

## طراحی یک سیستم فشرده سازی/بازسازی تصاویر متنی با درجه‌ی تفکیک مکانی بالا مبتنی بر فرا تفکیک پذیری

سعید مرادی<sup>۱</sup>، فارغ التحصیل کارشناسی ارشد الکترونیک، هادی گرایلو<sup>۲</sup>، استادیار

۱- دانشکده مهندسی برق - دانشگاه صنعتی شاهرود - شاهرود - ایران - saeid006@gmail.com

۲- دانشکده مهندسی برق - دانشگاه صنعتی شاهرود - شاهرود - ایران - grailu@shahroodut.ac.ir

**چکیده:** در این مقاله یک سیستم فشرده سازی/بازسازی تصاویر متنی با درجه‌ی تفکیک مکانی بالا مبتنی بر فرا تفکیک پذیری پیشنهاد شده است. در روش پیشنهادی، برای رسیدن به میزان فشرده سازی بیشتر از ایده کاهش ابعاد در تصاویر متنی استفاده شده است. کاهش ابعاد در کنار عمل فشرده سازی ممکن است باعث تنزل در کیفیت تصویر شود. بنابراین باید روشی انتخاب شود که واحد بازسازی بتواند در کنار افزایش ابعاد تصویر، اثرات مخرب تأثیر گذار بر تصویر را نیز اصلاح کند. در مرحله بازسازی از روش فرا تفکیک پذیری استفاده شده است. در این روش، تصویر و ضوح پایین ورودی به سه لایه تقسیم و سپس هر لایه براساس اهمیت اطلاعاتی آن با یک روش خاص بزرگ‌نمایی شده است. در نهایت لایه‌های بزرگ‌نمایی شده با هم ترکیب و تصویر و ضوح بالای نهایی تشکیل شده است. یک ویژگی مهم روش پیشنهادی، قابلیت ترکیب آن با روش‌های فشرده سازی مختلف است. در این مقاله، ترکیب روش پیشنهادی با هر یک از روش‌های فشرده سازی JPEG، JPEG2000 و SPIHT بررسی و ملاحظه می شود، جواب قابل قبولی از نظر معیارهای باز شناسی متن (OCR) و متوسط امتیاز نظر سنجی (MOS) بدست آمده است؛ گرچه از نظر معیار قله سیگنال به نویز (PSNR) روش‌های دیگر بهتر از روش پیشنهادی عمل کرده‌اند.

**واژه‌های کلیدی:** فشرده سازی تصاویر متنی، فشرده سازی JPEG، فشرده سازی JPEG2000، فشرده سازی SPIHT، فراتفکیک پذیری، شناسایی نوری کاراکترها.

## Designing a Codec System for High Resolution Textual Images Based on Super Resolution

Saeid Moradi, Graduated<sup>1</sup>, Hadi Grailu, Assistant Professor<sup>2</sup>

1- Faculty of Electrical Engineering, Shahrood University of Technology, Shahrood, Iran, Email: saeid006@gmail.com

2- Faculty of Electrical Engineering, Shahrood University of Technology, Shahrood, Iran, Email: grailu@shahroodut.ac.ir

**Abstract:** In this paper, a CODEC system based on super resolution, is proposed for compression of high resolution textual images. It employs image resizing to decrease image dimensions and consequently, to improve the compression ratio; but at the other hand, it may reduce the image quality. Therefore, the decompression unit employs super resolution to simultaneously increase the reconstructed image dimensions and quality. In the employed interpolation-based super-resolution method, using an efficient textual image matting algorithm, the input low-resolution textual image is decomposed into three layers after which, each layer is enlarged using a particular method. Finally, the enlarged layers are combined to build the high resolution reconstructed textual image. An interesting feature of the proposed method is the ability to use existing compression methods such as JPEG, JPEG2000 and SPIHT. We have employed the aforementioned compression methods in the proposed CODEC system and evaluated the compression results with respect to OCR rate, Mean Opinion Score (MOS), and PSNR measures. Considering the OCR and MOS measures, the proposed method outperformed the others but not so with respect to PSNR.

**Keywords:** Textual image compression, JPEG compression, JPEG2000 compression, SPIHT compression, super resolution, optical character recognition.

تاریخ ارسال مقاله: ۱۴۰۰/۰۲/۱۵

تاریخ اصلاح مقاله: ۱۴۰۰/۱۱/۲۶

تاریخ پذیرش مقاله: ۱۴۰۰/۱۲/۱۳

نام نویسنده مسئول: هادی گرایلو

نشانی نویسنده مسئول: ایران - شاهرود - دانشگاه صنعتی شاهرود - دانشکده مهندسی برق - گروه الکترونیک.

## ۱- مقدمه

داد، زیرا چنین روش‌هایی از ویژگی‌های مخصوص به متن برای افزایش نرخ فشرده‌سازی استفاده نمی‌کنند.

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. پس می‌توان گفت که، طیف این نوع از تصاویر، طیفی میان‌گذر - یا شامل مولفه‌های فرکانسی میانی مهم - است اما طیف تصاویر طبیعی، پایین‌تر است [۳]. 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 [۴]. تفاوت دیگر بین تصاویر متنی و تصاویر طبیعی این است که تصاویر متنی، برخلاف بیشتر تصاویر طبیعی، شامل جزئیات دیداری مهم مانند شکل و انحنا، رنگ، بافت 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 [۵].

## ۲-۱- فشرده‌سازی تصاویر متنی

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) 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.

روش‌های فشرده‌سازی مبتنی بر تبدیل موجک جزو روش‌های امروزی و موثر محسوب می‌شوند [۶]. امروزه مشخص شده است که ترکیب تبدیل موجک و کدگذارهای بخش‌بندی مجموعه (SPC) [۸،۷] در حالت کلی، در بیشتر کاربردها کارایی بیشتری نسبت به دیگر انواع تبدیل‌ها (مانند DCT) و کدگذاری‌ها (مانند هافمن و حسابی و فقی) دارد [۶-۸]. کدگذارهای SPC عموماً از ویژگی وجود شباهت در زیرباندهای هم‌راسستا 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 [۹]. از جمله مهم‌ترین روش‌های کدگذاری SPC، روش کدگذاری SPIHT است [۱۰]. از این کدگذار تاکنون در

حجم عظیم تصاویر اسناد متنی که 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 این مقاله از سه روش فشرده‌سازی JPEG، JPEG2000 و SPIHT در ترکیب با روش پیشنهادی استفاده شده است. هر کدام از این روش‌ها در نرخ بیت‌های پایین می‌توانند موجب اثرات نامطلوبی مانند اثرات بلوکی یا نشتی روی تصاویر متنی شوند. از طرف دیگر در دو روش JPEG و JPEG2000 حداکثر میزان فشرده‌سازی قابل حصول محدود است. با توجه به اینکه تصاویر مورد بررسی، تصاویر با درجه‌ی تفکیک مکانی بالا بوده است و این تصاویر حجم فایل زیادی برای ذخیره‌سازی اشغال می‌کنند، هدف این است تا جایی که امکان دارد حجم فایل تصویر فشرده شده جهت ذخیره‌سازی کمتر شود.

با توجه به مشکلات ذکر شده برای روش‌های فشرده‌سازی استفاده شده، به‌منظور رسیدن به نرخ بیت‌های کمتر، ایده‌ی اصلی مورد استفاده در روش پیشنهادی کاهش ابعاد تصاویر متنی ورودی است؛ ابتدا تصویر ورودی را کاهش ابعاد داده و سپس آن را فشرده می‌کنیم. باید توجه داشت که روش‌های کاهش ابعاد نیز ممکن است موجب اثرات نامطلوبی در تصویر شوند. بنابراین در مرحله بازسازی نیاز به روشی است که بتواند هم اثرات نامطلوب ناشی از فشرده‌سازی و کاهش ابعاد را بهبود دهد و هم ابعاد تصویر را به ابعاد اولیه بازگرداند. برای این منظور در مرحله بازسازی از روش فراتفکیک پذیری استفاده شده است.

## ۲-۲- مرور روش‌های موجود

یک روش فشرده‌سازی تصاویر مبتنی بر روش فشرده‌سازی فراکتالی با استفاده از شبکه‌های عصبی مصنوعی در [۱] پیشنهاد شده است. استفاده از شبکه عصبی به‌منظور سرعت بخشیدن به روش فشرده‌سازی فراکتالی در مرحله فشرده‌سازی بوده است. یک روش فشرده‌سازی بدون اتلاف تصاویر متنی مبتنی بر ترکیب روش کدگذاری طول اجرا\* و روش کدگذاری هافمن در [۲] پیشنهاد شده است.

روش‌های امروزی موجود برای فشرده‌سازی تصاویر طبیعی را نمی‌توان برای فشرده‌سازی تصاویر اسناد با نرخ فشرده‌سازی زیاد مورد توجه قرار

مانند فراتفکیک پذیری مرتفع یا تعدیل ساخت، می توان به میزان فشرده سازی بیشتری دست یافت.

