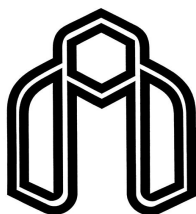


اللَّهُ الرَّحْمَنُ الرَّحِيمُ



دانشگاه صنعتی شاهرود

دانشکده مهندسی برق و رباتیک

پایان نامه دوره کارشناسی ارشد مهندسی برق-الکترونیک

طراحی سیستم کد کننده/کد گشا برای فشرده سازی تصاویر متنی مبتنی بر مدل محتوای

ترکیبی

نگارش:

آیدین خدائشناس پرچینکی

استاد راهنما:

دکتر هادی گرایلو

استاد مشاور:

دکتر حسین خسروی

پایان نامه جهت اخذ درجه کارشناسی ارشد

بهمن ۹۱

دانشگاه صنعتی شاهرود

دانشکده برق و رباتیک

گروه الکترونیک

پایان نامه کارشناسی ارشد آقای آیدین خدشناس

تحت عنوان: طراحی سیستم کد کننده/کد گشا برای فشرده سازی تصاویر متنی مبتنی بر مدل محتوای ترکیبی

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

امضاء	اساتید مشاور	امضاء	اساتید راهنما
	نام و نام خانوادگی :		نام و نام خانوادگی :
	نام و نام خانوادگی :		نام و نام خانوادگی :

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

تقدیم به :

مادرم که با عاطفه سرشار شور و هم را از تنهایر و نومیدر رهایر مر
بفسد و پدرم که وجود پر مهرش در سردترین دورانی، تسلر بفسر
جانیهاست و همه کسانر که در راه علم آموزر مشوق مع بوده اند و
معلمانر که دانتر اند کم را دریوخ آخ ها هستم.

سپاس و قدردانی:

سپاس و حمد پروردگار را که توان اندر سیدرخ به ما عطا فرمود. اکنون که در ایغ مقطع از تحصیلات به درجه کارشناس ارشد نائل آمده‌ام بر خود لازم مردانم از کسانی که مرا در ایغ مسیر یار نموده اند تشکر و قدر دانم نمایم. از معلمان که مرابا الفبا را علمشان نمودند و به مع قدرت تفکر و تعمق آموختند.

از استاد دوره کارشناس ارشد علم الفصول جناب دکتر هادی گرایلو به پاس راهنمایی‌ها بر دروغشاخ و دکتر حسین خسرو که به عنوان استاد راهنما و مشاور همکار نزدیگر با بنده داشته و مرا از نظرات سودمندش بهره مند ساختند و همچنین از دوست عزیزم مهرداد مصطفی صفایر و خانواده مصطفی‌شاخ که در ایغ سالها از صحبت بیدروغشاخ بهره مند بودم، کمال تشکر و سپاس را دارم.

تعهد نامه

اینجانب آیدین خدانشناس دانشجوی دوره کارشناسی ارشد رشته الکترونیک-دیجیتال دانشکده برق و رباتیک دانشگاه صنعتی شاهرود نویسنده پایان نامه طراحی سیستم کد کننده/کد گشای فشرده سازی تصاویر متنی مبتنی بر مدل محتوای ترکیبی تحت راهنمایی جناب دکتر هادی گرایلو متعهد می شوم .

- تحقیقات در این پایان نامه 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 می باشد 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 پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- 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

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .

چکیده:

امروزه، بیشتر تولیدات تاریخی، هنری و علمی روی کاغذ منتشر می‌شوند. بسیاری از کتابخانه‌های بزرگ دنیا در حال تبدیل منابع خود به شکل دیجیتال می‌باشند زیرا ذخیره اطلاعات مربوط به اسناد مختلف به شکل دیجیتالی، تا حد زیادی مقرون به صرفه‌تر از نگهداری آنها در قفسه‌های کتابخانه است. بنا به دلایل اقتصادی، برای ذخیره چنین حجم عظیمی از اطلاعات به فشرده‌سازی آنها نیاز داریم.

هدف در این پایان‌نامه ارائه روشی مبتنی بر مدل محتوای ترکیبی پایه جهت فشرده سازی تصاویر متنی می‌باشد. تصویر سند در این روش معمولاً به سه لایه پیش‌زمینه، پس‌زمینه و پوشش تجزیه می‌شود.

در راستای اجرای این پایان‌نامه اقدام به جمع آوری ۱۴۴۹ تصویر متنی شد. این تصاویر طیف وسیعی از انواع رسم‌الخط و کیفیت و ترکیب متن و گرافیک را شامل می‌شود. برای ایجاد لایه پوشش مورد نیاز در مدل محتوای ترکیبی از یک روش جدا سازی متن از تصویر استفاده شد که جهت تصاویر متنی سطح خاکستری مورد استفاده قرار می‌گیرد. این روش مبتنی بر ایده‌ی ترکیب طبقه‌بندها (با استفاده از شبکه عصبی) و استفاده از تحلیل چندمقیاسی (با استفاده از فیلتربانک) است.

دو روش نیز جهت فشرده‌سازی تصاویر استخراج شده از روش فوق ارائه شد. روش اول مبتنی بر استفاده از تبدیل KLT به منظور فشرده‌سازی تصاویر متنی ارائه گردید. دومین روش پیش پردازش تصاویر متنی قبل از اعمال فرآیند فشرده‌سازی JPEG می‌باشد.

در روش پیشنهادی اول جهت فشرده سازی با استفاده از تبدیل KLT علاوه بر بیشتر بودن میزان فشرده‌سازی در روش پیشنهادی نسبت به دو روش JPEG و JPEG2000 کیفیت تصویر متنی بازسازی شده نیز تا حد قابل توجهی بیشتر است. همچنین ملاحظه می‌شود که روش پیشنهادی تمایل به کاهش کانتراست تصویر متنی دارد.

کلمات کلیدی : فشرده سازی تصاویر متنی-مدل محتوای ترکیبی- جداسازی متن از زمینه - دودویی کردن.

لیست مقالات پذیرفته شده/چاپ شده

[۱] هادی گرایلو، آیدین خداشناس، و مصطفی صفایی، "پیش پردازش تصاویر متنی به منظور استفاده موثر از استاندارد JPEG در کاربرد فشرده‌سازی تصاویر متنی"، ارائه شده در چهارمین کنفرانس فناوری اطلاعات و دانش، ۲ لغایت ۴ خرداد ۱۳۹۱.

[۲] هادی گرایلو، آیدین خداشناس، و مصطفی صفایی، "دودویی سازی سریع تصاویر متنی دستنویس فارسی سطح خاکستری مبتنی بر تحلیل چندمقیاسی و ترکیب طبقه‌بندها" ارائه شده در هفتمین کنفرانس ماشین بینایی و پردازش تصویر ایران، ۲۵ و ۲۶ آبان ۱۳۹۰.

[۳] هادی گرایلو، آیدین خداشناس، و مصطفی صفایی، "فشرده‌سازی تصاویر متنی سطح خاکستری با استفاده از تبدیل KLT" پذیرفته شده در اولین کنفرانس بازشناسی الگو و تحلیل تصویر ایران، ۱۶ تا ۱۸ اسفند ۱۳۹۱.

