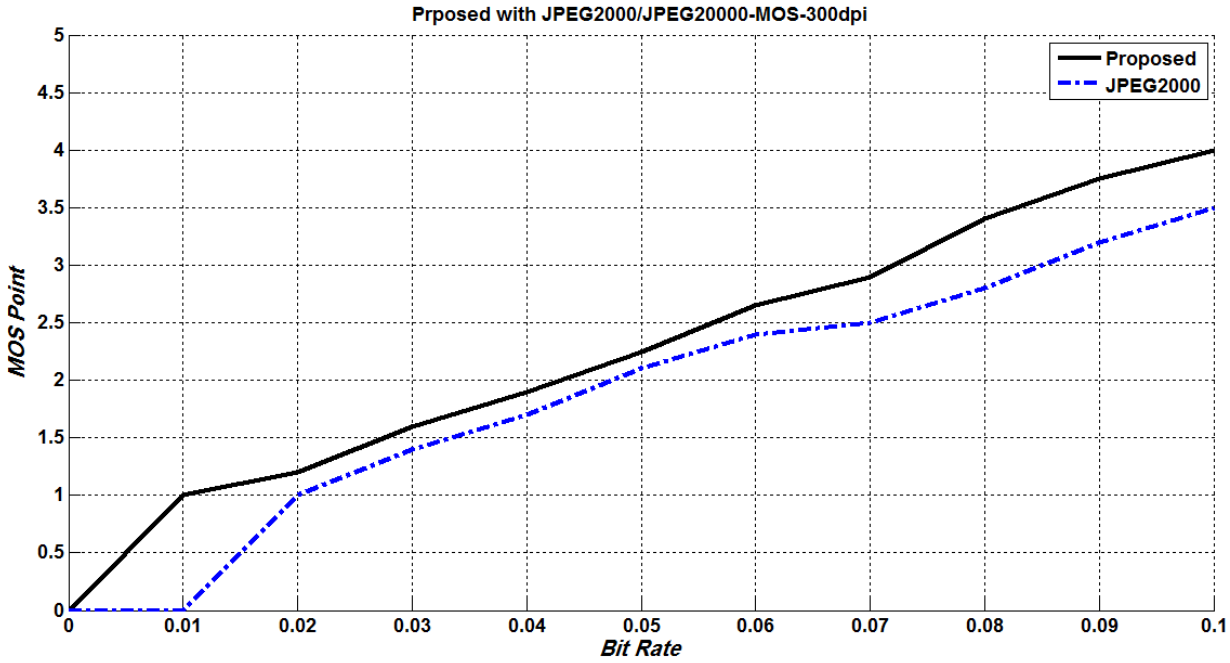
همان‌طور که از شکل‌های ۵-۲۰ و ۵-۲۱ دیده می‌شود، ترکیب روش پیشنهادی با روش فشرده-سازی JPEG از نظر نرخ بیت توانسته به نرخ بیت کمتری برسد. همچنین در این شکل‌ها مشخص است که ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG از نظر خوانایی انسانی در جامعه ۲۰ نفره ذکر شده، نتایج مطلوب‌تری نسبت به روش فشرده‌سازی JPEG به‌تنهایی داشته است.

در شکل‌های ۵-۲۲ و ۵-۲۳ نتایج حاصل از ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG2000 روی دو تصویر گفته شده که به‌ترتیب با درجه‌ی تفکیک مکانی ۳۰۰dpi و ۶۰۰dpi روبش شده‌اند، نشان داده شده است.

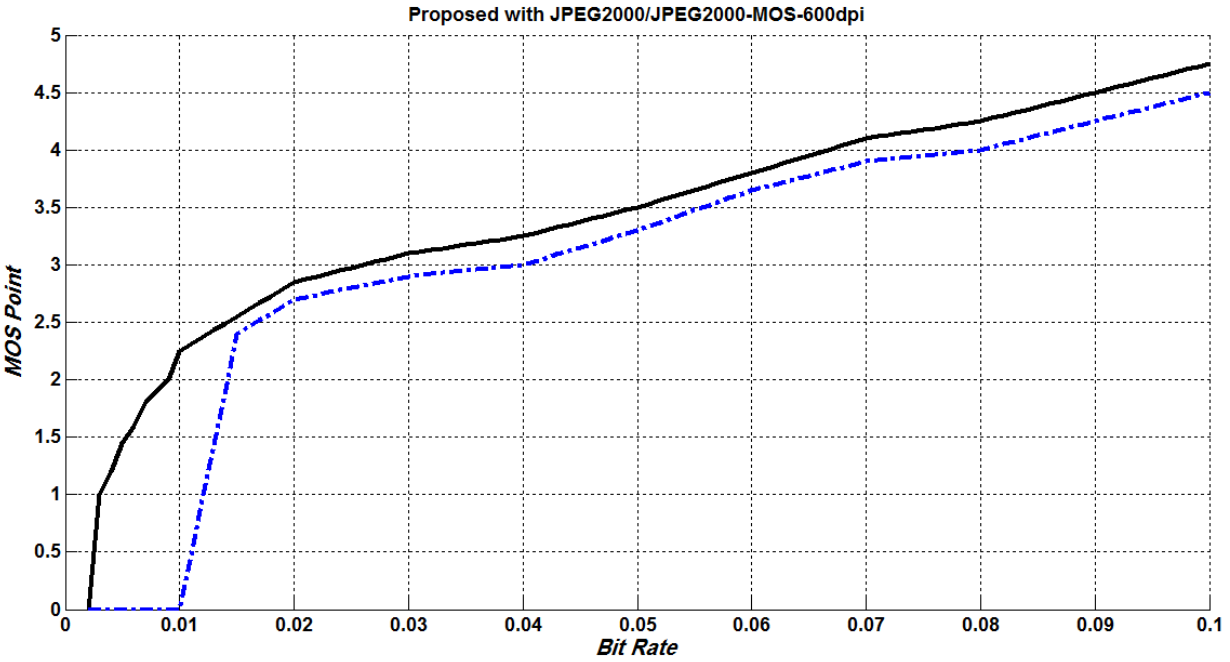
همان‌طور که از شکل ۵-۲۲ مشخص است روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG2000 توانسته است خوانایی انسانی تصویر فشرده‌شده ورودی را بهبود دهد. در این شکل نمودار برای تصویر با درجه تفکیک مکانی ۳۰۰dpi رسم شده است.

در شکل ۵-۲۳ نیز نمودار برای تصویر با درجه‌ی تفکیک مکانی ۶۰۰dpi رسم شده است. همان‌طور که از این شکل نیز دیده می‌شود روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG2000 برای این تصویر نیز توانسته است خوانایی انسانی را بهبود دهد. در این شکل می‌توان به نرخ بیت کمتر بدست آمده در ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG2000 در مقایسه با روش فشرده‌سازی JPEG2000 نیز توجه کرد.

در هر دو شکل ۵-۲۲ و ۵-۲۳، خط ممتد مربوط به روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG2000 و خط-نقطه مربوط به روش فشرده‌سازی JPEG2000 به تنهایی است.