## ۲-۲- فرا تفکیک پذیری

فرا تفکیک پذیری فرایند بدست آوردن یک یا چند تصویر وضوح بالا از روی یک یا چند تصویر وضوح پایین است. فرا تفکیک پذیری را می توان براساس معیارهای مختلفی طبقه بندی کرد. اگر براساس معیار تعداد تصاویر با وضوح پایین روش ها را طبقه بندی کنیم، می توان دو گروه شامل روش های مبتنی بر چند تصویر و روش های مبتنی بر تک تصویر را داشت. فراتفکیک پذیری تک تصویر در دو دهه اخیر توجه بسیاری از محققان را به خود جلب کرده است. الگوریتم های فرا تفکیک پذیری تک تصویر در حوزه ی تصاویر متنی را می توان به دو دسته تقسیم کرد: (۱) روش های مبتنی بر درون یابی (۲) روش های مبتنی بر یادگیری.

## ۲-۲-۱- روش های مبتنی بر یادگیری

در روش های مبتنی بر یادگیری، رابطه ی میان وصله های وضوح بالا و وصله های وضوح پایین با استفاده از یک پایگاه داده خارجی یاد گرفته می شود. سپس این رابطه به وصله های وضوح پایین تصویر ورودی اعمال می شود و یک تصویر وضوح بالا تولید می شود. این مسئله همانند حل یک دستگاه معادلات است به طوری که تعداد مجهولات از تعداد معادلات بیشتر باشد. در این حالت بینهایت جواب برای دستگاه معادله وجود خواهد داشت و همین موضوع مسئله فرا تفکیک پذیری را بدحالت می کند. برای حل این مشکل به یک سری فرضیات و اطلاعات پیشین نیاز است. روش های متفاوتی برای یادگیری رابطه ی میان وصله های وضوح بالا و پایین ارائه شده اند. موفقیت روش های فرا تفکیک پذیری مبتنی بر یادگیری به دو فاکتور مهم وابسته است: (۱) جمع آوری یک پایگاه داده مناسب شامل جفت وصله های وضوح بالا و پایین (۲) یاد گرفتن رابطه بین وصله های وضوح بالا و پایین. در [۱۷] یک روش فراتفکیک پذیری برای تصاویر متنی با استفاده از لایه بندی تصویر در کنار شبکه های مولد تخا صمی<sup>†</sup> پیشنهاد شده است. منبع [۱۸] یک روش فراتفکیک پذیری مبتنی بر آموزش برای تصاویر متنی پیشنهاد می دهد. در این روش از اطلاعات مخصوص به تصاویر متنی که از یک مدل تشخیص متن و به صورت دنباله ای از احتمالات پیشین هر کاراکتر انتخاب شده است استفاده می کنند. استفاده از اطلاعات پیشین متنی در مرحله بازسازی تصویر با درجه تفکیک مکانی بالا مفید است. در نهایت از فراتفکیک پذیری مبتنی بر اطلاعات پیشین چند سطحی برای روش پیشنهادی استفاده می شود. در [۱۹] یک مدل فراتفکیک پذیری تک تصویر براساس مدل واقعی تصویر برداری دوربین های دیجیتال ارائه شده است.

فشرده سازی تصاویر مختلفی مانند تصاویر هوایی، تصاویر پزشکی، تصاویر عمومی و نیز برای سیگنال هایی مانند سیگنال های قلبی و مغزی استفاده شده است [۱۸]. یک روش فشرده سازی تصاویر متنی مبتنی بر تبدیل موجک و کدگذاری SPIHT به منظور حفظ یا بهبود کیفیت تصویر، خوانایی و تشخیص نوری کاراکترها در [۱۱] پیشنهاد شده است.

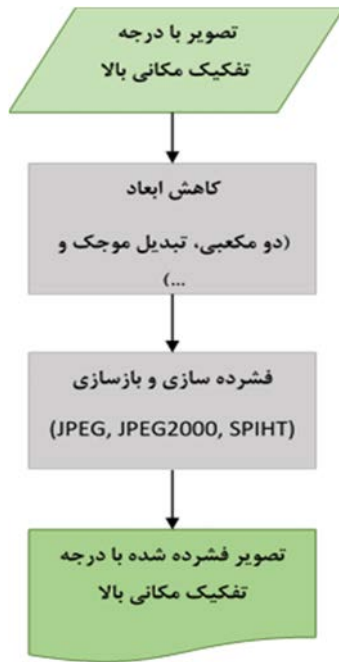
در [۱۲] یک روش قطعه بندی بلوکی تصاویر متنی به منظور فشرده سازی این نوع از تصاویر معرفی شده است. در این روش از استاندارد JPEG برای کدگذاری استفاده شده است. روش پیشنهاد شده در [۱۳] به دنبال کاهش یا حذف اثرات نامطلوب ایجاد شده از اعمال استاندارد JPEG که ویژه ی فشرده سازی تصاویر طبیعی است، روی تصاویر متنی می باشد. استانداردهایی که تاکنون در کاربرد فشرده سازی تصاویر متنی استفاده شده اند، محدود به استانداردهای فشرده سازی تصاویر نمی شوند. برای نمونه، در [۱۴] از کدک ویدیویی H.264/AVC برای فشرده سازی تصاویر متنی استفاده شده است. یک روش فشرده سازی تصاویر متنی در نرخ بیت های پایین بر اساس کدگذاری ناحیه مورد علاقه در [۱۵] ارائه شده است. در این روش از کدگذار SPIHT در چارچوب فشرده سازی ناحیه مورد علاقه استفاده شده است. به منظور بهبود اثرات ناشی حاصل از روش SPIHT از روش های بهبود تصاویر استفاده شده است.

در تصاویر متنی، کیفیت لبه های بازیابی شده در تصاویر بازسازی شده (پس از فشرده سازی) مهم است. روش های عمومی فشرده سازی تصاویر و نیز برخی روش های تخصصی فشرده سازی تصاویر متنی به این نکته توجهی نشان ندادند. برای مثال در فشرده سازی به کمک استاندارد JPEG، اثر بلوکی و حذف فرکانس های میانی به بالا مشاهده می شود. لبه های تصاویر متنی عمده تاً متناظر با فرکانس های میانی می باشند. در خروجی روش JPEG که مبتنی بر تبدیل DCT است، به دلیل بلوک بندی تصویر، اثر بلوکی شدن مشاهده می شود. همچنین این روش به دلیل کوانتیزاسیون انجام شده در آن، عملکردی معادل با یک فیلتر پایین گذر دارد [۳].

در استاندارد JPEG2000 از تبدیل موجک و کدگذار EBCOT که نوعی کدگذار بخش بندی مجموعه است [۱۶]، استفاده شده است. موجک ها از ویژگی حضور محلی و نیز ویژگی متراکم سازی انرژی برخوردار هستند؛ بنابراین در JPEG2000، میزان فشرده سازی بیشتر شده و کیفیت لبه ها نیز بهتر حفظ می شود؛ اما متأسفانه به دلیل وقوع اثر ناشی در نرخ بیت های پایین، کیفیت لبه ها و نیز کیفیت دیداری تصویر، آفت پیدا می کنند [۳].

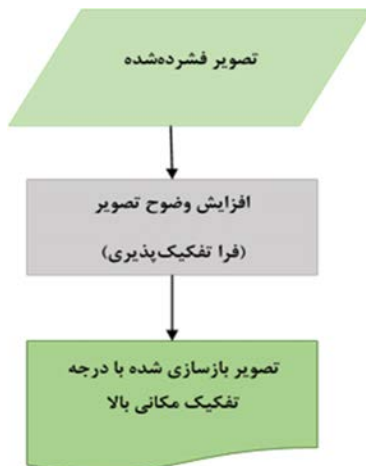
در روش های فشرده سازی JPEG2000 و SPIHT مهم ترین عامل محدودیت ساز در رسیدن به مقادیر بزرگ میزان فشرده سازی، اثر ناشی است. لذا اگر بتوان این عامل آفت کیفیت را با برخی روش ها

<sup>†</sup> Generative Adversial Networks



شکل ۱. مرحله فشرده سازی روش پیشنهادی

دهد. مطابق با شکل ۲، تصویر فشرده شده بعنوان ورودی مرحله بازسازی در نظر گرفته می شود. در تنها بلوک مهم مورد استفاده مرحله بازسازی، به منظور افزایش وضوح و نیز افزایش ابعاد تصویر فشرده شده از روش فرا تفکیک پذیری استفاده می شود. در مرحله بازسازی از روش فرا تفکیک پذیری مبتنی بر درون یابی استفاده شده است. در روش استفاده شده ابتدا تصویر ورودی به سه زیر لایه شامل؛ (۱) لایه شامل اطلاعات لایه تصویر متن (۲) لایه پیش زمینه و (۳) لایه پس زمینه تجزیه می شود. سپس با توجه به اهمیت هر لایه، آن لایه به روشی خاص بزرگ نمایی می شود. سپس لایه ها با هم ترکیب شده و تصویر وضوح بالای نهایی بدست می آید.