[۴] هادی گرایلو، امین قنبرزاده، آیدین خداشناس، و مصطفی صفایی، "استفاده از تکنیک انطباق الگوی بهبود یافته و مدل سازی چندجمله‌ای برای فشرده سازی با/بدون ائتلاف سیگنالهای گفتار" ارائه شده در بیستمین کنفرانس مهندسی برق ایران، ۲۶ لغایت ۲۸ اردیبهشت ۱۳۹۱

فهرست عناوین:

فصل اول: مقدمه

۱-۱

۱.....پیش‌گفتار

۲-۱.....تاریخچه کلی

۳-۱.....مشکلات موجود

۴-۱.....ساختار پایان‌نامه

فصل دوم: مروری بر روش‌های موجود

۱-۲.....مقدمه

۲-۲.....روش‌های فشرده‌سازی استاندارد بین‌المللی ویژه تصاویر ثابت و متنی

۱-۲-۲.....کدگذاری طول تداوم

۲-۲-۲.....فشرده‌سازی CCITT یک‌بعدی

۳-۲-۲.....فشرده‌سازی CCITT دو‌بعدی

۴-۲-۲.....کدگذاری مبتنی بر نماد

۵-۲-۲.....فشرده‌سازی JBIG2

۶-۲-۲.....برش‌بندی صفحه‌بیتی

۲۸.....کد گذاری صفحه بیتی.....۷-۲-۲

۳۳.....مروری بر روشهای پایه فشرده سازی تصاویر متنی دودویی.....۳-۲

۳۹.....۴-۲ مروری بر روشهای جداسازی و فشرده سازی تصاویر اسناد مبتنی بر مدل محتوای ترکیبی.....

۴۶.....۵-۲ جمع بندی و نتیجه گیری.....

فصل سوم: مبانی نظری

۴۸.....مقدمه.....

۴۸.....۱-۳ تبدیل KLT.....

۴۸.....۱-۱-۳ مبانی تبدیل KLT.....

۵۰.....۲-۱-۳ تبدیل KLT به طور کامل سیگنال را ناهمبسته می کند.....

۵۱.....۳-۱-۳ تبدیل KLT به طور بهینه انرژی را متراکم می کند.....

۵۳.....۴-۱-۳ کاربرد KLT در فشرده سازی.....

۵۴.....۲-۳ آنالیز موجک.....

۵۵.....۱-۲-۳ تبدیل موجک (WAVELET TRANSFORM).....

۵۸.....۳-۳ JPEG.....

۶۱.....۴-۳ شبکه عصبی.....

۶۱.....۱-۴-۳ پرسپترون.....

- ۶۳.....۲-۴-۳ شبکه ی پس انتشار استاندارد.....
- ۶۳.....۱-۲-۴-۳ ساختار.....
- ۶۴.....۲-۲-۴-۳ الگوریتم.....
- ۶۷.....۳-۲-۴-۳ تابع فعال سازی.....
- ۶۹.....۴-۲-۴-۳ الگوریتم آموزش.....
- ۷۲.....۵-۳ آستانه گذاری.....
- ۷۳.....۱-۵-۳ مبانی آستانه گذاری شدت روشنایی.....
- ۷۵.....۲-۵-۳ نقش نویزدر، آستانهگذاری تصویر.....
- ۷۶.....۳-۵-۳ نقش شدت روشنایی وانعکاس.....
- ۷۷.....۴-۵-۳ آستانه گذاری سراسری ساده.....
- ۷۸.....۵-۵-۳ آستانه چند گانه.....
- ۷۸.....۶-۵-۳ آستانه گذاری متغیر.....
- ۷۸.....۱-۶-۵-۳ آستانه گذاری متغیر بر مبنای خصوصیات محلی تصویر.....
- ۸۰.....۲-۶-۵-۳ استفاده از جابجایی میانگینها.....

فصل چهارم: روش های پیشنهادی جهت ایجاد تصویر دودویی و فشرده سازی آنها

- ۸۱.....۱-۴ مقدمه.....

- ۲-۴ روش پیشنهادی جهت دودویی سازی تصاویر متنی فارسی..... ۸۳
- ۱-۲-۴ روش دودویی سازی پیشنهادی اول..... ۸۵
- ۲-۲-۴ روش دودویی سازی پیشنهادی دوم..... ۸۷
- ۳-۲-۴ ترکیب دو روش پیشنهادی و تصمیم گیری..... ۸۸
- ۳-۴ روشهای پیشنهادی جهت فشرده سازی تصاویر متنی..... ۸۸
- ۱-۳-۴ روش اول..... ۸۸
- ۲-۳-۴ روش دوم..... ۸۹
- ۱-۲-۳-۴ بررسی اثر تبدیل DCT..... ۹۱
- ۴-۴ روش پیش پردازش پیشنهادی..... ۹۷
- ۱-۴-۴ نحوه تعیین وزن ضرایب تبدیل..... ۹۸
- ۲-۴-۴ بازچینی ضرایب تبدیل..... ۹۹
- ۵-۴ جمع بندی و نتیجه گیری..... ۱۰۱

فصل پنجم: پایگاه داده و نتایج آزمایش ها

- ۱-۵ پایگاه داده..... ۱۰۲
- ۱-۱-۵ ساخت افزار استفاده شده..... ۱۰۲
- ۲-۱-۵ دسته بندی تصاویر..... ۱۰۳
- ۳-۱-۵ درصد فراوانی انواع تصاویر در پایگاه داده..... ۱۰۴
- ۲-۵ نتایج مربوط به دودویی سازی تصاویر متنی..... ۱۰۷

- ۳-۵ نتایج روش پیشنهادی اول مربوط به فشردن سازی تصاویر متنی با استفاده از تبدیل KLT... ۱۱۰
- ۴-۵ نتایج روش دوم جهت فشردن سازی تصاویر متنی توسط پیش پردازش تصاویر جهت استفاده بهینه از استاندارد JPEG..... ۱۲۰
- ۵-۵ جمع بندی و نتیجه گیری..... ۱۲۴

فصل ششم: جمع بندی و نتیجه گیری

- ۱-۶ خلاصه و جمع بندی مطالب گفته شده..... ۱۲۶
- ۲-۶ نتیجه گیری..... ۱۲۸
- ۳-۶ پیشنهادهایی برای ادامه کار..... ۱۲۸
- فهرست مراجع..... ۱۳۰

فهرست اشکال:

شکل (۱-۱): یک نمونه از تجزیه تصویر یک سند به سه لایه پیش‌زمینه، پوشش و پس‌زمینه به کمک مدل محتوای ترکیبی..... ۳

شکل (۲-۱): (الف) قسمتی از یک تصویر اصلی که در آن به دلیل اثر ضد‌همپوشانی، نواحی مرزی دارای گذر نرم می‌باشند؛ (ب) نتیجه حاصل از جداسازی تصویر (الف) که در آن مرزها مضرس شده‌اند..... ۶