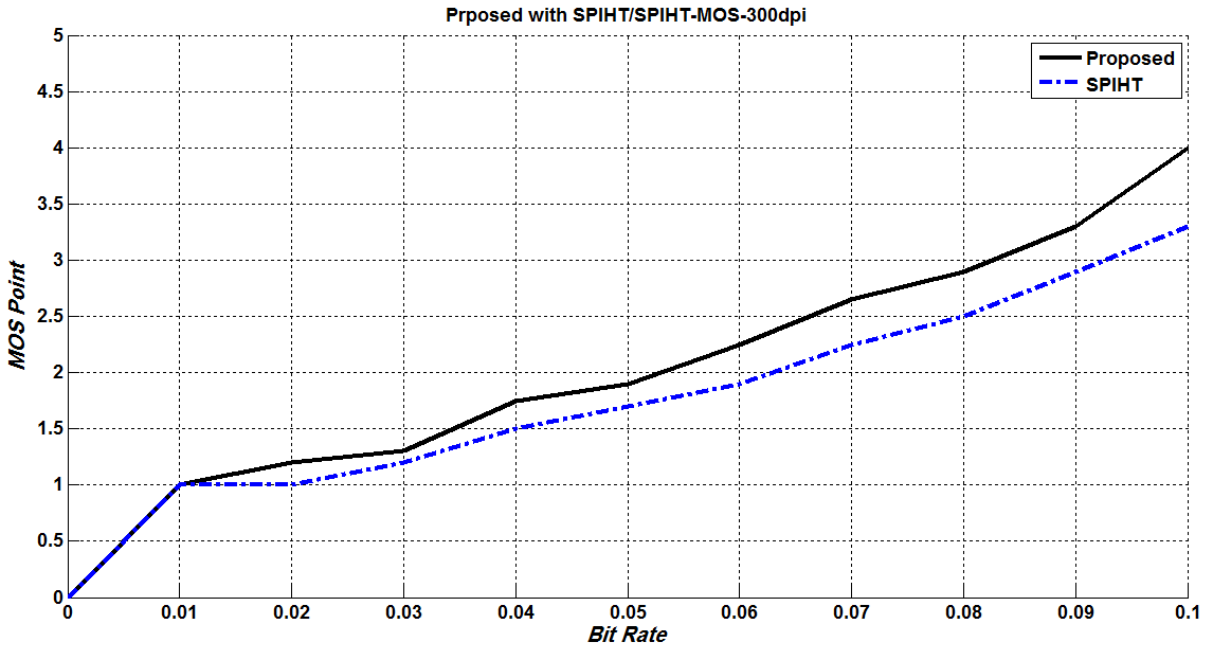


شکل ۵-۲۲ نتایج MOS روش پیشنهادی در ترکیب با روش JPEG2000 روی تصویر با درجه تفکیک مکانی ۳۰۰dpi

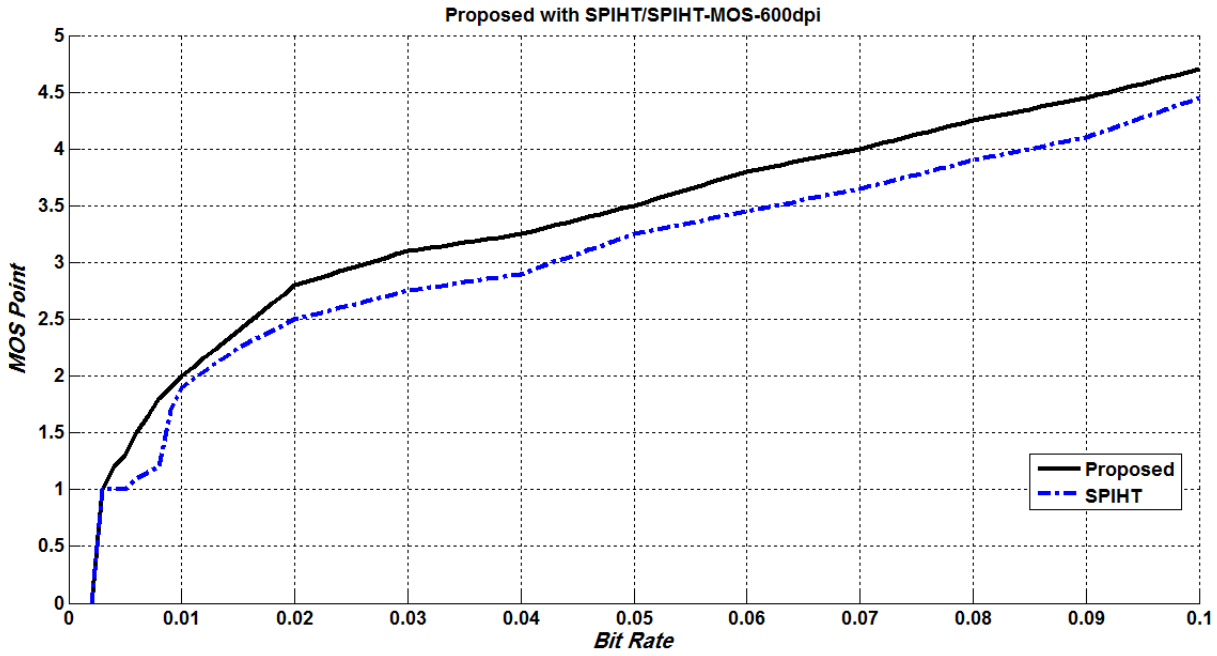


شکل ۵-۲۳ نتایج MOS روش پیشنهادی در ترکیب با روش JPEG2000 روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

آخرین نمودارهای مربوط به معیار MOS به ترکیب روش پیشنهادی با روش فشرده‌سازی SPIHT در مقایسه با روش فشرده‌سازی SPIHT به تنهایی مربوط می‌شود. این نمودارها برای تصاویر با درجه‌ی تفکیک مکانی ۳۰۰dpi و ۶۰۰dpi به‌ترتیب در شکل‌های ۵-۲۴ و ۵-۲۵ نشان داده شده است. در این شکل‌ها خط ممتد نمودار مربوط به روش پیشنهادی در ترکیب با روش SPIHT و خط-نقطه مربوط به روش فشرده‌سازی SPIHT به‌تنهایی است.



شکل ۵-۲۴ نتایج MOS روش پیشنهادی در ترکیب با روش SPIHT روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



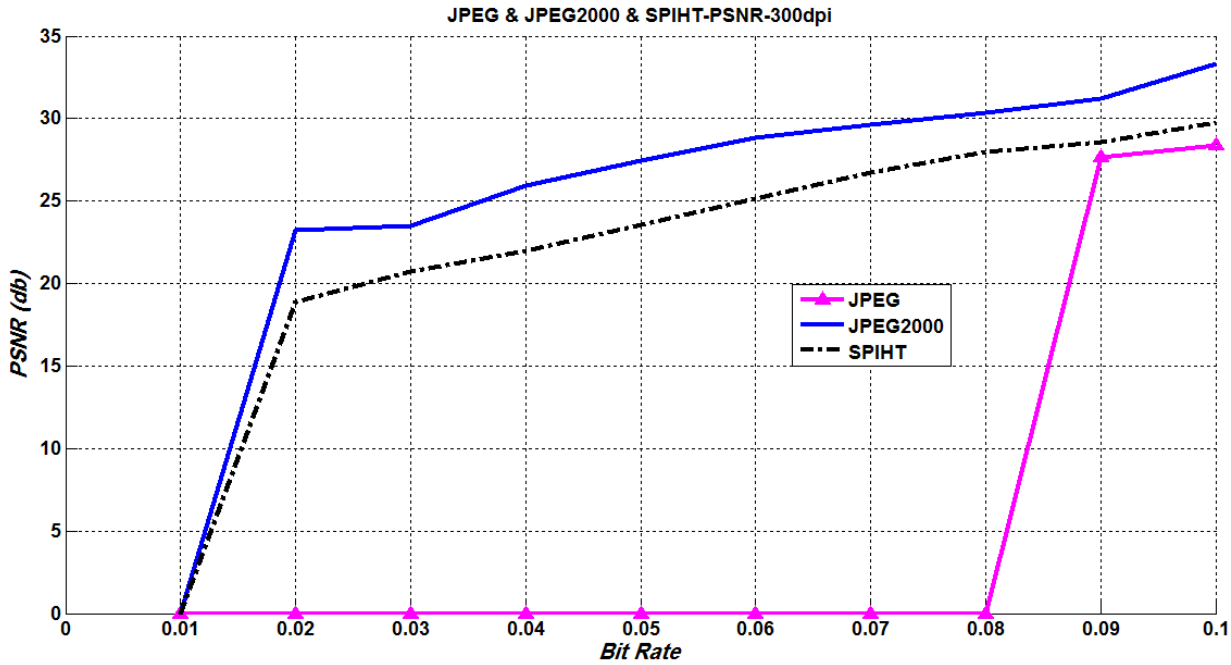
شکل ۵-۲۵ نتایج MOS روش پیشنهادی در ترکیب با روش SPIHT روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

همان‌طور که از شکل‌های ۵-۲۴ و ۵-۲۵ مشخص است روش پیشنهادی در ترکیب با روش SPIHT برای هر دو تصویر از نظر MOS جواب بهتری را داشته است.