شکل ۲. مرحله برگشت

## ۲-۲-۲- روش های مبتنی بر درون یابی

این روش ها، روش های ساده و از لحاظ محاسباتی، روش های مؤثری هستند. در این روش ها مدل تصویربرداری ساده در نظر گرفته می شود [۲۱، ۲۰]. در واقع در این روش ها به نوعی ترکیبی از الگوریتم های درون یابی و بهبود تصاویر را داریم. عبارت دیگر یا تصویر را با روش های معمول درون یابی مانند دو مکعبی بزرگ می کنند و سپس تصویر بزرگ شده را با تأکید بر لایه های متن بهبود می دهند و یا برعکس، ابتدا تصویر را بهبود می دهند و سپس بزرگ نمایی را انجام می دهند.

اخیراً از روش های درون یابی مبتنی بر ویژگی های آماری لایه به منظور فرا تفکیک پذیری مبتنی بر درون یابی استفاده شده است. اما این روش ها، روش هایی زمان بر هستند [۲۲]. در [۲۳] روش افزایش ابعاد ساده اما مؤثری برای بهبود خودکار وضوح تصویر/ویدئو، در حالی که اطلاعات ساختاری اساسی را حفظ می کند، پیشنهاد شده است. مزیت اصلی روش پیشنهاد شده چارچوب بازخورد کنترلی است که به خوبی اطلاعات تصویر وضوح بالا را بدون در نظر گرفتن هیچ آموزشی از داده ورودی باز یابی می کند. این امر باعث می شود که روش پیشنهادی هیچ گونه وابستگی به کیفیت و تعداد نمونه های انتخاب شده که مسئله رایج در روش های مبتنی بر آموزش است، نداشته باشد. مزیت دیگر روش پیشنهاد شده در [۲۳] سرعت بالای آن است.

در [۲۴] یک الگوریتم جدید زمان واقعی لایه بندی مبتنی بر مدل خطی-محلی به منظور تجزیه تصویر متنی ورودی به لایه های مختلف آن طراحی شده است. در این روش برای افزایش ابعاد هر لایه یک روش بزرگ نمایی خاص در نظر گرفته شده است. در نهایت تصویر متنی با وضوح بالای نهایی حاصل از فرا تفکیک پذیری با ترکیب لایه های وضوح بالای ایجاد شده، بدست می آید.

## ۳- روش پیشنهادی

روش پیشنهادی شامل دو بخش، مرحله اول شامل فشرده سازی تصویر متنی با درجه تفکیک مکانی بالا (مرحله فشرده سازی) و مرحله دوم شامل بازسازی تصویر متنی فشرده شده و سپس رسیدن به تصویر با درجه تفکیک مکانی بالای اولیه است (مرحله بازسازی). در مرحله فشرده سازی مطابق با شکل ۱، ابتدا تصویر ورودی که با درجه تفکیک مکانی بالا روبش شده است بوسیله یکی از روش های کاهش ابعاد، ابعادش کاهش داده می شود. سپس، تصویر حاصل از مرحله کاهش ابعاد با یکی از روش های فشرده سازی، فشرده می شود. از جمله مزیت های روش پیشنهادی قابلیت ترکیب این روش با انواع مختلف روش های فشرده سازی موجود است.

در رسیدن به نرخ بیت های پایین، اعمال فشرده سازی و تکنیک پیشنهادی کاهش ابعاد، اثرات مخرب و نامطلوبی روی تصویر متنی ورودی ایجاد می کنند. بنابراین در مسیر بازسازی نیاز به روشی است که بتواند بطور همزمان هم اثرات مخرب و نامطلوب ایجاد شده را بهبود دهد و هم ابعاد تصویر فشرده شده را به ابعاد تصویر ورودی اولیه افزایش

پایین نواحی متن  $\alpha'$  است. بدست آوردن  $\alpha'$ ،  $F'$  و  $B'$  با داشتن تنها تصویر ورودی  $I'$  یک مسأله نامعین است. برای حل این مسأله روش‌های لایه‌بندی متفاوتی ارائه شده است [۲۶، ۲۴]. این روش‌ها به‌طور معمول زمان‌بر می‌باشند. در [۲۵] از یک روش کارا به‌منظور لایه‌بندی استفاده می‌شود.

### ۳-۱-۱- 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. از روش اتسو به منظور دودویی نمودن تصویر ورودی استفاده می‌شود [۲۵]. به‌منظور استخراج اطلاعات لبه تصویر متن اصلی، از فیلتر حفظ لبه یعنی فیلتر هدایت [۲۷] استفاده می‌شود. کلید اصلی فیلتر هدایت 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:

$$q_i = a_k I_i + b_k, \quad \forall i \in w_k \quad (2)$$

در رابطه (۲)  $a_k$  و  $b_k$  ضرایب خطی ثابت در  $w_k$  هستند. پنجره  $w_k$  به صورت مربعی و به شعاع  $r$  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  $n$  مانند نویز از ورودی  $p$  مدل می‌شود.

$$q_i = p_i - n_i \quad (3)$$

راه‌حلی که در نظر گرفته شده است، روشی است که در رابطه (۳) اختلاف بین  $p$  و  $q$  را کمینه کند در حالیکه مدل خطی (۲) را شامل شود. به‌طور خاص، تابع هزینه ارائه شده در رابطه (۴) در پنجره  $w_k$  کمینه شده است.

$$E(a_k, b_k) = \sum_{i \in w_k} ((a_k I_i + b_k - p_i)^2 + \epsilon a_k^2) \quad (4)$$

$\epsilon$  پارامتر تنظیم جهت  $a_k$ ‌های بزرگ است. رابطه (۵) رگرسیون خطی ستیغی<sup>‡</sup> است که در [۲۹، ۲۸] مدل شده است و راه‌حل آن بصورت زیر است:

$$a_k = \frac{1}{|w| \sum_{i \in w_k} I_i p_i - \mu_k \overline{p_k}} \quad (5)$$

$$b_k = \overline{p_k} - a_k \mu_k \quad (6)$$

در روابط (۵) و (۶)،  $\mu_k$  و  $\sigma_k^2$  به ترتیب میانگین و واریانس  $I$  در  $w_k$  هستند.  $|w|$  تعداد پیکسل‌ها در  $w_k$  و  $\overline{p_k} = 1/|w| \sum_{i \in w_k} p_i$  میانگین

در  $P$  است. با بدست آوردن ضرایب خطی  $a_k$  و  $b_k$  می‌توان خروجی  $q_i$  فیلتر را با استفاده از رابطه (۲) محاسبه کنیم. شکل ۴ فرایند

### ۳-۱- 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  $\alpha$  به‌عنوان وزن مطابق رابطه (۱) مدل شود [۲۵].

$$I = \alpha F + (1 - \alpha) B \quad (1)$$

در رابطه (۱)،  $\alpha$  ماتریسی است که هر پیکسل آن می‌تواند مقداری در بازه  $[0, 1]$  داشته باشد. اگر  $\alpha_{i,j} = 0$  در این صورت پیکسل واقع در مختصات  $(i,j)$  پس‌زمینه و اگر  $\alpha_{i,j} = 1$  پیکسل واقع در مختصات  $(i,j)$  را پیش‌زمینه می‌نامند. در غیر این صورت پیکسل ترکیبی از پیش‌زمینه و پس‌زمینه است. لایه‌بند تصویر با هدف بدست آوردن لایه  $\alpha$ ، پیش‌زمینه  $F$  و پس‌زمینه  $B$  روی تصویر ورودی  $I$  اعمال می‌شود.

شکل ۳ مثالی از لایه‌بندی تصویر متنی را نشان می‌دهد. شکل ۳(الف) تصویر ورودی را نشان می‌دهد. شکل ۳(ب) لایه  $\alpha$  را بعد از لایه‌بندی تصویر نشان می‌دهد. شکل‌های ۳(ج) و ۳(د) به ترتیب لایه‌های پیش‌زمینه و پس‌زمینه را نشان می‌دهند. لایه  $\alpha$  شامل اطلاعات مهم مربوط به نواحی متنی و لایه‌های پیش‌زمینه و پس‌زمینه نیز به ترتیب شامل اطلاعات مربوط به پیش‌زمینه و پس‌زمینه هستند. همان‌طور که از شکل ۳ مشخص است، لایه  $\alpha$  شامل نواحی هموار متن می‌باشد. لایه‌های پیش‌زمینه و پس‌زمینه تنها شامل اطلاعات رنگ متن و پس‌زمینه هستند. این نکته برای فرا تفکیک‌پذیری تصاویر متنی بسیار مفید است و به جلوگیری از مصنوعات و ماتی با پردازش جداگانه روی لبه متن و اجزای رنگ کمک می‌کند.