شکل (۳-۱): (الف) قسمتی از یک تصویر اصلی؛ (ب) نسخه فشرده شده تصویر (الف) به کمک روش DjVu؛ (ج) و (د) چند کلمه مشخص شده در تصاویر (الف) و (ب) با مقیاس بزرگتر جهت مقایسه بهتر..... ۹

شکل (۱-۲): (الف) یک نمونه تصویر متنی؛ (ب) و (ج) قسمتی از نتایج فشرده‌سازی تصویر (الف) به کمک روشهای به ترتیب JPEG2000 و j2b..... ۱۳

شکل (۲-۲): برخی استانداردهای رایج فشرده‌سازی تصاویر، قالبهای فایل و حاملها. اقلام مورد تحریم بین‌المللی (منسوخ شده) با تیره پر رنگ نشان داده شده‌اند؛ در حالی که بقیه به صورت خاکستری کمرنگ هستند..... ۱۴

شکل (۳-۲) روال کدگذاری READ دوبعدی CCITT. نماد گذاری $a_1 b_1$ نشان دهنده مقدار مطلق فاصله بین عناصر تغییر یابنده a_1 و b_1 است..... ۱۹

شکل (۴-۲): پارامترهای کدگذاری CCITT مربوط به (الف) مد گذر و (ب) مد افقی و عمودی..... ۲۰

شکل (۵-۲): کد دو بعدی CCITT..... ۲۱

شکل (۲-۶): (الف) سند دوسطحی (ب) دیکشنری نماد (ج) سه تایی های استفاده شده برای یافتن مکان نمادها در سند..... ۲۳

شکل (۲-۷): مقایسه فشرده سازی JBIG2 (الف) فشرده سازی بدون اتلاف و بازسازی (ب) بدون اتلاف از نظر ادراکی (ج) تفاضل مقیاس دهی شده بین آن دو..... ۲۵

شکل (۲-۸): نمایش صفحه‌بیتی یک تصویر ۸ بیتی..... ۲۶

شکل (۲-۹): یک تصویر ۸ بیتی خاکستری به اندازه $1192 * 500$. (ب) تا (ت) صفحات بیتی ۱ تا ۸ که صفحه ۱ مربوط به کم اهمیت ترین بیت است. هر صفحه بیتی یک تصویر دودویی است..... ۲۷

شکل (۲-۱۰): تصویر بازسازی شده با استفاده از (الف) صفحات بیت ۷ و ۸ (ب) صفحات بیت ۶ و ۷ و ۸ (ج) صفحات بیت ۵ و ۶ و ۷ و ۸..... ۲۸

شکل (۲-۱۱) (الف) تصویر تک رنگ ۲۵۶ بیتی (ب) تا (ح) چهار صفحه بیتی مرتبه بالا با کد دودویی و کدگری از تصویر الف..... ۳۱

شکل (۲-۱۲): (الف) تا (ح) چهار صفحه پایین مرتبه با کد دودویی (ستون چپ) و کدگری (ستون راست) از تصویر شکل ۲-۱۱ (الف)..... ۳۲

شکل (۳-۱): یک تصویر و تغییرات هیستوگرام محلی آن..... ۵۴

شکل (۳-۲): (الف) تبدیل موجک گسسته با استفاده از توابع پایه هارکه تغییرات هیستوگرام محلی آن نشان داده شده. (ب) - (د) چند تغریب مختلف $64 * 128$ ، $64 * 256$ ، $128 * 256$ که می‌توانند از الف به دست آیند..... ۵۷

شکل (۳-۳) یک تغییر کوچک در یک زمینه ثابت..... ۵۸

- شکل (۳-۴): بلوک دیاگرام تبدیل JPEG..... ۵۹
- شکل (۳-۵): (الف) ماتریس مربوط به شدت روشنایی و (ب) ماتریس مربوط به اطلاعات رنگ..... ۶۰
- شکل (۳-۶): محل قرار گرفتن ضرایب DC در تبدیل که به روش DPCM کد گذاری می‌شوند..... ۶۱
- شکل (۳-۷): نحوه نمایش ضرایب DC پس از کد گذاری توسط DPCM..... ۶۱
- شکل (۳-۸) ساختار یک شبکه ی عصبی پس انتشار با یک لایه ی مخفی ۶۴
- شکل (۳-۹) تابع سیگموید دودویی با دامنه $(0,1)$ ۶۸
- شکل (۳-۱۰) تابع سیگموید دو قطبی، دامنه ی $(-1,1)$ ۶۸
- شکل (۳-۱۱): هیستوگرام‌های شدت روشنایی که قابل جداسازی هستند (الف) توسط یک آستانه جدا کرد. (ب) به وسیله دو آستانه جدا کرد. ۷۴
- شکل (۳-۱۲): الف: تصویر ساختگی بدون نویز، تصویر اصلی بانویز گوسی میانگین صفر و انحراف معیار ۱۰، تصویر اصلی با نویز گوسی میانگین صفر و انحراف معیار ۵۰ ۷۵
- شکل (۳-۱۳): (الف) تصویر نویزی (ب) شیب شدت روشنایی در بازه $[0.2-0.6]$ (ج) حاصل ضرب (الف) و (ب). (د) - (و) هیستوگرام‌های متناظر..... ۷۷
- شکل (۴-۱): دیاگرام بلوکی روش دودویی سازی پیشنهادی..... ۸۴
- شکل (۴-۲) بلوک دیاگرام روش فشرده سازی پیشنهادی اول..... ۸۸
- شکل (۴-۳) محل های قرار گیری ۱۶ طیف تبدیل DCT..... ۹۲
- شکل (۴-۴): تصویر اول: یک تصویر متنی تصویر دوم: یک تصویر طبیعی (تصویر ابر)..... ۹۲

- شکل (۴-۵): اثر حذف نواحی طیفی ۲ تا ۱۶ در تصویر اول اثر حذف نواحی طیفی ۲ تا ۱۶ در تصویر دوم..... ۹۳
- شکل (۴-۶): تصویر اول مربوط به حذف فرکانسهای میانی تصویر دوم مربوط به حذف فرکانسهای میانی..... ۹۳
- شکل (۴-۷): تصویر اول مربوط به حذف فرکانسهای بالا تصویر دوم مربوط به حذف فرکانسهای بالا..... ۹۴
- شکل (۴-۸): تصویر اول مربوط به صفرکردن نواحی ۱۲ و ۱۳ و ۱۴ تصویر دوم مربوط به صفرکردن نواحی ۱۲ و ۱۳ و ۱۴..... ۹۵
- شکل (۴-۹): جابجایی محل فرکانس های تصویر متنی پس از تبدیل DCT..... ۹۶
- شکل (۴-۱۰): میزان تقویت هر یک از بازه های فرکانسی..... ۹۶
- شکل (۴-۱۱) بلوک دیاگرام روش پیشنهادی..... ۹۷
- شکل (۴-۱۲): نمونه ای از تصویر لبه ها..... ۹۸
- شکل (۴-۱۳): (الف) نواحی نواری شکل در ماتریس تبدیل DCT (ب) یک نمونه نمودار انرژی نواحی لبه بر حسب پارامتر L ۹۹
- شکل (۴-۱۴): (الف) تصویر اصلی، (ب) نتیجه اعمال پیش پردازش پیشنهادی روی تصویر (الف)..... ۱۰۰
- شکل (۵-۱) نمونه هایی از تصاویر موجود در پایگاه داده..... ۱۰۷
- شکل (۵-۲): یک نمونه تصویر متنی اصلی در درجه ی تفکیک مکانی ۳۰۰ نقطه بر اینچ..... ۱۰۸
- شکل (۵-۳): نتیجه ی دودویی سازی تصویر متنی شکل (۵-۲) به کمک روش پیشنهادی..... ۱۰۹
- شکل (۵-۴): نتیجه ی دودویی سازی تصویر متنی شکل (۵-۲) به کمک روش اُتسو..... ۱۰۹