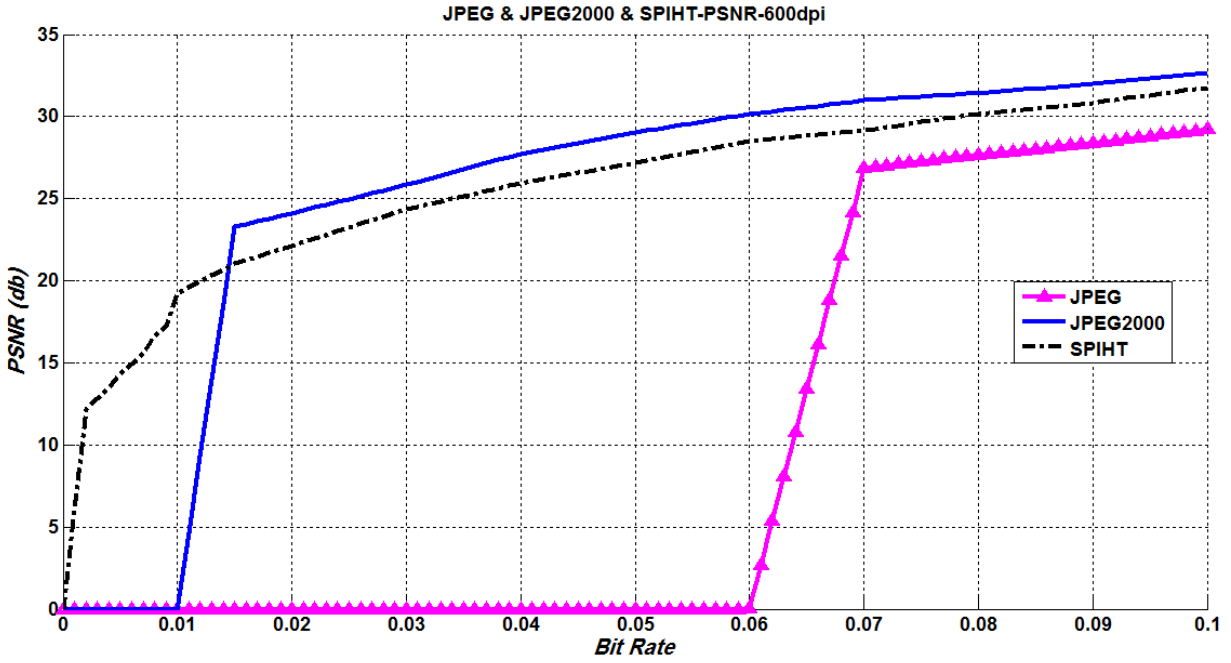
۵-۴ ارزیابی روش پیشنهادی براساس معیار PSNR

در ادامه بررسی معیارهای ارزیابی روش پیشنهادی به آخرین معیار یعنی معیار PSNR می‌رسیم. این معیار در واقع میزان خطای بین تصویر بازسازی شده و تصویر اصلی را نشان می‌دهد. نتایج گزارش شده برای این معیار براساس رابطه (۵-۱) بدست آمده‌است. در ابتدا نتایج حاصل برای مقایسه بین روش‌های فشرده‌سازی مختلف گزارش شده‌اند. سپس نتایج حاصل از ترکیب روش پیشنهادی با هر یک از روش‌های فشرده‌سازی گفته شده گزارش شده‌اند.

شکل‌های ۵-۲۶ و ۵-۲۷ نتایج حاصل از اعمال روش‌های فشرده‌سازی مختلف روی دو تصویر به‌ترتیب با درجه‌ی تفکیک مکانی ۳۰۰dpi و ۶۰۰dpi را براساس معیار PSNR نشان می‌دهند.



شکل ۵-۲۶ مقایسه روش‌های فشرده‌سازی از نظر معیار PSNR برای تصویر با درجه‌ی تفکیک مکانی ۳۰۰dpi

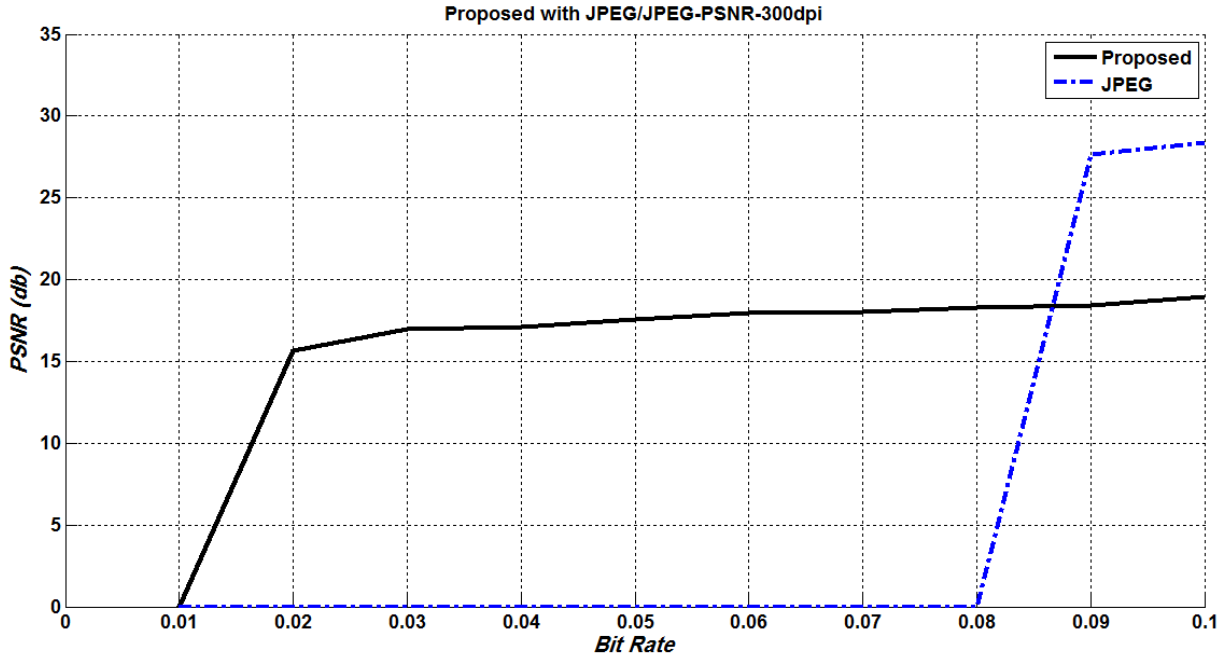


شکل ۵-۲۷ مقایسه روش‌های فشرده‌سازی از نظر معیار PSNR برای تصویر با درجه‌ی تفکیک مکانی ۶۰۰dpi

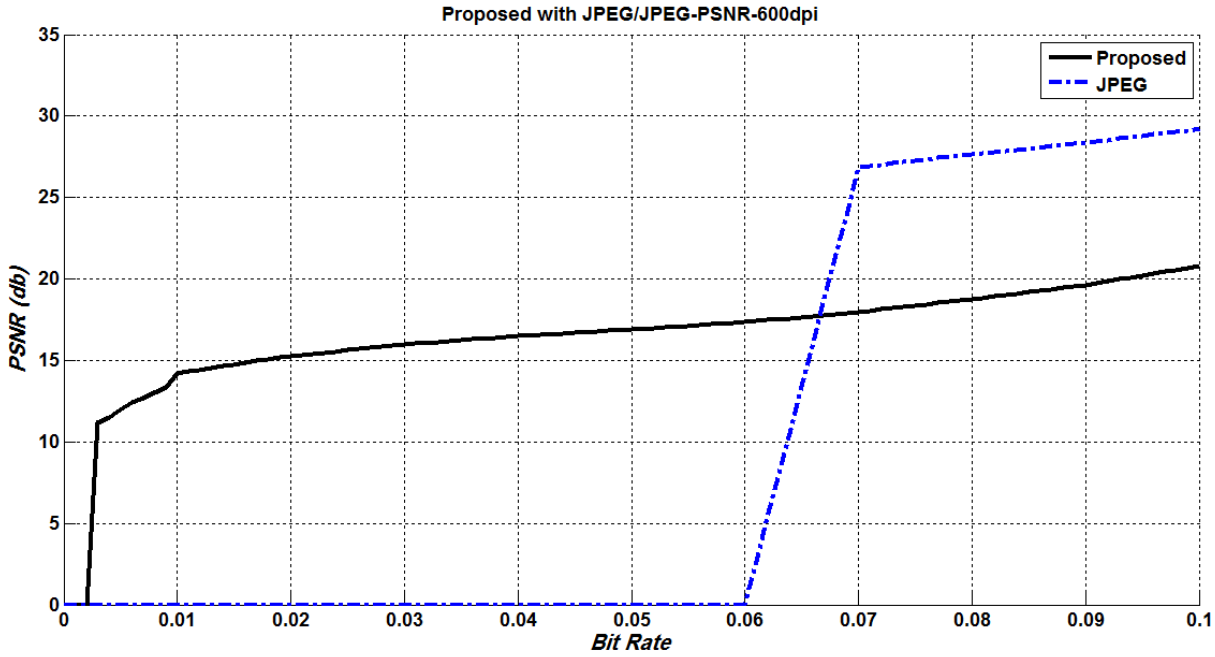
از نتایج بدست آمده در شکل‌های ۵-۲۶ و ۵-۲۷ دیده می‌شود که روش JPEG2000 نسبت به دو روش دیگر از نظر معیار PSNR جواب بهتری داشته است. در این نمودارها باز هم محدودیت فشرده‌سازی

برای روش‌های JPEG و JPEG2000 قابل مشاهده است. در ادامه روش پیشنهادی در ترکیب با روش‌های فشرده‌سازی دیگر از نظر معیار PSNR ارزیابی شده‌است.