همان‌طور که گفته شد اولین مرحله از روش ارائه شده، تجزیه تصویر ورودی  $I'$  به سه لایه متفاوت شامل لایه وضوح پایین رنگ پیش‌زمینه  $F'$ ، لایه وضوح پایین رنگ پس‌زمینه  $B'$  و لایه وضوح



شکل ۳ مثالی از لایه‌بندی تصویر متنی. (الف) تصویر ورودی. (ب) لایه  $\alpha$ . (ج) لایه پیش‌زمینه. (د) لایه پس‌زمینه. [۲۵]

با کمینه کردن خطای بازسازی و عبارت تنظیم، بازسازی پیش‌زمینه و پس‌زمینه می‌تواند با کمینه کردن تابع انرژی رابطه (۱۲) حل شود [۲۵]:

$$\sum_{i \in W_k} ((F_k - B_k) \alpha_i + B_k - I_i)^2 + \gamma (F_k - B_k)^2 \gamma (F_k - B_k)^2 \quad (12)$$

رابطه (۱۲) بیان‌گر عبارت تنظیم است و  $\gamma$  ثابتی است که معمولاً برابر ۰.۰۱ تنظیم می‌شود. تابع انرژی نمایش داده شده در رابطه (۱۲) می‌تواند در شکل بسته حل شود و حل آن به صورت رابطه (۱۳) و (۱۴) است [۳۰]:

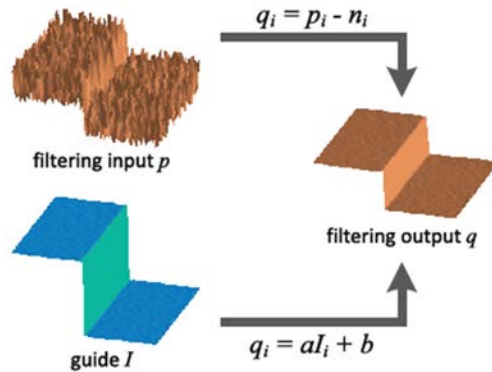
$$F_k - B_k = \frac{1}{\sigma_k^2 + \gamma} \sum_{i \in W_k} \alpha_i I_i - \mu_k \bar{I}_k \quad (13)$$

$$B_k = \bar{I}_k - (F_k - B_k) \mu_k \quad (14)$$

در رابطه (۱۳)  $W_k$  پنجره  $3 \times 3$  است زیرا استفاده از پنجره بزرگ‌تر ممکن است فرض خطی محلی را نقض کند.  $\mu_k$  و  $\sigma_k^2$  در روابط (۱۳) و (۱۴) به ترتیب میانگین و واریانس  $\alpha$  در  $W_k$  می‌باشند، تعداد پیکسل‌ها در پنجره محلی است،  $i$  شماره پیکسل‌ها در همسایگی  $W_k$  و میانگین شدت روشنایی پیکسل‌های  $I$  در همسایگی  $W_k$  می‌باشد.

### ۳-۲-۳- بزرگ‌نمایی جداگانه لایه‌ها و ترکیب آن‌ها

این قسمت از روش فرا تفکیک‌پذیری شامل افزایش ابعاد لایه‌های به‌دست آمده از قسمت قبل است. همان‌طور که در شکل ۳(ج) و ۳(د) دیده می‌شود، لایه‌های پیش‌زمینه و پس‌زمینه به‌طور معمول هموار هستند و اطلاعات مفید کمتری دارند درحالی‌که لایه‌ی حاوی اطلاعات لایه متن نشان داده شده در شکل ۳(ب) شامل عمده اطلاعات مربوط به شکل متن است. بنابراین برای افزایش ابعاد هر یک از سه لایه پیش‌زمینه، پس‌زمینه و لایه حاوی اطلاعات متن آلفا، روش‌های مختلفی در نظر گرفته می‌شود. با توجه به اهمیت لایه شامل اطلاعات لایه متن، ابعاد این لایه با تأکید بر تقویت لایه‌ها بزرگ شده‌است. از طرف دیگر به دلیل اینکه لایه‌های پیش‌زمینه و پس‌زمینه از اهمیت اطلاعاتی کمی در رابطه با ویژگی‌های متن برخوردار هستند، از روش درون‌یابی دومکعبی برای بزرگ‌نمایی آنها استفاده می‌شود. چون درون‌یابی دومکعبی لایه‌های تصویر متنی را مات می‌کند بنابراین لایه‌ی حاوی اطلاعات لایه متن را قبل از درون‌یابی ابتدا بهبود می‌دهیم. برای بهبود لایه‌های تصویر متنی از فیلتر تیگر [۳۱] استفاده می‌کنیم. فیلتر تیگر، فیلتری با تأکید بر تقویت لایه‌های تصویر متنی و از بین بردن نویز است. این فیلتر به منظور تقویت لایه‌های متن مورد استفاده قرار گرفته شده است. فیلتر تیگر برای یک پیکسل تصویر  $l$  به مختصات  $(i, j)$  در رابطه (۱۵) معرفی شده است:



شکل ۴ فیلتر هدایت

فیلتر هدایت را نشان می‌دهد. 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 لایه‌های تصویر متن اصلی در نظر بگیرد و اطمینان دهد که در نواحی محلی تغییرات تصویر ورودی و تصویر پوشش با هم برابر است یعنی  $\nabla A = \nabla I$ . اگر  $G$  را معرف فیلتر هدایت در نظر بگیریم، لایه‌ی حاوی اطلاعات لایه متن  $a'$  را می‌توان از رابطه (۷) به‌دست آورد:

$$a' = GAI; r, \epsilon \quad (7)$$

در رابطه (۷) دو پارامتر  $r$  و  $\epsilon$  به ترتیب شعاع همسایگی و پارامتر تنظیم ضرایب فیلتر هدایت هستند [۲۷]. شکل ۳(ب) لایه حاوی اطلاعات لایه متن  $a'$  را نشان می‌دهد.

### ۳-۱-۳- لایه‌های پیش‌زمینه و پس‌زمینه

مدل خطی محلی بیان می‌کند که در پنجره‌های محلی کوچک، کانال  $\alpha$  می‌تواند به‌صورت ترکیب خطی از تصویر  $I$  نمایش داده شود.

$$\alpha_k = ZI_k^c + H \quad k \in \omega_i \quad (8)$$

در رابطه (۸)  $\omega_i$  پنجره کوچکی حول پیکسل  $i$  و  $c$  بیان‌گر کانال رنگ است.  $H$  و  $G$  در پنجره محلی، ثابت می‌باشند. برای سادگی، در ادامه شاخص کانال  $c$  حذف می‌شود. براساس رابطه (۱) و (۸)، می‌توان  $Z$  و  $H$  را به‌صورت زیر به‌دست آورد. با جای‌گذاری رابطه (۸) در رابطه (۱) داریم:

$$I = ZIF + HF + B - ZIB - HB \quad (9)$$

با در نظر گرفتن  $I = 0$  برای مقادیر پس‌زمینه و  $I = 1$  برای مقادیر پیش‌زمینه و جای‌گذاری این مقادیر در رابطه بالا داریم:

$$I = 0 \rightarrow H(F - B) = -B \rightarrow H = \frac{B}{B - F} \quad (10)$$

$$I = 1 \rightarrow 1 = Z(F - B) + H(F - B) + B \quad (11)$$

$$1 = Z(F - B) + \frac{-B}{F - B}(F - B) + B \rightarrow Z = \frac{1}{F - B}$$

در این مقاله از سه روش فشرده سازی JPEG2000، JPEG و SPIHT در ترکیب با روش پیشنهادی استفاده شده است. به منظور فشرده سازی تصاویر از نرم افزارهای Advanced JPEG Compressor، JPEG2000 Compressor و روش SPIHT [۱۰] استفاده شده است. به منظور ارزیابی روش پیشنهادی، معیارهای باز شنا سی متن (OCR)، متوسط امتیاز نظرسنجی (MOS) و پیک سیگنال به نویز (PSNR) مورد استفاده قرار گرفته اند.

#### معیار OCR

برای OCR، از نرم افزار ABBY Fine Reader استفاده شده است. برای بدست آوردن صحت OCR، تعداد کاراکترهایی که نرم افزار معرفی شده به درستی تشخیص داده است به تعداد کل کاراکترها تقسیم می شود.

#### معیار MOS

برای معیار MOS از ۲۰ نفر خواسته شد که به تصاویری که به آنها نشان داده می شود از نظر خوانایی و کیفیت از ۱ تا ۵ نمره دهند. به این صورت که ۵ بهترین نمره ای است که می توان به یک تصویر با کیفیت عالی داد و ۱ ضعیف ترین نمره ای است که می توان به یک تصویر داد. در نهایت برای هر تصویر کل ۲۰ نمره ای که به آن اختصاص داده شده است را میانگین گیری کرده و نمره نهایی برای تصویر مورد نظر گزارش می شود.