شکل (۵-۵): برخی از تصاویر متنی موجود در پایگاه مورد استفاده در این پایان نامه..... ۱۱۱

شکل (۵-۶): (الف تا ه) تصویر اصلی، تصویر فشرده شده با روش پیشنهادی، افزایش کانتراست تصویر ، تصویر فشرده شده با JPEG، تصویر فشرده شده با JPEG2000..... ۱۱۵

شکل (۵-۷): نمونه‌ای از اثر نواری تولید شده در میزانهای فشرده‌سازی بزرگ..... ۱۱۵

شکل (۵-۸): بررسی اثر روشهای فشرده‌سازی پیشنهادی، JPEG، و JPEG2000 در مورد تصاویر عمدتاً شامل قسمتهای ترسیمات خطی..... ۱۱۹

شکل (۵-۹): برخی از تصاویر متنی موجود در پایگاه مورد استفاده در این روش..... ۱۲۱

شکل (۵-۱۰): (الف تا ج) تصویر اصلی، تصویر فشرده شده با اعمال روش پیش‌پردازش پیشنهادی و سپس روش JPEG، و تصویر فشرده شده با اعمال مستقیم JPEG..... ۱۲۳

فهرست جداول:

جدول (۱-۲): برخی استانداردهای رایج فشرده سازی تصاویر..... ۱۵

جدول (۱-۵): مقادیر پارامترهای مورد استفاده در روش دودویی سازی پیشنهادی..... ۱۰۸

فصل اول

مقدمه

۱-۱ پیش گفتار

پردازش دیجیتالی تصاویر

با توجه به اینکه امروزه، بیشتر تولیدات تاریخی، هنری و علمی روی کاغذ منتشر می‌شوند. بسیاری از کتابخانه‌های بزرگ دنیا در حال تبدیل منابع خود به شکل دیجیتالی می‌باشند زیرا ذخیره اطلاعات مربوط به اسناد مختلف به شکل دیجیتالی، تا حد زیادی مقرون به‌صرفه‌تر از نگهداری آنها در قفسه‌های کتابخانه است. بنا به دلایل اقتصادی، برای ذخیره چنین حجم عظیمی از اطلاعات به فشرده‌سازی آنها نیاز داریم. در گذشته چنین نیازی وجود نداشت زیرا حجم اطلاعات چندان زیاد نبود اما امروزه با توسعه سریع علم و پیشرفت تکنولوژی، روزانه به مقدار بسیار زیادی به حجم اطلاعات افزوده می‌شود. بنابراین، فشرده‌سازی اطلاعات تصاویر اسناد (مرکب^۱) موضوع جدیدی است که در دهه اخیر توجه زیادی به آن شده است. تصاویر اسناد عمدتاً شامل متن، گرافیک و ترسیمات خطی^۲ است.

روشهای امروزی موجود برای فشرده‌سازی تصاویر طبیعی^۳ را نمی‌توان برای فشرده‌سازی تصاویر اسناد با نرخ فشرده‌سازی زیاد مورد استفاده قرار داد؛ زیرا چنین روشهایی از ویژگیهای

^۱ Compound Document Image

^۲ Line Drawings

^۳ Natural Raster content

مخصوص به متن برای افزایش نرخ فشرده‌سازی استفاده نمی‌کنند. در دهه اخیر مدل محتوای ترکیبی یا MRC¹ برای رسیدن به نرخ فشرده‌سازی بالا و همزمان حفظ کیفیت آن تا حد ممکن، پیشنهاد شده است. در این مدل، اطلاعات مهم مختلف از تصویر سند جدا شده و در لایه جداگانه‌ای قرار داده می‌شوند. سپس هر لایه به طور جداگانه و مستقل از بقیه لایه‌ها، کدگذاری می‌شود.

در این پایان نامه، روشی مبتنی بر مدل محتوای ترکیبی برای فشرده‌سازی اسناد ارایه می‌شود که در آن سعی می‌شود از تبدیل‌های ریاضی خاص و پیش پردازش های لازم در جهت افزایش نرخ فشرده‌سازی بهره گرفته شود. رسم‌الخط‌های مورد توجه ما در این پایان‌نامه، شامل رسم‌الخط‌های فارسی، عربی و لاتین (انگلیسی) می‌باشند.

۱-۲ تاریخچه کلی

موضوع فشرده‌سازی تصاویر اسناد، تاریخچه کوتاهتری نسبت به فشرده‌سازی تصاویر عمومی دارد. روش‌های مهم موجود برای فشرده‌سازی تصاویر اسناد بر مبنای مدل محتوای ترکیبی عمل می‌کنند. همان طور که قبلاً نیز اشاره شد در این مدل، تصویر سند ورودی به تصویرها یا لایه‌های مختلفی تجزیه شده که هر لایه شامل اطلاعات مهم و خاصی از تصویر سند اولیه است. هر لایه معمولاً با دیگر لایه‌ها هم‌پوشانی ندارد. در مدل محتوای ترکیبی پایه، تصویر سند معمولاً به سه لایه پیش‌زمینه^۲، پس‌زمینه^۳ و پوشش^۴ تجزیه می‌شود. لایه پوشش، یک تصویر دودویی است که مشخص کننده نواحی

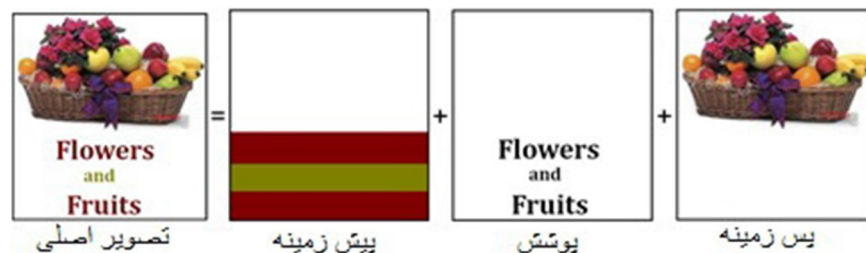
¹ Mixed Raster Content

² Foreground

³ Background

⁴ Mask

متنی است. لایه پیش‌زمینه معرف رنگ نواحی متنی است و لایه پس‌زمینه نیز نشان‌دهنده نواحی گرافیک و ترسیمات خطی است. از لایه پوشش جهت بازسازی تصویر سند فشرده شده به کمک لایه‌های پیش‌زمینه و پس‌زمینه استفاده می‌شود. یک نمونه از تجزیه یک تصویر سند به کمک مدل محتوای ترکیبی در شکل ۱-۱ نشان داده شده است.