در شکل‌های ۵-۲۸ و ۵-۲۹ نتایج حاصل از ترکیب روش پیشنهادی با روش JPEG بر مبنای معیار PSNR روی تصاویر ذکر شده نشان داده شده است. همان‌طور که از این شکل‌ها دیده می‌شود، نتایج بدست آمده از روش پیشنهادی براساس این معیار در مقایسه با روش فشرده‌سازی JPEG خوب نبوده است. دلیل این نتایج این است که در روش پیشنهادی در مرحله بازسازی ضخامت حروف در تصویر متنی بازسازی شده بیشتر از ضخامت حروف در تصویر متنی اصلی است. این سبب می‌شود که در زمان محاسبه‌ی PSNR نقاطی که در تصویر بازسازی شده به ضخامت حروف افزوده شده‌اند، خطای بازسازی را نسبت به تصویر اصلی زیاد کنند.

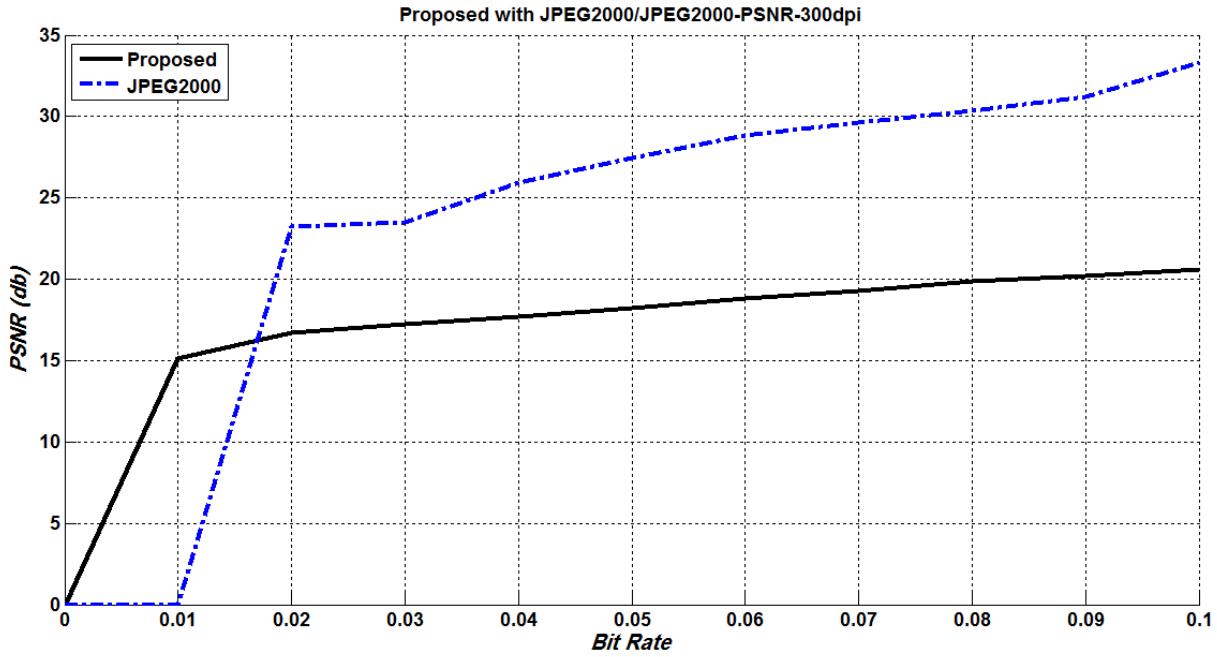


شکل ۵-۲۸ نتایج PSNR روش پیشنهادی در ترکیب با روش JPEG روی تصویر با درجه تفکیک مکانی ۳۰۰dpi

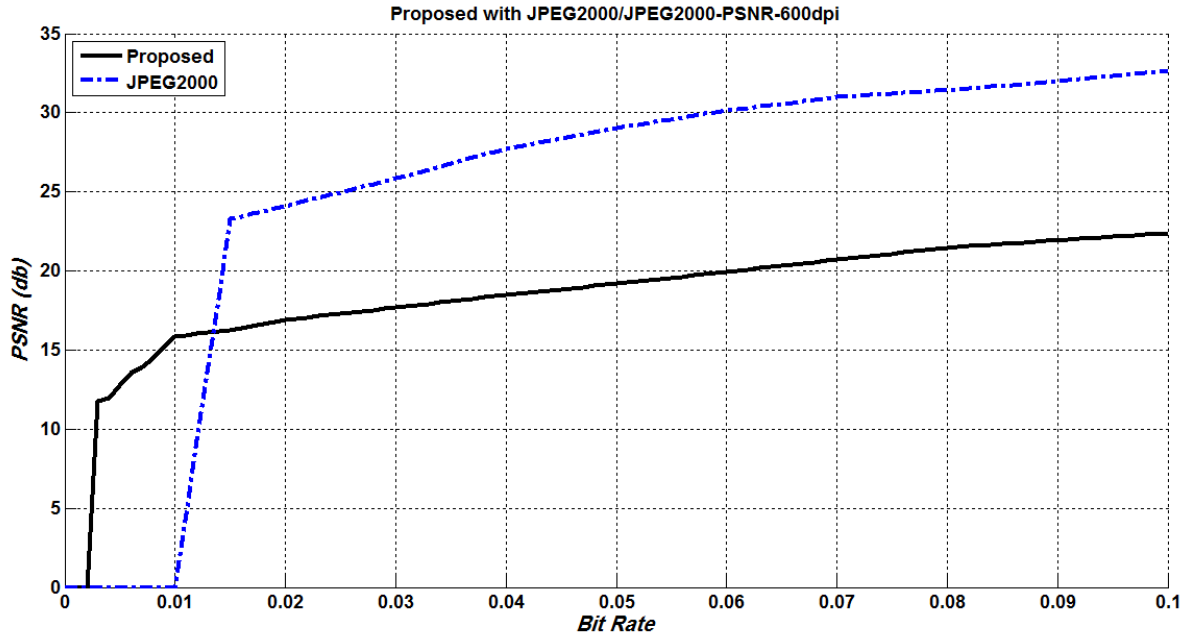


شکل ۵-۲۹ نتایج PSNR روش پیشنهادی در ترکیب با روش JPEG روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

شکل‌های ۵-۳۰ و ۵-۳۱ نتایج حاصل از ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG2000 را بر مبنای معیار PSNR نشان می‌دهند. از این شکل‌ها نیز می‌توان مانند نتایج گزارش شده در شکل ۵-۲۸ و ۵-۲۹ دید که روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG2000 نیز از نظر معیار PSNR نتایج ضعیف‌تری نسبت به روش فشرده‌سازی JPEG2000 به تنهایی دارد.