#### معیار PSNR

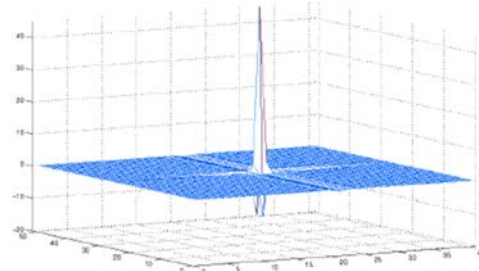
این معیار میزان نزدیک بودن تصویر بازسازی شده به تصویر اصلی را از لحاظ عددی نشان می دهد. برای محاسبه PSNR از رابطه (۱۵) استفاده می شود:

$$psnr = 10 \log_{10} \left( \frac{255^2 mn}{\|\hat{X} - I\|^2} \right) \quad (15)$$

در رابطه (۱۵)،  $I$  و  $\hat{X}$  به ترتیب تصویر اصلی و تصویر بازسازی شده و  $mn$  تعداد پیکسل های تصویر اصلی هستند.

$$y^T(i, j) = 3y^2(i, j) - \frac{1}{2}y(i+1, j+1)y(i-1, j-1) - \frac{1}{2}y(i+1, j-1)y(i-1, j+1) - y(i+1, j)y(i-1, j) - y(i, j+1)y(i, j-1) \quad (15)$$

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



شکل ۵ فیلتر تیگر [۳۱]

فیلتر تیگر یک فیلتر غیرخطی است که ما را قادر می سازد تا لبه های تصویر متن را تقویت و نویز را حذف کنیم. پس از بهبود لایه حاوی اطلاعات لبه متن، به منظور بزرگنمایی این لایه از درون یابی دومکعبی استفاده می کنیم. افزایش نمونه های لایه لبه متن به صورت رابطه (۱۶) انجام می شود:

$$\alpha^h = \text{Bicubic}(\alpha^l + \beta \text{Teager}(\alpha^l)) \quad (16)$$

در رابطه (۱۶)  $\beta$  ضریب وزن دهی برای کنترل تیزی لبه ها است که معمولاً برابر ۱ قرار داده می شود [۲۵].  $\text{Teager}()$  فیلتر تیگر و  $\text{Bicubic}()$  درون یابی دومکعبی را نشان می دهد. پس از آنکه لایه های وضوح بالای  $\alpha^h$  و  $F^h$  به دست آمدند، تصویر وضوح نهایی با ترکیب این لایه ها و به صورت رابطه (۱۷) بدست می آید [۲۵]:

$$I^h = \alpha^h F^h + (1 - \alpha^h) B^h \quad (17)$$

#### ۴- نتایج

جهت تولید داده های آموزشی، پایگاه داده ای شامل ۱۰۱ تصویر از تصاویر متنی واقعی در درجه ی تفکیک مکانی ۶۰۰ dpi و ۳۰۰ dpi ایجاد شد. در شکل های ۶ و ۷ به ترتیب تصاویری از پایگاه داده که با درجه ی تفکیک ۳۰۰ dpi و ۶۰۰ dpi روبش شده اند، نشان داده شده اند. برای نشان دادن اختلاف بین این تصاویر، قسمتی از این تصاویر در شکل های ۸ و ۹ بزرگنمایی شده اند.

به منظور اجرای الگوریتم فراتفکیک پذیری از کد نوشته شده در نرم افزار متلب استفاده شده است که در سیستمی با مشخصات پردازنده Intel Core i5-3210M @ 2.5GHz و 4GB RAM اجرا شده است و ضریب کاهش ابعاد برای نمودارهای گزارش شده در این تحقیق ۵ می باشد. زمان اجرای روش پیشنهادی به طور میانگین ۵.۲۳ ثانیه بوده است. نتایج ارائه شده در این مقاله بهترین پاسخ به دست آمده از اجرای روش پیشنهادی روی تصاویر پایگاه داده می باشد.



شکل ۶- تصویر ۳۰۰ dpi از پایگاه داده ایجاد شده

در شکل‌های ۱۱ تا ۱۳ نتایج حاصل از ترکیب روش پیشنهادی به ترتیب با روش‌های فشرده‌سازی JPEG، JPEG2000 و SPIHT در نرخ بیت‌های ۰/۱، ۰/۱۵ و ۰/۱ نشان داده شده است. با دقت در این شکل‌ها که برای دیدن بهتر اثر روش‌های فشرده‌سازی و کارایی روش پیشنهادی بزرگ‌نمایی شده است، مشخص است که روش پیشنهادی در ترکیب با هر یک از روش‌های فشرده‌سازی بکار گرفته شده، توانسته است اثرات مخرب ناشی از این روش‌ها و همچنین اثرات مخرب ناشی از کاهش ابعاد را به خوبی بهبود دهد. در شکل ۱۱ روش پیشنهادی توانسته است بخوبی اثرات بلوکی ناشی از روش فشرده‌سازی JPEG را بهبود دهد. شکل ۱۲ اثرات ناشی از اعمال روش فشرده‌سازی JPEG2000 روی تصویر ورودی را در نرخ بیت ۰/۱۵ نشان می‌دهد. کارایی روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG2000 در این شکل دیده می‌شود. در شکل ۱۳ اثرات ناشی از اعمال روش فشرده‌سازی مبتنی بر تبدیل موجک و سپس کدگذاری SPIHT روی تصویر ورودی در نرخ بیت ۰/۱ قابل مشاهده است. روش پیشنهادی در ترکیب با روش فشرده‌سازی SPIHT توانسته است اثرات ناشی مخرب ناشی از این روش فشرده‌سازی را بهبود دهد.

sentences

الف

sentences

ب

sentences

ج

شکل ۱۱ بزرگ‌نمایی بخشی از تصویر در ترکیب روش پیشنهادی با JPEG (الف) تصویر اصلی (ب) تصویر فشرده‌شده با JPEG. در نرخ بیت ۰/۱ (ج) روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG در نرخ بیت ۰/۱

sentences

الف

sentences

ب

sentences

ج

شکل ۱۲ بزرگ‌نمایی بخشی از تصویر در ترکیب روش پیشنهادی با JPEG2000 (الف) تصویر اصلی (ب) تصویر فشرده‌شده با JPEG2000 در نرخ بیت ۰/۱۵ (ج) ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG2000 در نرخ بیت ۰/۱۵

- $k$ -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 clustering 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  $\theta_j$  (see the practical hints that follow). If the initial values for  $\text{arg } m_j$  of  $\theta_j$  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 one  $\theta_j$ 's.
- Accurate estimation of the number of clusters (representatives) is crucial for the algorithm, since a poor estimate will prevent it from unravelling 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 fit 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  $\theta_j$ 's are (a) random initialization, (b) random selection of  $m$  data vectors from  $X$  as the initial estimates of  $\theta_j$ '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  $n_{\text{trials}}$  times (each time using different initial values) and determine the clustering (among the  $n_{\text{trials}}$  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 (other than some 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 FMM algorithm, discussed later).

Example 7.5.1. Generate and plot a data set,  $X_N$ , 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 ۷- تصویر ۶۰۰dpi از پایگاه داده ایجاد شده

This procedure is

شکل ۸ تصویر ۳۰۰dpi بزرگ‌نمایی شده از پایگاه داده

This procedure is

شکل ۹ تصویر ۶۰۰dpi بزرگ‌نمایی شده از پایگاه داده

در شکل ۱۰ به ترتیب اثر روش‌های فشرده‌سازی JPEG، SPIHT و زمانی که این روش‌ها بطور مستقیم روی تصاویر متنی با درجه تفکیک ۶۰۰dpi اعمال شوند، نشان داده شده است. همان‌طور که از این شکل مشخص است، در نرخ بیت‌های پایین روش فشرده‌سازی JPEG روی تصویر متنی اثرات بلوکی و روش‌های فشرده‌سازی JPEG2000 و SPIHT اثر ناشی ایجاد می‌کنند.

sentences

الف

sentences

ب

sentences

ج

sentences

د

شکل ۱۰ اثرات ناشی از روش‌های فشرده‌سازی مختلف روی تصاویر متنی (الف) تصویر اصلی (ب) اثر بلوکی حاصل از JPEG با نرخ بیت ۰/۰۹ (ج) اثر ناشی حاصل از JPEG2000 با نرخ بیت ۰/۰۸ (د) اثر ناشی فشرده‌سازی به کمک SPIHT با نرخ بیت ۰/۰۸





جدول ۱: مقادیر میانگین OCR ترکیب روش پیشنهادی با روش های فشرده-

**سازي روی تصاویر با درجه تفکیک مکانی ۳۰۰dpi**

نرخ بیت	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰
JPEG	۹۹.۰	۹۹.۰	۹۹.۰	۹۸.۰	۹۸.۰	۹۷.۰	۹۷.۰	۹۶.۰	۹۶.۰	۰
JPEG 2000	۹۹.۰	۹۹.۰	۹۹.۰	۹۹.۰	۹۹.۰	۹۹.۰	۹۹.۰	۹۸.۰	۹۶.۰	۶۲
SPIHT	۹۹.۰	۹۹.۰	۹۹.۰	۹۹.۰	۹۹.۰	۹۸.۰	۹۸.۰	۹۸.۰	۹۷.۰	۷۱

جدول ۲: مقادیر انحراف معیار OCR ترکیب روش پیشنهادی با روش های

**فشرده سازی روی تصاویر با درجه تفکیک مکانی ۳۰۰dpi**

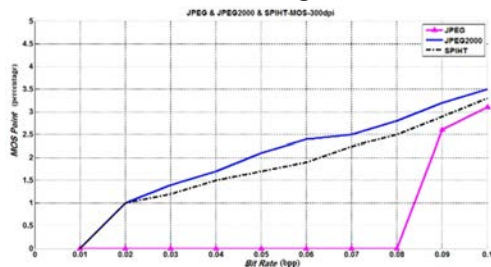
نرخ بیت	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰	۰.۰
JPEG	۰.۴	۰.۵	۰.۷	۰.۷	۰.۸	۰.۹	۱.۰	۱.۱	۰	۰
JPEG 2000	۰.۱	۰.۲	۰.۲	۰.۳	۰.۳	۰.۳	۰.۵	۰.۶	۰.۸	۰.۸
SPIHT	۰.۰	۰.۲	۰.۴	۰.۵	۰.۶	۰.۶	۰.۷	۰.۷	۰.۸	۰.۸

شده است. در این حالت هم استفاده از روش پیشنهادی موجب بهبود عملکرد فشرده سازی و بازشناسی شده است.