شکل (۱-۱): یک نمونه از تجزیه تصویر یک سند به سه لایه پیش‌زمینه، پوشش و پس‌زمینه به کمک مدل محتوای ترکیبی

مدل محتوای ترکیبی توسط سازمان ITU به صورت یک استاندارد، به نام استاندارد T.44 درآمده است. در این استاندارد هر یک از سه لایه پوشش، پیش‌زمینه و پس‌زمینه به کمک یک کدگذار مناسب فشرده و کدگذاری می‌شود. لایه پوشش معمولاً به کمک روشها یا استانداردهای مبتنی بر تکنیک انطباق الگو، مانند JBIG2 و j2b فشرده و کدگذاری می‌شود. لایه پس‌زمینه معمولاً به کمک روشهای فشرده‌سازی تصاویر طبیعی مانند JPEG یا JPEG2000 فشرده و کدگذاری می‌شود.

مهم‌ترین روشهای امروزی جهت فشرده‌سازی تصاویر دودویی، مبتنی بر تکنیک انطباق الگو می‌باشند. در این تکنیک، ابتدا تمام الگوهای تصویر متنی، که به صورت مجموعه‌های به هم پیوسته‌ای از پیکسلهای سیاه تعریف می‌شوند، استخراج می‌شوند. سپس الگوهای مشابه شناسایی، دسته‌بندی و برای هر دسته (یا مجموعه) از آنها یک نماینده به نام الگوی نمونه^۱ انتخاب می‌شود. الگوی نماینده معمولاً شبیه‌ترین الگو به الگوهای مجموعه متناظر خود است. ولی هر الگو لزوماً به طور کامل شبیه به

1 Prototype

نماینده خود نیست بلکه تاحدی متفاوت است. تصویر اختلاف آنها الگوی باقیمانده^۱ یا نقشه خطا^۲ و پیکسلهای آن نیز پیکسلهای باقیمانده یا پیکسلهای خطا نامیده می‌شوند [۱۱ مقاله IJECE]. مجموعه تمام الگوهای نمونه، کتابخانه نامیده می‌شود. هر الگوی نمونه دارای یک شماره در کتابخانه است که شاخص^۳ نامیده می‌شود. کتابخانه در هنگام شروع روش انطباق الگو خالی است. الگوهای تصویر از بالا به پایین به ترتیب مارپیچی استخراج و با هریک از الگوهای نمونه مقایسه می‌شوند. به عمل مقایسه، انطباق کلیشه^۴ نیز گفته می‌شود. اگر الگویی شبیه به یکی از اعضای کتابخانه باشد، الگوی باقیمانده متناظر آن به همراه شاخص آن عضو کتابخانه، و نیز موقعیت مکانی الگوی مذکور نسبت به آخرین الگوی پردازش شده ذخیره می‌شوند؛ اما اگر این الگو شبیه به هیچیک از اعضای کتابخانه نباشد، به عنوان یک الگوی نمونه جدید به کتابخانه اضافه شده و برای آن یک شاخص جدید در نظر گرفته می‌شود. در پایان، دنباله‌های عددی، الگوهای نمونه و تمام یا قسمتی از الگوهای باقیمانده فشرده و کدگذاری می‌شوند. در فشرده‌سازی با اتلاف، معمولاً قسمتی از الگوهای نمونه حذف شده و نیازی به کدگذاری آنها نیست. از آنجائیکه الگوهای باقیمانده، و حتی پیکسلهای باقیمانده، دارای اهمیت یکسانی نسبت به یکدیگر نیستند [۱، ۲]، بهتر است برای کدگذاری آنها اولویت در نظر گرفته شود. با این کار کیفیت تصویر تا حد ممکن حفظ می‌شود.

موارد اختلاف روشهای فشرده‌سازی امروزی مبتنی بر تکنیک انطباق الگو عمدتاً شامل نحوه انجام انطباق کلیشه، نحوه برخورد با الگوهای باقیمانده جهت رسیدن به فشرده‌سازی بهتر (از جنبه‌های مختلف مانند نرخ فشرده‌سازی و داشتن فشرده‌سازی پیش‌رونده^۵) در فشرده‌سازی با اتلاف و بالاخره، نوع کدگذارهای مورد استفاده جهت کدگذاری دنباله‌های عددی مختلف، الگوهای نمونه و

1 Residual Pattern
2 Error Map
3 Index
4 Template Matching
5 Progressive

الگوهای باقیمانده می‌باشند. در مورد روشهای مبتنی بر مدل محتوای ترکیبی، عمده موارد اختلاف روشهای موجود شامل نحوه جداسازی متن از پس‌زمینه آن، نحوه کدگذاری لایه‌های پوشش، پس‌زمینه و پیش‌زمینه و نحوه بازسازی گذر آهسته در نواحی مرزی بین نواحی متنی و غیرمتنی است؛ مورد آخر (یعنی بازسازی گذر آهسته نواحی مرزی) اخیراً مورد توجه قرار گرفته است و تاکنون کار چندانی روی آن انجام نشده است.

روش پیشنهادی در این پایان‌نامه برای فشرده‌سازی تصاویر اسناد مبتنی بر مدل محتوای ترکیبی پایه عمل می‌کند که در آن تصویر ورودی به سه لایه پوشش، پیش‌زمینه و پس‌زمینه تجزیه می‌شود. در اجرای این پایان‌نامه از چندین روش یا تکنیک پیشنهادی استفاده می‌کند که مهمترین آنها شامل روش فشرده‌سازی پیشنهادی برای تصاویر دودویی اسناد و روش جداسازی پیشنهادی می‌باشد. روش پیشنهادی برای فشرده‌سازی تصاویر دودویی اسناد مبتنی بر یک نوع تکنیک پیش پردازش روی تصویر ورودی و استفاده از فشرده ساز JPEG استاندارد و در روش دیگر استفاده از تبدیل ریاضی KLT می‌باشد. روش پیشنهادی نیز برای جداسازی متن از تصاویر اسناد متنی معرفی شده.

روش پیشنهادی در این پایان‌نامه برای فشرده‌سازی همه نوع تصاویر اسناد متنی و دودویی (از نظر نوع رسم‌الخط و محتویات متنی آن) قابل استفاده است اما در تحقیق این پایان‌نامه کارایی آن برای رسم‌الخطهای فارسی و لاتین و با تاکید بر تصاویر متنی چاپی بررسی می‌شود. همچنین فرض می‌شود تصاویر مورد بررسی دارای کیفیت معمولی و عاری از آسیبهای جدی می‌باشند. گرچه می‌توان کارایی آن را برای این نوع تصاویر جداگانه بررسی کرد.