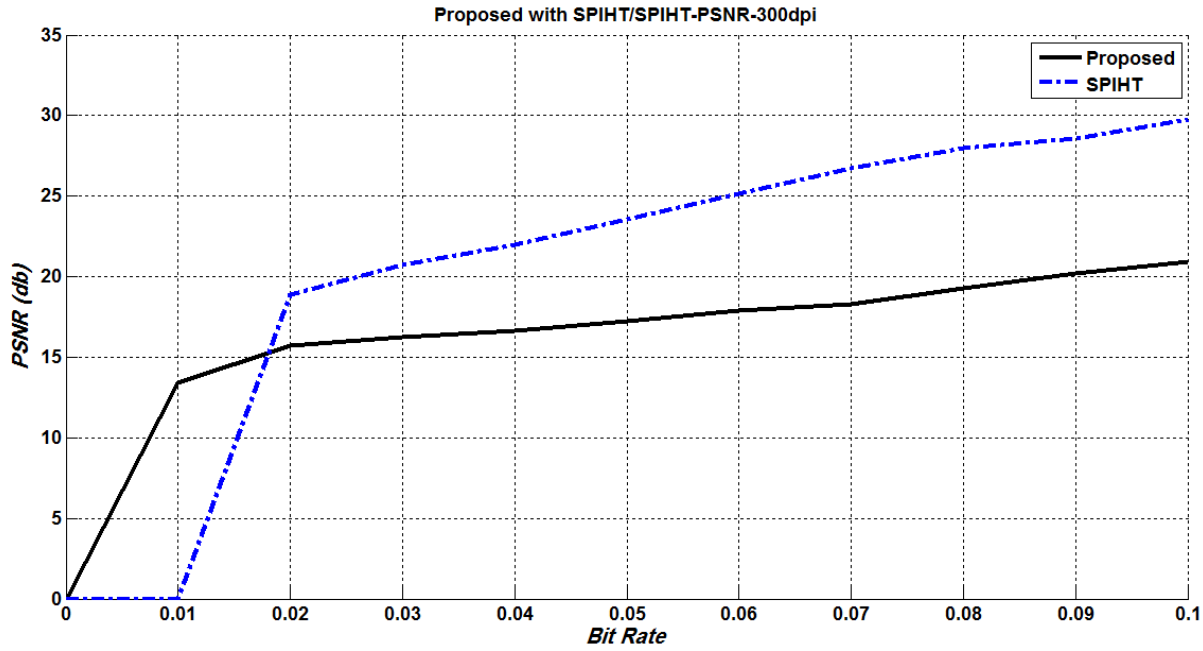


شکل ۵-۳۰ نتایج PSNR روش پیشنهادی در ترکیب با روش JPEG2000 روی تصویر با درجه تفکیک مکانی ۳۰۰dpi

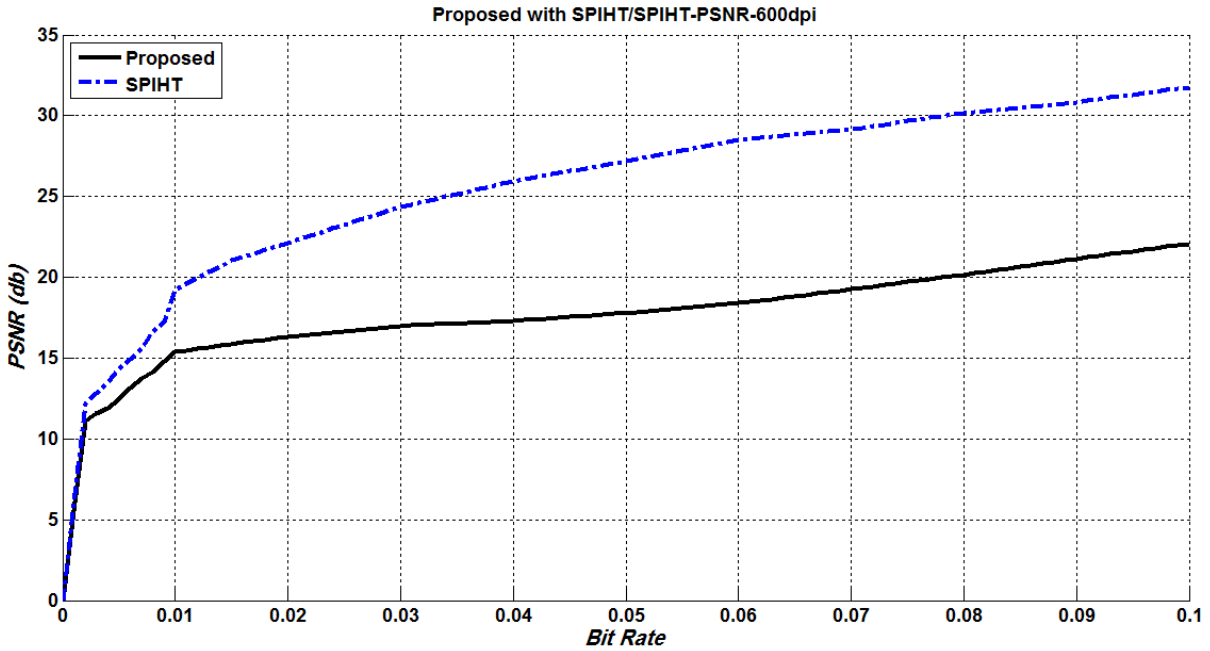


شکل ۵-۳۱ نتایج PSNR روش پیشنهادی در ترکیب با روش JPEG2000 روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

در آخر نیز نتایج حاصل از ترکیب روش پیشنهادی با روش SPIHT بر مبنای معیار PSNR در شکل‌های ۵-۳۲ و ۵-۳۳ نشان داده شده است.



شکل ۵-۳۲ نتایج PSNR روش پیشنهادی در ترکیب با روش SPIHT روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



شکل ۵-۳۳ نتایج PSNR روش پیشنهادی در ترکیب با روش SPIHT روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

از شکل‌های ۵-۳۲ و ۵-۳۳ نیز دیده می‌شود که روش فشرده‌سازی SPIHT به‌تنهایی از نظر معیار PSNR نسبت به روش پیشنهادی در ترکیب با روش SPIHT جواب بهتری داشته است.

در پایان نکته‌ی دیگری که ذکر می‌شود در رابطه با زمان اجرای روش پیشنهادی است. با توجه به اینکه تصاویر مورد استفاده تصاویری با درجه‌ی تفکیک مکانی بالایی هستند، اجرای روش پیشنهادی زمان‌بر است. با ایده‌ی کاهش ابعاد که ایده‌ی اصلی در افزایش میزان فشرده‌سازی در این پایان‌نامه بوده‌است، می‌توان سرعت اجرای الگوریتم را نیز بهبود داد. اما روی کاهش ابعاد محدودیت وجود دارد.

تا جایی می‌توان ابعاد تصویر با درجه‌ی تفکیک بالای ورودی را کاهش داد که نتیجه‌ی روش پیشنهادی بر مبنای معیارهای بحث شده در بالا قابل قبول باشد. به‌طور مثال تصویری با درجه‌ی تفکیک مکانی بالای ورودی با ضرایب ۴ و ۵ ابعاد آن کاهش داده شده است (یعنی تعداد سطر و ستون‌های آن به‌ترتیب به ۴ و ۵ تقسیم شده‌است). سپس با استفاده از روش‌های موجود هر دو تصویر در یک نرخ بیت برابر فشرده شده‌اند. در مرحله بازسازی تصویر فشرده‌شده به ابعاد اولیه برگردانده شده‌اند. پس از بازسازی هر دو تصویر بازسازی شده جهت OCR به نرم‌افزار معرفی شده داده شدند و نتایج آن با تصویر اصلی اولیه به‌منظور صحت OCR مقایسه شد. برای تصویری که ابعاد آن با ضریب ۴ کاهش یافته بود، صحت OCR برابر ۹۹/۷۱ درصد و برای تصویر که با ضریب ۵ کاهش ابعاد داده شده بود، صحت OCR برابر ۹۹/۶۳ درصد بدست آمد. در نتیجه OCR تفاوت زیادی ایجاد نشده است اما زمان اجرای بدست آمده برای بازسازی تصویری که با ضریب ۴ کاهش ابعاد داده شده بود برابر ۴۰۱ ثانیه و برای تصویری که با ضریب ۵ کاهش ابعاد داده شده بود برابر ۶۲ ثانیه بدست آمد. نتایج بدست آمده روی یک سیستم با مشخصات فرکانس کاری پردازنده ۲/۵GHz و حافظه‌ی ۴GB اجرا شده‌اند.

پس می‌توان گفت که با توجه به کاربرد و دقت مورد نظر می‌توان زمان‌های اجرای متفاوتی را برای روش پیشنهادی داشت.