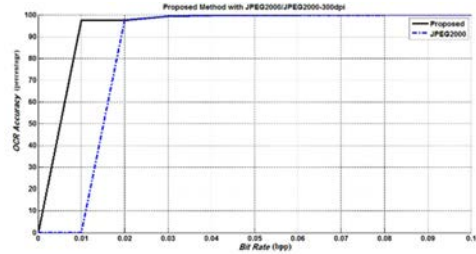
در جداول ۱ و ۲ به ترتیب مقادیر مربوط به میانگین و انحراف معیار نتایج حاصل از OCR روش پیشنهادی در ترکیب با سه روش فشرده سازی JPEG، JPEG2000 و SPIHT روی تصاویر با درجه تفکیک مکانی ۳۰۰ dpi ارائه شده است. جداول ۳ و ۴ به ترتیب مقادیر مربوط به میانگین و انحراف معیار نتایج حاصل از OCR روش پیشنهادی در ترکیب با سه روش فشرده سازی JPEG، JPEG2000 و SPIHT روی مجموعه تصاویر با درجه تفکیک مکانی ۶۰۰ dpi را ارائه می دهند.

#### ۴-۲- ارزیابی روش پیشنهادی براساس معیار MOS

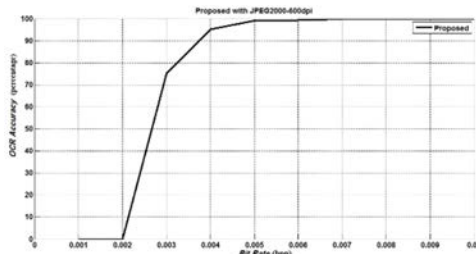
در شکل های ۲۲ و ۲۳ نتایج محاسبه معیار MOS حاصل برای حالت اعمال روش های JPEG، JPEG2000 و SPIHT به تنهایی روی دو تصویر به ترتیب با درجه تفکیک مکانی ۳۰۰ dpi و ۶۰۰ dpi نشان داده شده است. نتایج نشان می دهند که افراد کیفیت روش JPEG2000 را در مقایسه با دو روش دیگر در نرخ بیت های قابل مقایسه، مطلوب تر



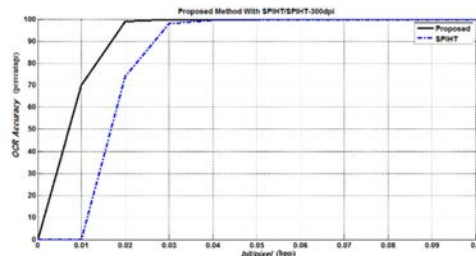
شکل ۲۲ مقایسه روش های فشرده سازی از نظر معیار MOS برای تصویر با درجه تفکیک مکانی ۳۰۰dpi



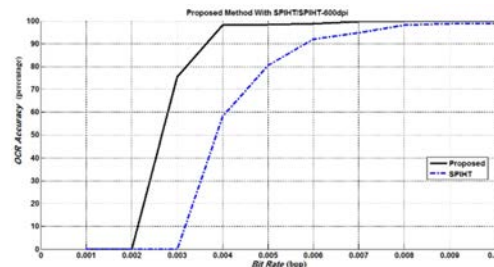
شکل ۱۸ نتایج OCR روش پیشنهادی در ترکیب با فشرده سازی JPEG2000، روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



شکل ۱۹ نتایج OCR روش پیشنهادی در ترکیب با فشرده سازی JPEG2000، روی تصویر با درجه تفکیک مکانی ۶۰۰dpi



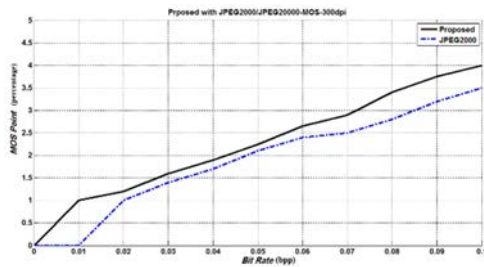
شکل ۲۰ نتایج OCR روش پیشنهادی در ترکیب با فشرده سازی SPIHT روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



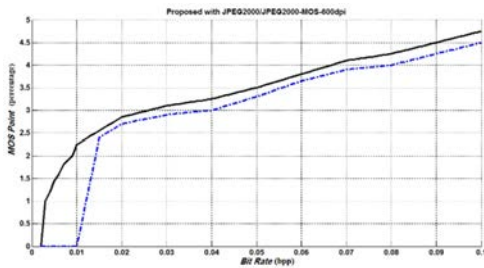
شکل ۲۱ نتایج OCR روش پیشنهادی در ترکیب با فشرده سازی SPIHT روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

از مقایسه شکل های ۱۹ و ۱۵ می توان دید روش پیشنهادی در ترکیب با JPEG2000 در مقایسه با اعمال روش فشرده سازی JPEG2000 به تنهایی روی تصویری با درجه تفکیک مکانی ۶۰۰ dpi، توانسته نرخ بیت را از ۰/۰۱۵ به ۰/۰۰۴ برساند.

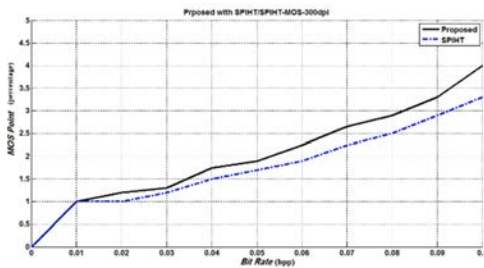
در شکل های ۲۰ و ۲۱ نتایج حاصل از ترکیب روش پیشنهادی با روش SPIHT در مقایسه با اعمال روش SPIHT به تنهایی نشان داده



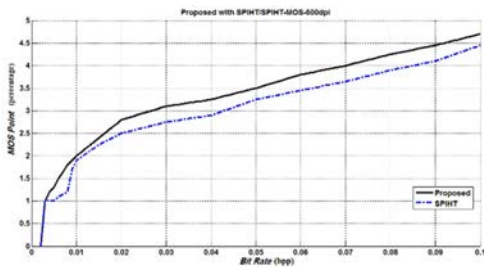
شکل ۲۶ نتایج MOS روش پیشنهادی در ترکیب با روش JPEG2000 روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



شکل ۲۷ نتایج MOS روش پیشنهادی در ترکیب با روش JPEG2000 روی تصویر با درجه تفکیک مکانی ۶۰۰dpi



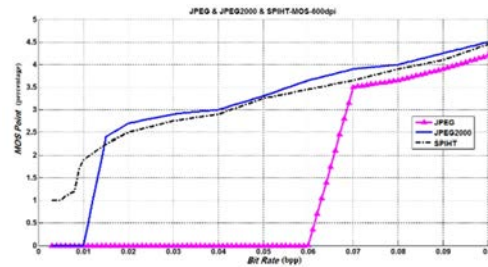
شکل ۲۸ نتایج MOS روش پیشنهادی در ترکیب با روش SPIHT روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



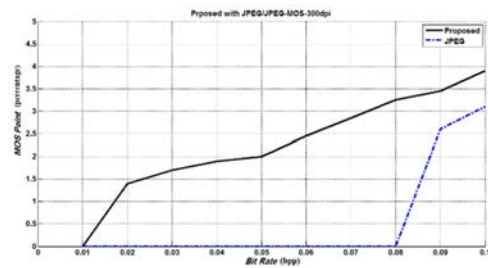
شکل ۲۹ نتایج MOS روش پیشنهادی در ترکیب با روش SPIHT روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

نمودارهای مربوط به ارزش‌های معیار MOS روش پیشنهادی در ترکیب با روش فشرده‌سازی SPIHT برای تصاویر با درجه تفکیک مکانی ۳۰۰ dpi و ۶۰۰ dpi به ترتیب در شکل‌های ۲۸ و ۲۹ 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 به تنهایی است.