روش پیشنهادی در کاربردهایی که نیاز به نرخ فشرده‌سازی بالا، کیفیت مناسب و زمان کدگذاری کم (نسبت به زمان کدگذاری) می‌باشد، قابل استفاده است. برخی نمونه‌های چنین مواردی شامل بایگانی دیجیتال تصاویر اسناد، مرور تصاویر اسناد کتابخانه‌ها می‌باشد.

۳-۱ مشکلات موجود

یک موضوع مربوط به جداسازی متن از پس‌زمینه می‌شود. جداسازی متن از پس‌زمینه در مدل محتوای ترکیبی تأثیر زیادی روی کیفیت تصویر و نرخ فشرده‌سازی دارد [۸]. برای بررسی بهتر این مشکل توجه خود را روی روش DjVu [۹] معطوف می‌کنیم. این روش یکی از بهترین روشهای موجود امروزی برای فشرده‌سازی تصاویر اسناد است و کارایی آن بهتر از بسیاری از روشهای متداول امروزی است [۱۰-۱۲]. اولین مسأله مربوط به اثر تخریب است. در این اثر، گذر هموار نواحی مرزی (ناشی از اثر ضدهمپوشانی^۱ موجود در فرآیند جاروب کردن^۲) تبدیل به گذر سریع و مرز بین نواحی نیز حالت مضرّس پیدا می‌کند که باعث آفت کیفیت و غیرطبیعی جلوه داده شدن نواحی مرزی می‌شود. این اثر در کدگذاری باکیفیت مبتنی بر مدل محتوای ترکیبی نامطلوب است [۱۰]. شکل ۱-۲-الف قسمتی از یک تصویر اصلی و شکل ۱-۲-ب نتیجه حاصل از جداسازی روش DjVu را نشان می‌دهد.



ب

الف

شکل (۱-۲): (الف) قسمتی از یک تصویر اصلی که در آن به دلیل اثر ضدهمپوشانی، نواحی مرزی دارای گذر نرم می‌باشند؛ (ب) نتیجه حاصل از جداسازی تصویر (الف) که در آن مرزها مضرّس شده‌اند

دومین مسأله این است که روشهای موجود در حین جداسازی متن از پس‌زمینه، توجهی به خوانایی، خوش‌آیندی و میزان قابلیت بازشناسی متن و تصحیح نتایج جداسازی طبق این معیارها نکرده‌اند. منظور ما از خوانایی که یک معیار کمی است، توانایی یک خواننده در خواندن متن

1 Anti-Aliasing Effect
2 Scanning

جداسازی شده است و این مطلب به معنای آن است که تمامی اجزای متنی تأثیرگذار در خواندن متن به درستی از تصویر سند جدا شده باشند. برای ارزیابی این معیار می‌توان از پارامترهای دقت^۱ و یادآوری^۲ استفاده کرد. منظور ما از خوش‌آیندی که یک معیار کیفی است، میزان رضایتمندی خواننده از متن جداسازی شده است که برای ارزیابی و کمی کردن آن، از معیار متوسط امتیاز نظرسنجی^۳ می‌توان استفاده کرد. منظور ما از میزان بازشناسی متن، مقدار نرخ بازشناسی پس از اعمال تصویر متنی جداسازی شده به یک سیستم بازشناسی حروف است. ترکیب استفاده از بازشناسی حروف با مدل محتوای ترکیبی گرچه تاکنون به صورت جدی برای افزایش نرخ فشرده‌سازی مورد استفاده قرار نگرفته است اما به عنوان موضوع امیدبخشی می‌تواند مورد توجه قرار گیرد.

روشهای جداسازی موجود به معیارهای فوق توجهی نشان نداده‌اند بلکه فقط به جداسازی مناسب متن از پس‌زمینه و احتمالاً برقراری مصالحه مناسب در بده بستان بین نرخ فشرده‌سازی و اعوجاج^۴ توجه کرده‌اند. به علاوه، حتی در مسأله جداسازی نیز، روشهای مذکور عمدتاً قادر به جداسازی اجزای متنی کوچک اما مهم نیستند. در رسم‌الخط فارسی/عربی تعداد این اجزای کوچک (مانند نقطه، سرکش و علائمی مانند تنوین) و اهمیت آنها نسبتاً زیاد است و تأثیر زیادی روی کیفیت، خوانایی و معنای متن دارد. یک نمونه تصویر متنی فارسی به همراه نسخه فشرده شده آن به کمک روش DjVu در شکل ۱-۳ نشان داده شده است. در این شکل برخی کلمات به منظور مشخص کردن برخی معایب روش جداسازی DjVu با علائم مستطیلی مشخص شده‌اند. چند کلمه از تصویر اصلی و تصویر فشرده شده در شکل‌های ۱-۳-ج و ۱-۳-د با مقیاس بزرگتری نشان داده شده‌اند تا بهتر با یکدیگر مقایسه شوند. همان طور که از شکل ۱-۳-ب یا ۱-۳-د ملاحظه می‌شود، برخی نقاط و سرکش‌ها به اشتباه

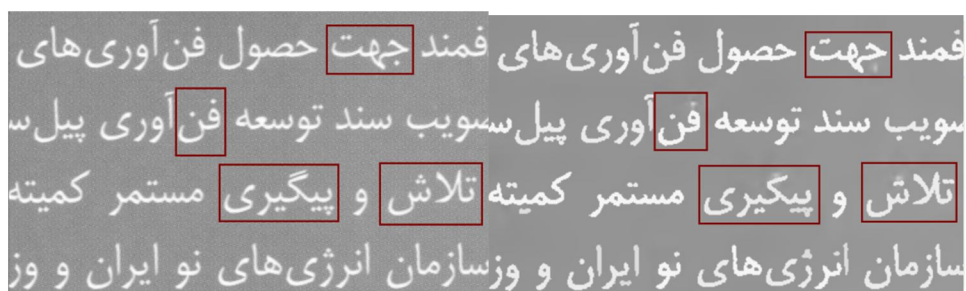
1 Precision

2 Recall

3 Mean Opinion Score (MOS)

4 Rate-Distortion Trade-off

جزو پس‌زمینه دستهبندی شده و بنابراین، به دلیل استفاده از یک کدگذار جداگانه برای لایه پس‌زمینه، مات یا هموار شده‌اند. همان طور که در تصویر کلمه سمت چپ شکل ۱-۳-۵ نشان داده شده است از بین دو نقطه منفرد و با پروفایل شدت روشنایی مشابه، به اشتباه، یکی از آنها جزو اجزای متنی و دیگری جزو اجزای غیر متنی دستهبندی شده‌اند. همچنین سرکش موجود در تصویر کلمه سمت راست شکل ۱-۳-۵ علی‌رغم داشتن شدت روشنایی و مساحت نسبتاً زیاد (همان طور که در شکل ۱-۳-۵ ج نشان داده شده است) به اشتباه، جزو پس‌زمینه تعیین شده است. به علاوه، در این تصویر کلمه، دو گروه نقاط دوتایی و سه‌تایی به اشتباه به یکدیگر چسبیده شده‌اند. شکل ۱-۳-۵ علاوه بر نشان دادن دستهبندی نادرست اجزای متنی و چسبانیدن ناخواسته گروه‌های نقاط، اثر نامطلوب تخریب را نیز نشان می‌دهد. اثر تخریب روی تصاویر اسناد فارسی و عربی تأثیر نامطلوبتری نسبت به اسناد لاتین دارد. چند دلیل برای این امر می‌توان ذکر کرد: اول این که، این اثر می‌تواند فرآیند جداسازی کلمه^۱ را که در اغلب سیستم‌های بازشناسی فارسی و عربی استفاده می‌شود [۱۳] مشکل‌تر کند. دوم این که، این اثر باعث هموار شدن یا اعوجاج جزئیات متنی مانند دندانها و قسمتهای کشیده و ظریف بدنه زیرکلمات می‌شود. این جزئیات در فرآیند بازشناسی اهمیت نسبتاً زیادی دارند. سوم این که، این اثر حتی ممکن است بدنه‌های نازک برخی زیرکلمات فارسی و عربی را شکسته و به چند قطعه تقسیم کند و در نتیجه، باعث کاهش کارایی بازشناسی حروف شود.