فصل ششم:

نتیجه‌گیری و پیشنهاد راه‌کارهای آینده

۶-۱ نتیجه‌گیری

در این پایان‌نامه یک سیستم فشرده‌سازی/ بازسازی برای تصاویر متنی با درجه‌ی تفکیک مکانی بالا مبتنی بر فرا تفکیک‌پذیری ارائه شد. در تصاویر متنی درجه تفکیک مکانی از اهمیت بالایی برخوردار است اما این ویژگی باعث می‌شود که حجم ذخیره‌سازی این تصاویر زیاد شود. با توجه به محدود بودن حافظه‌های دیجیتالی و حجم زیاد این تصاویر نیاز به فشرده‌سازی تصاویر متنی امری ضروری به‌نظر می‌رسد. در کاربرد مورد نظر چون نیاز به حجم خروجی کم به‌منظور ذخیره‌سازی تصاویر متنی بود، از روش‌های فشرده‌سازی با اتلاف استفاده شده‌است. اما مشکلی که پیش می‌آید این است که فشرده‌سازی می‌تواند موجب اثرات مخربی روی تصاویر متنی شود. بنابراین بسته به کاربرد تا جایی می‌توان تصاویر متنی را فشرده کرد که خوانایی انسانی/ماشینی این تصاویر کم نشود. علاوه‌بر مشکلات ذکر شده برای روش‌های فشرده‌سازی JPEG و JPEG2000، مشکل دیگر در این روش‌ها محدود بودن میزان فشرده‌سازی یا نرخ بیت خروجی است.

ایده اصلی در این پایان‌نامه این بوده‌است که با ارائه روشی مناسب بتوان میزان فشرده‌سازی را بیشتر و اثرات مخرب روی تصویر فشرده‌شده را نیز کم کرد. برای رسیدن به میزان فشرده‌سازی بیشتر از ایده کاهش ابعاد در تصاویر متنی استفاده شده‌است. به این معنا که ابتدا ابعاد تصویر متنی با درجه‌ی تفکیک مکانی بالا با استفاده از یک روش کاهش ابعاد، مانند دومکعبی، کاهش یافته و سپس به فشرده‌ساز داده می‌شود. نتایج نشان داد که با این روش می‌توان به نرخ بیت‌های کمتر یا میزان فشرده‌سازی بیشری دست پیدا کرد. البته باید به این موضوع توجه شود که کاهش ابعاد روی کیفیت تصویر بی تأثیر نیست و ممکن است باعث تنزل در کیفیت تصویر شود. بنابراین باید روشی انتخاب می‌شد که بتواند هم ابعاد تصویر را افزایش دهد و هم اثرات مخرب تأثیر گذار بر تصویر را اصلاح کند. به همین دلیل از ایده فرا تفکیک‌پذیری در ترکیب با روش‌های فشرده‌سازی استفاده شد.

مبانی نظری مورد استفاده در پایان‌نامه در فصل ۳ معرفی و توضیح داده شده‌اند. در فصل ۴ روش پیشنهادی به همراه مراحل انجام روش توضیح داده شد و چند مثال نیز در این فصل آورده شد. در فصل ۵ نتایج بدست آمده با روش پیشنهادی براساس معیارهای OCR، MOS و PSNR در مقایسه با دیگر روش‌های فشرده‌سازی مورد ارزیابی قرار گرفت.

پایگاه داده‌ای شامل تصاویر متنی با درجه تفکیک مکانی بالا و متوسط با استفاده از تصاویر متنی واقعی ایجاد و روش پیشنهادی روی آنها اعمال شد. نتایج براساس معیارهای گفته‌شده ارزیابی و با هم مقایسه شدند. از نتایج گزارش شده در فصل ۵ می‌توان چنین نتیجه گرفت که روش پیشنهادی از نظر معیارهای OCR و MOS که در تصاویر متنی مهم هستند جواب بهتری نسبت روش‌های فشرده‌سازی دیگر داشت اما از نظر PSNR جواب خوبی بدست نیامد.

۶-۲ پیشنهاد راه‌کارهای آینده

در روند اجرای روش پیشنهادی عوامل متعددی روی نتایج بدست آمده تأثیر گذار بودند. در ادامه براساس تأثیر این عوامل راه‌کارهایی ارائه شده که ممکن است برای ادامه‌ی این کار مفید باشد.

- روش‌های فشرده‌سازی مختلف ویژگی‌های متفاوتی دارند و هر کدام تأثیر متفاوتی روی تصویر می‌گذارند. برای ادامه کار می‌توان به جز روش‌های انجام شده در این پایان‌نامه از روش‌های فشرده-سازی دیگر مانند روش‌های مبتنی بر مدل محتوای ترکیبی استفاده کرد.
- در این پایان‌نامه از ایده‌ی کاهش ابعاد به‌منظور رسیدن به میزان فشرده‌سازی بیشتر استفاده شد. اما کاهش ابعاد تصویر خود می‌تواند روی تصویر اثرات مخربی را داشته باشد. می‌توان از روش‌های کاهش ابعاد مختلف استفاده کرد. سپس نتایج بدست آمده از روش‌های مختلف را با هم مقایسه کرد و تأثیر هر روش کاهش ابعاد را روی تصاویر متنی دید و سعی در بهبود اثر آنها داشت. می‌توان

از روش‌های دیگری مانند تبدیل موجک، نزدیک‌ترین همسایه[1] و ... به‌منظور کاهش ابعاد استفاده کرد.

- در فرا تفکیک‌پذیری روش‌های متفاوتی وجود دارد. در این پایان‌نامه از روش‌های مبتنی بر درون-یابی و مبتنی بر آموزش استفاده شد. با روش اول جواب قابل قبولی بدست آمد اما با روش دوم جواب خوبی بدست نیامد. پیشنهاد می‌شود با در نظر گرفتن ویژگی‌های مربوط به تصاویر متنی از روش‌های مبتنی بر آموزش استفاده شود.

[1] Nearest Neighbor

مراجع

[۱] H. Grailu, M. Lotfizad, and H. S. Yazdi, "Farsi and Arabic Document Images Lossy Compression Based on the Mixed Raster Content Model," *International Journal on Document Analysis and Recognition (IJDAR),* Vol. 12, No. 4, pp. 227-248, 2009.

[۲] N. C. Francisco, N. M. M. Rodrigues, E. A. B. Silva, M. B. Carvalho, S. M. M. Faria, and V. M. M. Silva, "Scanned Compound Document Encoding Using Multiscale Recurrent Patterns," *IEEE Transactions on Image Processing,* Vol. 19, No. 10, pp. 2712-2724, 2010.

[۳] L. H. Sharpe and B. Manns, "JPEG 2000 Options for Document Image Compression," *Document Recognition and Retrieval IX*, Paul B. Kantor, Tapas Kanungo, Jiangying Zhou, Editors, Proceedings of SPIE, vol. 4670, pp. 167-173, 2002.

[۴] S. Dhawan, "A Review of Image Compression and Comparison of its Algorithms," *International Journal of Electronics and Communication Technology,* Vol. 2, No. 1, pp. 22-26, 2011.

[۵] J.Xiao, Ch.Wang, X.Hu, "Single Image Super-Resolution in Compressed Domain Based on Field of Expert Prior", 5th International Congress on Image and Signal Processing (CISP), PP. 607-611, China, Sichuan, 2012.

[۶] Nasrollahi, Kamal; Moeslund, Thomas B, "Super-resolution: A comprehensive survey", *Machine Vision & Applications,* Vol. 5, No. 3, pp. 223-226, 2014.

[۷] R.W.Gerchberg, "Super-resolution through error energy reduction", *International Journal of Optic*, Vol. 21, NO. 9, PP. 709–720, 1974.

[۸] P.D.Santis, F.Gori, "On an iterative method for super-resolution", *International Journal of Optic*, Vol. 22, NO. 8, PP. 691–695, 1975.

[٩] F.Champagnat, G.L.Besnerais, "A Fourier interpretation of super-resolution techniques", Proceedings of IEEE International Conference on Image Processing, Italy 1, pp. 865–868, 2005.

[١٠] D.O.Walsh, P.A.Nielsen-Delaney, "A direct method for superresolution", *International Journal of Optic*, Vol. 11, NO. 8, PP. 572–579, 1994.

[١١] R. Tsai and T. S. Huang, "Multiframe image restoration and registration," *Advances in computer vision and Image Processing*, vol. 1, pp. 317-339, 1984.