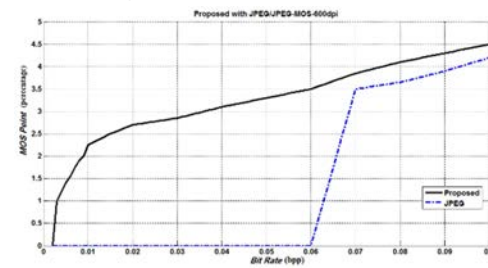
همان‌طور که از این شکل‌ها مشخص است روش پیشنهادی در ترکیب با روش SPIHT برای هر دو تصویر از نظر MOS جواب بهتری را داشته است.



شکل ۲۳ مقایسه روش‌های فشرده‌سازی از نظر معیار MOS برای تصویر با درجه تفکیک مکانی ۶۰۰dpi



شکل ۲۴ نتایج MOS روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



شکل ۲۵ نتایج MOS روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

می‌دانسته‌اند. در شکل‌های ۲۴ و ۲۵ نتایج حاصل از ترکیب روش پیشنهادی با فشرده‌سازی JPEG در مقایسه با اعمال روش JPEG به تنهایی 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.

همان‌طور که از این شکل‌ها دیده می‌شود، ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG از نظر خوانایی انسانی در جامعه ۲۰ نفره ذکر شده، نتایج مطلوب‌تری نسبت به روش فشرده‌سازی JPEG به تنهایی داشته است.

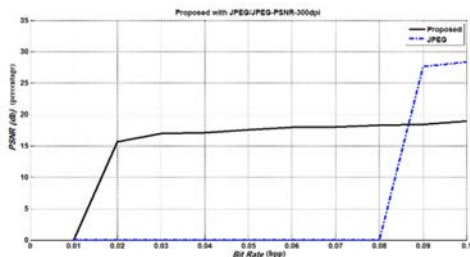
در شکل‌های ۲۶ و ۲۷ نتایج حاصل از ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG2000 در مقایسه با اعمال روش فشرده‌سازی JPEG2000 به تنهایی نشان داده شده است. همان‌طور که از این شکل‌ها مشخص است روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG2000 توانسته است خوانایی تصویری فشرده‌شده ورودی را بهبود دهد. در شکل‌های ۲۶ و ۲۷، خط ممتد مربوط به روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG2000 و خط-نقطه مربوط به روش فشرده‌سازی JPEG2000 به تنهایی است.

### ۳-۴- ارزیابی روش پیشنهادی براساس معیار PSNR

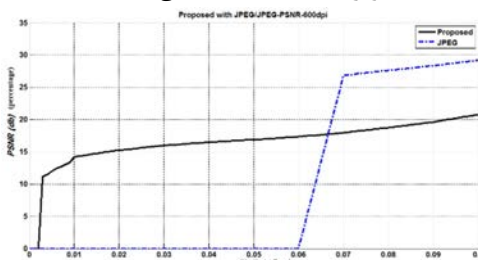
این معیار در واقع میزان نزدیک بودن تصویر بازسازی شده و تصویر اصلی را نشان می‌دهد. نتایج گزارش شده برای این معیار براساس رابطه (۱۵) بدست آمده است. در ابتدا نتایج محاسبه معیار PSNR حاصل از اعمال روش‌های فشرده‌سازی JPEG2000، SPIHT و JPEG به‌تنهایی گزارش شده‌اند. سپس نتایج حاصل از ترکیب روش پیشنهادی با هر یک از روش‌های فشرده‌سازی گفته شده گزارش شده است.

شکل‌های ۳۰ و ۳۱ نتایج حاصل از اعمال روش‌های فشرده‌سازی گفته‌شده به‌تنهایی را روی تصاویر براساس معیار PSNR نشان می‌دهند. از نتایج بدست آمده در شکل‌های ۳۰ و ۳۱ 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.

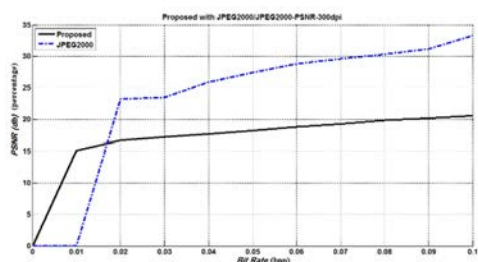
در شکل‌های ۳۲ و ۳۳ نتایج حاصل از ترکیب روش پیشنهادی با روش JPEG در مقایسه با اعمال روش فشرده‌سازی JPEG به‌تنهایی بر مبنای معیار PSNR نشان داده شده است. نتایج بدست آمده برای روش پیشنهادی براساس معیار PSNR در مقایسه با اعمال روش فشرده‌سازی JPEG به‌تنهایی ضعیف‌تر بوده است. دلیل این نتایج این است که در روش 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 محاسبه PSNR نقاطی که در تصویر بازسازی شده به ضخامت حروف افزوده شده‌اند، خطای بازسازی را نسبت به تصویر اصلی زیاد کنند.



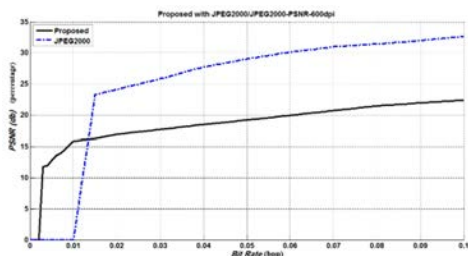
شکل ۳۲ نتایج PSNR روش پیشنهادی در ترکیب با روش JPEG، روی تصویر با درجه تفکیک مکانی ۳۰۰dpi



شکل ۳۳ نتایج PSNR روش پیشنهادی در ترکیب با روش JPEG، روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

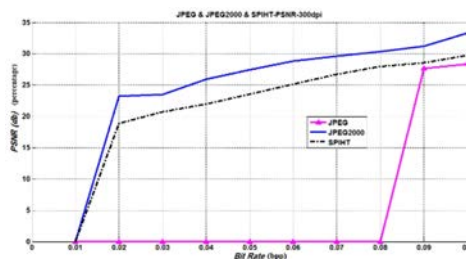


شکل ۳۴ نتایج PSNR روش پیشنهادی در ترکیب با روش JPEG2000، روی تصویر با درجه تفکیک مکانی ۳۰۰dpi

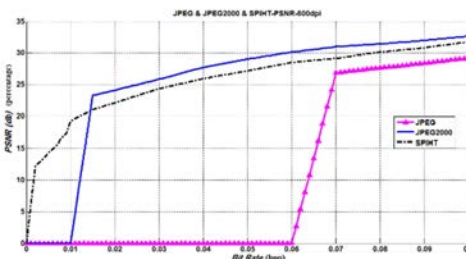


شکل ۳۵ نتایج PSNR روش پیشنهادی در ترکیب با روش JPEG2000، روی تصویر با درجه تفکیک مکانی ۶۰۰dpi

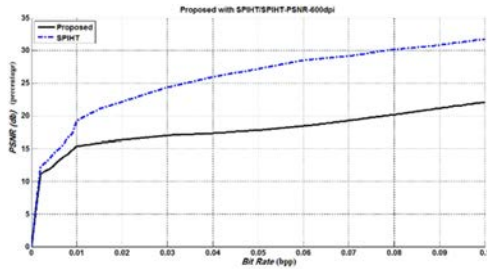
شکل‌های ۳۴ و ۳۵ نتایج حاصل از ترکیب روش پیشنهادی با روش فشرده‌سازی JPEG2000 را در مقایسه با اعمال روش فشرده‌سازی JPEG2000 به‌تنهایی بر مبنای معیار PSNR نشان می‌دهند. از این شکل‌ها نیز می‌توان دید که به دلیل گفته شده در بالا، روش پیشنهادی در ترکیب با روش فشرده‌سازی JPEG2000 از نظر معیار PSNR نتایج ضعیف‌تری نسبت به روش فشرده‌سازی JPEG2000 به‌تنهایی دارد. نتایج حاصل از ترکیب روش پیشنهادی با روش SPIHT بر مبنای معیار PSNR در شکل‌های ۳۴ و ۳۵ نشان داده شده است.