ب

الف

فن پیگیری فن پیگیری

د

ج

شکل (۱-۳): (الف) قسمتی از یک تصویر اصلی؛ (ب) نسخه فشرده شده تصویر (الف) به کمک روش DjVu؛ (ج) و (د) چند کلمه مشخص شده در تصاویر (الف) و (ب) با مقیاس بزرگتر جهت مقایسه بهتر

بازسازی لبه سومین مسأله مربوط به روشهای جداسازی و فشرده‌سازی امروزی مبتنی بر مدل محتوای ترکیبی (از جمله روش DjVu) می‌باشد. همان طور که بیان شد، لایه پوشش یک تصویر دودویی است که تمایل به ازبین بردن و اعوجاج جزئیات کوچک متنی مانند خطوط باریک و لبه‌های متنی دارد [۱۰]. در حقیقت یک محدودیت ذاتی لایه پوشش این است که فقط قادر به نمایش گذرهای سریع بین متن، ترسیمات خطی و گرافیک است حال آنکه در عمل، تصاویر اسناد در چنین نواحی‌ای گذرهای آهسته و نرم دارند. این گذرهای نرم اثرات مطلوبی در ضدهمپوشانی دارند و باعث می‌شوند که بتوان تصویر را در درجات تفکیک مکانی پایین‌تر کدگذاری کرده و در نتیجه، به نرخ فشرده‌سازی بالاتری دست یافت.

۴-۱ ساختار پایان نامه

در فصل دوم پایان نامه در زیربخشهای جداگانه‌ای، روشهای موجود برای فشرده‌سازی تصاویر دودویی، روشهای موجود برای جداسازی متن از پس‌زمینه و روشهای موجود برای فشرده‌سازی تصاویر اسناد را مرور می‌کند. در فصل سوم مبانی نظری مورد استفاده در این پایان نامه شرح داده می‌شود، در فصل چهارم روشهایی را برای فشرده‌سازی تصاویر متنی دودویی پیشنهاد می‌دهد. روش پیشنهادی شامل جداسازی متن از پس‌زمینه آن به همراه روشهای پیشنهادی برای فشرده‌سازی تصاویر اسناد می‌باشد.

فصل پنجم پایان نامه نیز نتایج حاصل از پیاده‌سازی بخشهای مختلف و انجام آزمایشها روی تصاویر موجود در پایگاه تصاویر مورد استفاده در پایان نامه را نشان داده و آنها را با نتایج روشهای استاندارد متناظر موجود مقایسه می‌کند. در پایان، فصل ششم به جمع‌بندی، نتیجه‌گیری و بررسی پایگاه داده ایجاد شده در راستای این پایان نامه و همچنین ارائه پیشنهادهایی برای ادامه کار می‌پردازد.

فصل دوم

مروری بر روش های موجود

۱-۲ مقدمه

همان طور که در فصل قبل بیان شد، فشردده سازی اطلاعات تصاویر اسناد جدیدی است که در دهه اخیر توجه زیادی به آن شده است. بنابراین تعداد روشهای مربوط به این زمینه در مقایسه با روشهای مربوط به فشردده سازی تصاویر طبیعی چندان زیاد نیست.

برای فشردده سازی تصاویر طبیعی روشهای متعددی مانند چندی سازی برداری [۱۴]، روشهای مبتنی بر تبدیل [۱۵] و فرکتالها [۱۶] وجود دارند. این روشها تزاید را در سطح پیکسلی کاهش می دهند، اما تزاید موجود در تصاویر متنی عمدتاً در سطح نماد است. بنابراین، اینگونه روشها در مورد تصاویر متنی کارایی فشردده سازی متوسط یا ضعیفی دارند و روشهای امروزی موجود برای فشردده سازی تصاویر طبیعی را نمی توان برای فشردده سازی تصاویر اسناد با نرخ فشردده سازی زیاد مورد استفاده قرار داد.

دلیل دیگری برای عدم کارایی مناسب فشردده سازی برخی از روشهای فشردده سازی تصاویر طبیعی، رفتار پایین گذری آنها است. روشهایی مانند JPEG ضرایب فرکانس بالای تبدیل DCT را صفر یا بسیار کوچک در نظر می گیرند. این عمل باعث هموارشدن تغییرات زیاد محلی می شود و بنابراین مانند یک فیلتر پایین گذر رفتار می کند. اطلاعات تصاویر طبیعی در تمام نواحی اعم از نواحی لبه ها و نواحی داخلی اشیاء توزیع شده است. بنابراین رفتار پایین گذر مذکور تأثیر نامطلوب چندان روی کیفیت

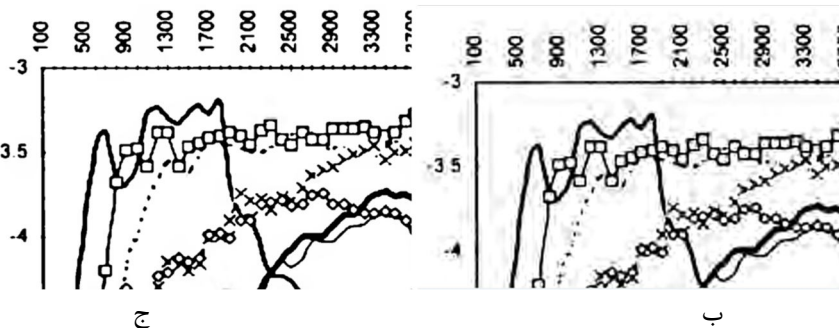
تصویر بازسازی شده ندارد. اما در تصاویر متنی تقریباً تمام اطلاعات در لبه‌ها قرار دارد. بنابراین رفتار پایین‌گذر مذکور تأثیر نامطلوب قابل توجهی روی کیفیت تصویر بازسازی شده دارد. نسخه‌های جدیدتر مانند JPEG2000 نیز اثرات نامطلوبی مانند اثرات نوسانی و مات شدن دارند [۱۷] که آنها نیز کیفیت تصاویر متنی را به طور قابل توجهی کاهش می‌دهند.