[١٢] V.H.Patil, D.S.Bormane, V.S.Pawar, "Super-resolution using neural network", Proceedings of IEEE 2nd Asia International Conference on Modeling and Simulation, Malaysia 2008.

[١٣] H.Demirel, G.Anbarjafari, "Image resolution enhancement by using discrete and stationary wavelet decomposition", *IEEE Transaction. Image Processing*, Vol. 20, NO. 5, PP. 1458–1460, 2011.

[١٤] P. Milanfar, "*Super-resolution imaging"*, CRC Press, 2010.

[١٥] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction", *Image Processing, IEEE Transactions on*, vol. 18, pp. 36-51, 2009.

[١٦] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution", *Image processing, IEEE Transactions on*, vol. 13, pp. 1327-1344, 2004.

[١٧] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction", *Image Processing, IEEE Transactions on*, vol. 16, pp. 349-366, 2007.

[١٨] M. Cristani, D. S. Cheng, V. Murino, and D. Pannullo, "Distilling information with super-resolution for video surveillance", in *Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks*, 2004, pp. 2-11.

[١٩] F. C. Lin, C. B. Fookes, V. Chandran, and S. Sridharan, "Investigation into optical flow super-resolution for surveillance applications", 2005.

[٢٠] F .Li, X. Jia, and D. Fraser, “Universal HMT based super resolution for remote sensing images”, in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, 2008, pp. 333-336.

[٢١] J. A. Maintz and M. A. Viergever, "A survey of medical image registration”, *Medical image analysis,* vol. 2, pp. 1-36, 1998.

[٢٢] L. M. Novak, G. J. Owirka, and A. L. Weaver, “Automatic target recognition using enhanced resolution SAR data”, *Aerospace and Electronic Systems, IEEE Transactions on,* vol. 35, pp. 157-175, 1999.

[٢٣] W. A.Pearlman, A.Said, “Set Partitioning Coding: Part I of Set Partition Coding and Image Wavelet Coding Systems”, Foundations and Trends in Signal Processing, Vol. 2, No. 2, pp. 95-180, 2008.

[٢٤] W.A.Pearlman, A.Said, “Image Wavelet Coding Systems: Part II of Set Partition Coding and Image Wavelet Coding Systems,” Foundations and Trends in Signal Processing, Vol. 2, No. 3, pp. 181-246, 2008.

[٢٥] W.A.Pearlman, A.Said, “Digital Signal Compression: Principles and Practice”, Cambridge University Press, New York, 2011.

[٢٦] A. Said and W. A. Pearlman, “A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 6, No. 3, pp. 243-250, 1996.

[٢٧] P.Mahesh, P.Rajesh, and I.Suneetha, “Improved Block-Based Segmentation for JPEG Compressed Document Images”, *International Journal of Research in Engineering and Technology*, Vol. 2, No. 11, pp. 669-673, 2013.

[٢٨] B.Oztan, A.Malik, Z.Fan, and R.Eschbach, “Removal of Artifacts from JPEG Compressed Document Images”, Proceedings of SPIE-IS&T Electronic Imaging, Vol. 6493, pp. 649306-1: 649306-9, 2007.

[٢٩] A.Zaghetto and R. L.Queiroz, "Scanned Document Compression Using a Block-Based Hybrid Video Codec", *IEEE Transactions on Image Processing*, Vol. 22, No. 6, pp. 2420-2428, 2010.

[٣٠] D.Taubman, "High Performance Scalable Image Compression with EBCOT", *IEEE Transactions on Image Processing*, 2000, Vol. 9, No. 7, pp. 1151-1170.

[٣١] G.Feng, C. A.Bouman, and H.Cheng, "high quality MRC document coding, Image Processing", Image Quality, Image Capture Systems Conference, PP. 320-325, 2001.

[٣٢] E.Y.Lam, "Compound Document Compression with Model-Based Biased Reconstruction", *Journal of Electronic Imaging (JEI)*, 2004, Vol. 13, No. 1, pp. 191-197.

[٣٣] R.L.deQueiroz, R.R.Buckley, M.Xu, "Mixed Raster Content (MRC) Model for Compound Image Compression", *Visual Communications and Image Processing*, Vol. 3653, pp. 1106-1117, 1998.

[٣٤] S.V.Khangar and L.G.Malik, "Handwritten Text Image Compression for Indic Script Document", *International Journal of Computer Applications*, Vol. 47, No. 5, pp. 11-16, 2012.

[٣٥] G.Feng and C. A.Bouman, "High-Quality MRC Document Coding", *IEEE Trans. Image Proc.*, Vol. 15, No. 10, pp. 3152-3169, 2006.

[٣٦] R.de Queiroz, R.Buckley, M.Xu, "Mixed Raster Content (MRC) Model for Compound Image Compression", *Proc. IS&T/SPIE Symposium on Electronic Imaging Science & Technology Visual Communications and Image Processing, San Jose, CA*, Vol. 3653, pp. 1106-1117, 1999.

[٣٧] Z.Fan and T.Jacobs, "Segmentation for Mixed Raster Contents with Multiple Extracted Constant Color Areas", *Proc. of SPIE-IS&T Electronic Imaging*, Vol. 5667, pp. 251-262, 2005.

[۳۸] D.P.Huttenlocher, P.F.Felzenswalb, W.Rucklidge, "DigiPaper: a Versatile Color Document Image Representation", International Conference on Image Processing (ICIP), pp. 219-223, 1999.

[۳۹] H.Grailu, M.Lotfizad, and H. S.Yazdi, "1-D Chaincode Pattern Matching for Compression of Bi-level Printed Farsi and Arabic Textual Images", *Image and Vision Computing (IVC)*, Vol. 27, No. 10, pp. 1615-1625, 2009.

[۴۰] H.Grailu, M.Lotfizad, and H.S.Yazdi, "An Improved Pattern Matching Technique for Lossy/Lossless Compression of Binary Printed Farsi and Arabic Textual Images", *International Journal of Intelligent Computing and Cybernetics*, Vol. 2, No. 1, pp. 120-147, 2009.

[۴۱] H.Grailu, M.Lotfizad, and H.S.Yazdi, "A Lossy/Lossless Compression Method for Printed Typeset Bi-level Text Images Based on Improved Pattern Matching", *International Journal on Document Analysis and Recognition (IJDAR)*, Vol. 11, No. 4, pp. 159-182, 2009.

[۴۲] A.Awajan, "Multilayer Model for Arabic Text Compression", *The International Arab Journal of Information Technology*, Vol. 8, No. 2, pp. 188-196, 2011.

[۴۳] L.Bottou, P.Haffner, P.G.Howard, P.Simard, Y.Bengio, and Y.LeCun, "High Quality Document Image Compression with DjVu", *J. Elec. Imaging*, Vol. 7, No. 3, pp. 410-425, 1998.

[۴۴] D.Huttenlocher, P.Felzenszwalb and W.Rucklidge, "*DigiPaper*: A Versatile Color Document Image Representation," *Proc. of IEEE Int'l Conf. on Image Proc.*, pp. 219–223, Kobe, Japan, October 1999.

[۴۵] K. U.Barthel, S. M.Partlin, and M.Thierschmann, "New Technology for Raster Document Image Compression", Part of the IS&T/SPIE Conference on Document Recognition and Retrieval VII, San Jose, CA, Vol. 3967, pp. 286-290, 2000.

[۴۶] M.Thierschmann, K.U.Barthel, and U.E.Martin, "A Scalable DSP-Architecture For High-Speed Color Document Compression", *Document Recognition and Retrieval VIII*, Paul B. Kantor, Daniel Lopresti P., Jiangying Z., Editors, Proceedings of SPIE, Vol. 4307, pp. 158-166, 2001.

[۴۷] B.F.Wu, C.C.Chiu, and Y.L.Chen, "Algorithms for Compressing Compound Document Images with Large Text/Background Overlap", *IEE Proc.-Vis. Image Signal Process.*, Vol. 151, No. 6, pp. 453-459, 2004.

[۴۸] E.Haneda, J.Yi, and C. A.Bouman, "Segmentation for MRC Compression", *Color Imaging XII: Processing, Hardcopy, and Applications*, edited by Reiner Eschbach, Gabriel G. Marcu, Proc. of SPIE-IS&T Electronic Imaging, Vol. 6493, pp. 252-262, 2007.