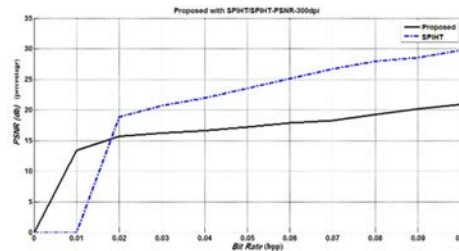
شکل ۳۰ مقایسه روش‌های فشرده‌سازی از نظر معیار PSNR برای تصویر با درجه تفکیک مکانی ۳۰۰dpi



شکل ۳۱ مقایسه روش‌های فشرده‌سازی از نظر معیار PSNR برای تصویر با درجه تفکیک مکانی ۶۰۰dpi



شکل ۳۷ نتایج PSNR روش پیشنهادی در ترکیب با روش SPIHT روی تصویر با درجه تفکیک مکانی ۶۰۰dpi



شکل ۳۶ نتایج PSNR روش پیشنهادی در ترکیب با روش SPIHT روی تصویر با درجه تفکیک مکانی ۳۰۰dpi

جدول ۳: مقادیر میانگین OCR ترکیب روش پیشنهادی با روش‌های فشرده‌سازی روی تصاویر با درجه تفکیک مکانی ۶۰۰dpi

نرخ بیت	۰.۰۰۴	۰.۰۰۵	۰.۰۰۶	۰.۰۰۷	۰.۰۰۸	۰.۰۰۹	۰.۰۱	۰.۰۱۵	۰.۰۲	۰.۰۳	۰.۰۴	۰.۰۵	۰.۰۶	۰.۰۷	۰.۰۸
JPEG	۸۶.۷	۸۷.۴۱	۸۷.۸۴	۹۷.۶۵	۹۷.۸۴	۹۸.۱۳	۹۸.۴۸	۹۸.۸۱	۹۹.۰۵	۹۹.۲۱	۹۹.۴۳	۹۹.۶۱	۹۹.۷	۹۹.۸۴	۹۹.۹۴
JPEG2000	۹۳.۷۶	۹۷.۶۷	۹۷.۹۸	۹۸.۵۶	۹۸.۷۸	۹۹.۰۴	۹۹.۲۵	۹۹.۵۵	۹۹.۶۴	۹۹.۷۱	۹۹.۷۸	۹۹.۸۴	۹۹.۹	۹۹.۹۸	۹۹.۹۹۸
SPIHT	۹۶.۹۵	۹۷.۷۶	۹۷.۹۹	۹۸.۶۱	۹۸.۸۸	۹۸.۹۶	۹۹.۱۴	۹۹.۳۴	۹۹.۴۸	۹۹.۵۸	۹۹.۶۵	۹۹.۸۴	۹۹.۸۸	۹۹.۹۷	۹۹.۹۹

جدول ۴: مقادیر انحراف معیار OCR ترکیب روش پیشنهادی با روش‌های فشرده‌سازی روی تصاویر با درجه تفکیک مکانی ۶۰۰dpi

نرخ بیت	۰.۰۰۴	۰.۰۰۵	۰.۰۰۶	۰.۰۰۷	۰.۰۰۸	۰.۰۰۹	۰.۰۱	۰.۰۱۵	۰.۰۲	۰.۰۳	۰.۰۴	۰.۰۵	۰.۰۶	۰.۰۷	۰.۰۸
JPEG	۱.۲۴۳	۱.۱۲۷	۰.۹۷۸	۰.۷۵۲	۰.۷۰۱	۰.۶۵۳	۰.۵۷۲	۰.۵۲۰	۰.۴۶۹	۰.۴۱۵	۰.۳۶۶	۰.۳۰۳	۰.۲۳۸	۰.۱۲۴	۰.۰۴
JPEG2000	۱.۱۱۱	۰.۷۶۱	۰.۷۰۳	۰.۶۵۱	۰.۶۱۳	۰.۵۵۸	۰.۵۰۳	۰.۴۴۱	۰.۳۵۲	۰.۲۸۱	۰.۲۱۳	۰.۱۱۳	۹۹.۹	۰.۰۱۱	۰.۰۰۲
SPIHT	۱.۰۰۳	۰.۸۷۱	۰.۸۰۱	۰.۷۲۱	۰.۶۶۲	۰.۶۰۱	۰.۵۴۱	۰.۵۰۰	۰.۴۵۲	۰.۳۴۱	۰.۲۶۳	۰.۱۳۴	۰.۱۰۱	۰.۰۱	۰.۰۰۵

نتایج بدست آمده با روش پیشنهادی براساس معیارهای OCR، MOS و PSNR در مقایسه با دیگر روش‌های فشرده‌سازی شامل JPEG، JPEG2000 و SPIHT مورد ارزیابی قرار گرفت. روش پیشنهادی براساس معیارهای OCR و MOS و افزایش میزان فشرده‌سازی تصاویر متنی نسبت به روش‌های فشرده‌سازی استفاده شده نتایج بهتری را داشته است اما از نظر معیار PSNR روش پیشنهادی نسبت به روش‌های استفاده شده ضعیف‌تر بوده است که دلیل این امر ضخیم شدن حروف در مرحله بازسازی است.

## مراجع

- [1] D. W. Naaman, "Image Compression Technique Based on Fractal Image Compression Using Neural Network – A Review", *AJRCoS*, vol. 10, no. 4, pp. 47-57, Jul. 2021.
- [2] H. H Huda, "Lossless Text Image Compression using Two Dimensional Run Length Encoding", *Journal Online Anformatika (JOIN)*, vol. 4, n0. 2, pp.75-78, 2019.
- [3] 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 (IJ DAR)*, vol. 12, no. 4, pp. 227-248, 2009.
- [4] 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.
- [5] 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.

از شکل‌های ۳۶ و ۳۷ نیز دیده می‌شود که روش فشرده سازی SPIHT به‌تنهایی از نظر معیار PSNR نسبت به روش پیشنهادی در ترکیب با روش SPIHT جواب بهتری داشته است.

## ۵- نتیجه‌گیری و پیشنهادات

فشرده سازی 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. علاوه بر مشکلات ذکر شده، برای روش‌های فشرده سازی مانند JPEG و JPEG2000 مشکل دیگر، محدود بودن میزان فشرده سازی یا نرخ بیت خروجی است. در این مقاله روشی جهت فشرده سازی تصاویر متنی با درجه تفکیک مکانی بالا پیشنهاد شد. ایده اصلی روش پیشنهادی در این مقاله این بوده‌است که با ارائه روشی مناسب بتوان میزان فشرده سازی را بیشتر و اثرات مخرب روی تصویر فشرده شده را نیز کم کرد. برای رسیدن به میزان فشرده سازی بیشتر از ایده کاهش ابعاد در تصاویر متنی استفاده شد. 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 در این مقاله از ایده فرا تفکیک‌پذیری در ترکیب با روش‌های فشرده سازی استفاده شد.

- [18] J. Ma, Sh. Guo, L. Zhang, "Text Prior Guided Scene Text Image Super-resolution", arXiv:2106.15368, 2021.
- [19] J. Cai, et al., "Toward Real-World Single Image Super-Resolution: A New Benchmark and A New Model", arXiv:1904.00523, 2019.
- [20] M. C. Chiang and T. E. Boult, "Efficient super-resolution via image warping", *Image and Vision Computing*, vol. 18, no. 10, pp. 761-771, 2000.
- [21] S. Farsiu, M. D. Robinson, M. Elad and P. Milanfar, "Fast and robust multiframe super resolution", *IEEE Transactions on Image processing*, vol. 13, pp. 1327-1344, 2004.
- [22] R. Fattal, "Image upsampling via imposed edges statistics", *ACM Transaction on Graphics*, vol. 26, no. 3, pp. 95:1-95:8, 2007.
- [23] Q. Shan, et al., "Fast image/video upsampling", *ACM Transaction on Graphics*, vol. 25, no. 5, pp. 153:1-153:7, 2008.
- [24] A. Levin, et al., "A closed-form solution to natural image matting", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228-242, 2008.
- [25] Y. zheng, X. Kang Shato Li and Y. He Jun Sun, "Real-time document image super-Resolution by fast matting" 11<sup>th</sup> IAPR International Workshop on Document Analysis System, 2014.
- [26] J. Wang and F. Cohen, "Optimized color sampling for robust matting", International Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.
- [27] K. He, J. Sun, X. Tang, "Guided image filtering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397-1409, 2013.
- [28] Khangar S. V. and Malik L. G., Handwritten Text Image Compression for Indic Script Document, 2012, International Journal of Computer Applications, Vol. 47, No. 5, pp. 11-16.
- [29] de Queiroz R., Buckley R. and Xu M., 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. 1106-1177, 1999.
- [30] J. Yang, J. Wright, T. Huang and Y. Ma, "Image superresolution via sparse representation of raw image patches", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
- [31] 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.
- [6] 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.
- [7] W. A. Pearlman and A. Said, "Set partitioning coding: part I of set partitioning coding and image wavelet coding systems", *Foundations and Trends in Signal Processing*, vol. 2, no. 2, pp. 95-180, 2008.
- [8] W. A. Pearlman and A. Said, "Image wavelet coding systems: part II of set partitioning coding and image wavelet coding systems," *Foundations and Trends in Signal Processing*, vol. 2, no. 3, pp. 181-246, 2008.
- [9] W. A. Pearlman and A. Said, *Digital Signal Compression: Principles and Practice*, Cambridge University Press, New York, 2011.
- [10] 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.
- [11] H. Grailu, "Textual Image Compression for Maintaining or Improving the Recognition Performance", *Circuits Syst Signal Process*, no. 36, pp. 658-674, 2017.
- [12] 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.
- [13] 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.
- [14] A. Zagherro 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.
- [15] H. Grailu, "Textual image compression at low bit rates based on region-of-interest coding", *IJDAR*, no. 19, pp. 65-81, 2016.
- [16] D. Taubman, "High performance scalable image compression with EBCOT", *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1151-1170, 2000.
- [17] K. Lim, Y. Liu, T.H. Li, Sh. Liu, G. Li, . "Text Image Super-Resolution by Image Matting and Text Label Supervision", *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, 2019.