[۴۹] K.Hu, Z.Tang, L.Gao, and Y.Mu, "MC-JBIG2: an Improved Algorithm for Chinese Textual Image Compression", *International Journal on Document Analysis and Recognition (IJDAR)*, 2010, Vol. 13, No. 4, pp. 271-284.

[۵۰] R.L.deQueiroz, Z.Fan, and T. D.Tran, "Optimizing Block-Thresholding Segmentation for Multilayer Compression of Compound Images", *IEEE Trans. Image Proc.*, Vol. 9, No. 9, pp. 1461-1471, 2000.

[۵۱] H.Cheng and C. A.Bouman, "Document Compression Using Rate-Distortion Optimized Segmentation", *Journal of Electronic Imaging*, 2001, Vol. 10, No. 2, pp. 460–474.

[۵۲] H.Cheng, G.Feng, and C. A.Bouman, "Rate-Distortion Based Segmentation for MRC Compression", in *Proc. of SPIE Conf. on Color Imaging*: Device-Independent Color, Color Hardcopy and Applications, Vol. 4663, San Jose, CA, 21-23 January 2002.

[۵۳] P.Simard, H.Malvar, J.Rinker, and E.Renshaw, "A Foreground/Background Separation Algorithm for Image Compression", in *Data Compression Conference 2004*, pp. 498–507, 2004.

[۵۴] M. C. Chiang and T. E. Boulte, “Efficient super-resolution via image warping”, *Image and Vision Computing*, 18(10):761{771, 2000.

[۵۵] R. Fattal, “Image upsampling via imposed edges statistics”, *ACM Trans. on Graphics*, vol. 26, no. 3, pp. 95:1-95:8, 2007.

[۵۶] Q. Shan, et al., “Fast image/video upsampling”, *ACM Trans. on Graphics*, vol. 25, no. 5, pp. 153:1-153:7, 2008.

[۵۷] A. Levin, et al., “A closed-form solution to natural image matting”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228-242, 2008.

[۵۸] E. Mjolsness, “Fingerprint Hallucination”, *California Institute of Technology*, 1985.

[۵۹] D. Kong .M. Han, W. Xu, H. Tao, and Y. Gong, “A conditional random field model for video super-resolution”, in Pattern Recognition, ICPR, 18th International Conference on, pp. 619-622, 2006.

[۶۰] W. Rim, F. Drira, F. Lebourgeois, Chr. Garcia, M. Alimi, “Multiple Learned Dictionaries based Clustered Sparse Coding for the Super-Resolution of Single Text Image ”, 12th International Conference on Document Analysis and Recognition, 2013, pp. 484-488.

[۶۱] N. Nayef, J. Chazalon, P. Gomez-Krىmer, J-M. Ogier, “Efficient Example-based Super-Resolution of Single Text Images Based on Selective Patch Processing”, 11th IAPR International Workshop on Document Analysis Systems, 2014, pp. 227-231.

[۶۲] W. Rim, F. Drira, F. Lebourgeois, Chr. Garcia, M. Alimi, "Super-resolution of single text image by sparse representation", DAR 12, 2012, Mumbai, India.

[۶۳] W. Fan, J. Sun, S. Naoi, A. Minagawa, Y. Hotta, “Local Consistency Constrained Adaptive Neighbor Embedding for Text Image Super-Resolution”, 10th IAPR International Workshop on Document Analysis Systems, pp. 90-94, 2012.

[۶۴] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: a technical overview”, *Signal Processing Magazine, IEEE*, vol. 20, pp. 21-36, 2003.

[۶۵] T. Gotoh and M. Okutomi, “Direct super-resolution and registration using raw CFA images,” in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2004, pp. II-600-II-607 Vol. 2.

[۶۶] N. Otsu, “A threshold selection method from gray-level histograms”, *IEEE Trans. on Sys. Man. Cyber.* vol. 9, no. 1, pp. 62-66, 1979.

[۶۷] R.C.Gonzalez, R.E.Woods, Digital Image Processing, 3rd Edition, Pearson Prentice Hall, Upper Saddle River New Jersey, USA, 2008.

[۶۸] K. He, J. Sun, X. Tang, “Guided Image Filtering”, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 35, NO. 6, pp. 1397-1409, JUNE 2013.

[۶۹] N. Draper and H. Smith, “Applied Regression Analysis”, second ed. John Wiley, 1981.

[۷۰] T. Hastie, R. Tibshirani, and J.H. Friedman, “The Elements of Statistical Learning”. Springer, 2003.

[۷۱] C. Mancas-Thillou and M. Majid. ”Super-resolution text using the teager filter.” *First International Workshop on Camera-Based Document Analysis and Recognition*. pp. 10-16, 2005.

[۷۲] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation”, Technion—Israel Inst. of Technology, *Tech. Ref*, 2004.

[۷۳] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*, Springer, pp. 711-730, 2012.

[۷۴] A. N. Tikhonov and V. I. A. k. Arsenin, “Solutions of ill-posed problems”, Vh Winston, 1977.

[۷۵] E. W. Tramel and J. E. Fowler, "Video compressed sensing with multihypothesis," in *Data Compression Conference (DCC)*, pp. 193-202, 2011.

[۷۶] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259-268, 1992.

[۷۷] S. Farsiu, "A fast and robust framework for image fusion and enhancement," Citeseer, 2005.

[۷۸] Y.zheng, X.KangShatoLi, Y.HeJunSun, "Real-time Document Image Super-Resolution by Fast Matting" 11th IAPR International Workshop on Document Analysis System, 2014.

[۷۹] A. Levin, et al., "A closed-form solution to natural image matting", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228-242, 2008.

[۸۰] J. Wang and F. Cohen, "Optimized color sampling for robust matting", International Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[۸۱] J. Yang, J. Wright, T. Huang and Y. Ma, "Image superresolution via sparse epresentation of raw image patches", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[۸۲] J. Yang et al. "Image super-resolution via sparse representation", *IEEE Transactions on Image Processing*, Vol 19, Issue 11, pp2861-2873, 2010.

Abstract

In this thesis, a codec system is designed for high resolution textual images based on super resolution. Spatial resolution is very important in textual images, but this feature will increase the storage size of the images. Due to the limitation of digital memories and large size of these images, textual images compression seems necessary, But there are some problems; 1) Compression can cause destructive effects on textual images. 2) In methods like JPEG and JPEG2000, compression ratio or output bit rate, is limited.

The main idea in this thesis is to provide a good way to further compression ratio and also reduce the damaging effects on compressed image. To achieve greater compression ratio we use the idea of reducing the size of textual images. It should be noted that the resizing may has a effect on image quality and may cause degradation in image quality. So a procedure must be chosen that can increase the size of textual images and improve the devastating effects on images. In this result, the idea of super resolution combined with compression methods has been used.

The methods of super resolution based on interpolation and based on learning in the step of reconstruction, have been utilized. In the method based on interpolation, input low resolution image is divided into three layers and then each layer based on its importance, has been magnified with a special method. At last magnified layers have been combined to each other and have formed the final high resolution image.

In the method based on learning, at first a suitable data base of high resolution and low resolution textual images, have been caused. Fore making dictionary, 3*3 and 5*5 patches have been used. The dictionaries have been based on feature and four high pass filters for having high frequency components of images have been used to make dictionaries. Dictionaries have been learned for 50000 and 100000 atoms.

The created database including real document images with high and middle resolution. For an image that was scanning with spatial resolution 600dpi, JPEG and JPEG2000 methods can compression up to 0.07 and 0.015 bit rate respectively; proposed method for this image can compression up to 0.004 bit rate for both methods. The results based on Optical Character Recognition (OCR), Mean Opinion Score (MOS) and Peak

Signal-to-Noise Ratio (PSNR) criterions, have been evaluated and compared with each other. Proposed method based on OCR and MOS criterions that are important in textual images, has better results than all other methods but the PSNR for proposed method was not good.

Keywords: text image compression, JPEG compression, JPEG2000 compression, SPIHT compression, super resolution, Optical Character Recognition (OCR), High resolution text images.

Department of Electrical and Robatics Engineering

FOR THE DEGREE OF MASTER OF SCIENCE

Title:

Designing a CODEC system for high resolution textual images based on super resolution

Saeid Moradi

Supervisor

Dr. Hadi Grailu

September 2016