تفاوت دیگر تصاویر طبیعی و متنی در نیاز آنها به درجات تفکیک مختلف مکان و مقدار است. در تصاویر متنی، بر خلاف تصاویر طبیعی، به درجه تفکیک بیشتری برای نمایش مکان پیکسلها نسبت به مقدار روشنایی آنها نیاز است [۲۰-۱۸]. مهمترین روشهای فشرده‌سازی امروزی مانند Digipaper [21]، DjVu [۹] و LuraDocument [۲۲] ابتدا تصویر متنی را به تعدادی لایه تجزیه می‌کنند که هرلایه شامل اطلاعات ویژه‌ای مانند متن و گرافیک است [۲۰-۱۸]. سپس هرلایه با روش جداگانه‌ای فشرده‌سازی و کدگذاری می‌شود. معمولاً ابتدا نرخ نمونه‌برداری لایه‌های غیرمتنی به دلیل ویژگی بیان شده، کاهش داده شده و سپس این لایه‌ها فشرده‌سازی و کدگذاری می‌شوند. همچنین لایه پوشش که یک تصویر دودویی و بیانگر محل نواحی متنی است، به کمک استانداردها یا روشهای مبتنی بر انطباق الگو مانند JBIG2 و j2b، فشرده‌سازی و کدگذاری می‌شوند [۲۶-۲۳].

روشهای فشرده‌سازی تصاویر دودویی متنی چاپی امروزی مانند SPM و j2b، بر مبنای تکنیک انطباق الگو عمل کرده و کارایی فشرده‌سازی بسیار بالاتری نسبت به روشهای فشرده‌سازی تصاویر طبیعی مانند JPEG و JPEG2000 دارند. شکل ۲-۱-الف یک نمونه تصویر دودویی متنی فارسی-انگلیسی را نشان می‌دهد. قسمتی از تصویر فشرده شده با روشهای JPEG2000 و j2b به ترتیب در شکل‌های ۲-۱-ب و ۲-۱-ج نشان داده شده است. دو روش مذکور را به صورتی تنظیم کرده‌ایم که تصویر اصلی را به میزان ۳۰ برابر فشرده کنند. همان طور که ملاحظه می‌شود، کیفیت تصویر حاصل از روش مبتنی بر انطباق الگو بسیار بهتر از کیفیت تصویر حاصل از روش دیگر است؛ زیرا در آن اثر نامطلوب نوسانی در اطراف لبه‌ها مشاهده نمی‌شود.



الف



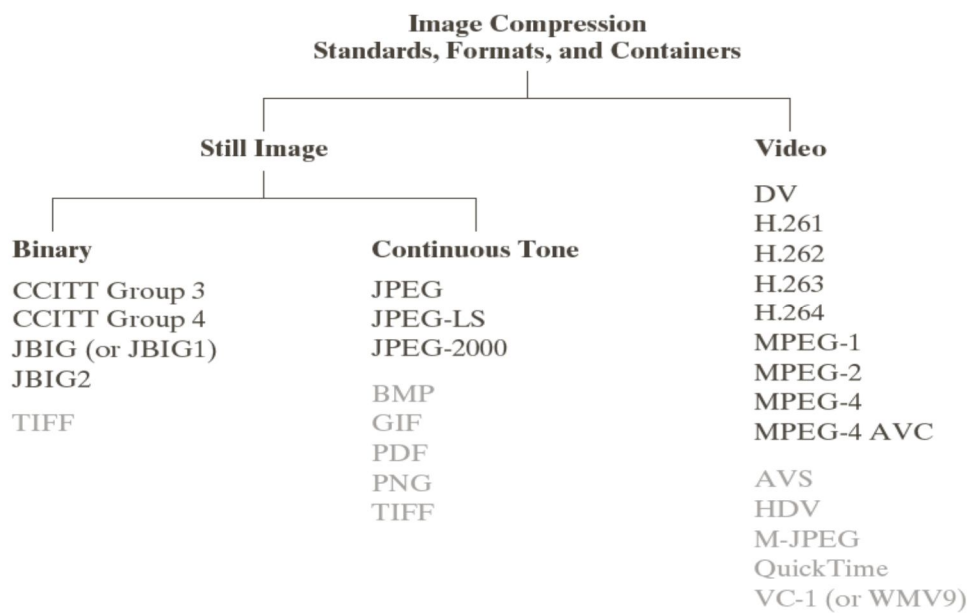
شکل (۱-۲): (الف) یک نمونه تصویر متنی؛ (ب) و (ج) قسمتی از نتایج فشردگی تصویر (الف) به کمک روشهای به ترتیب JPEG2000 و Jb2

زیربخشهایی که در ادامه می‌آیند، روشهای فشردگی تصاویر متنی دودویی، روشهای جداسازی متن از پس‌زمینه آن و روشهای موجود مبتنی بر مدل محتوای ترکیبی برای فشردگی تصاویر اسناد را مرور می‌کنند.

۲-۲ روش‌های فشردگی سازی استاندارد بین‌المللی ویژه تصاویر ثابت و متنی [۹۵]

به طور کلی تصاویر موجود را به دو دسته تصاویر متحرک و ثابت تقسیم می‌کنند، که برای هر یک از این انواع تصویر روش‌هایی جهت فشردگی سازی ارائه شده، تعدادی از این روش‌ها با توجه به کارایی بالاتر مورد اقبال همگانی قرار گرفته و تعدادی از این روش‌ها به صورت استاندارد پروتکل‌های بین

المللی در آمده اند. در نتیجه پس از انتشار روش هایی جهت فشرده سازی تصاویر متنی ، گروهی از این روش ها به صورت مبنایی برای ارسال و دریافت یا ذخیره سازی این نوع از تصاویر قرار گرفتند، پس از مدتی تعدادی از این روش ها بنا به دلایل خاص منسوخ شدند و با روش های جدید تر جایگزین شدند. در زیر نام تعدادی از این روش ها را در جدول ملاحظه می کنید ، که به اختصار به توضیح آن دسته از روش هایی می پردازیم که مربوط به تصاویر ثابت هستند:



شکل (۲-۲): برخی استاندارد های رایج فشرده سازی تصاویر، قالبهای فایل و حاملها. اقلام مورد تحریم بین المللی (منسوخ شده) با تیره پر رنگ نشان داده شده اند؛ در حالی که بقیه به صورت خاکستری کم رنگ هستند. [۹۵]

جدول (۲-۱)- [۹۵]

نام	سازمان	توضیحات شکل ۲,۲
CCITT Group3	ITU-T	بعنوان روشی برای فکس و انتقال اسناد دودویی از طریق خط تلفن طراحی شد . کدگذارهای هافمن و طول تداوم دوبعدی و یک بعدی را پشتیبانی می کند .
CCITT Group4	ITU-T	نسخه ساده تر و روانتر گروه CCITT3 است که فقط از طول تداوم دو بعدی پشتیبانی می کند.
IBGIG یا IBIGI	ISO/IEC ITU-T/	استانداردهای متعلق به گروه مشترک متخصصین تصویر دو سطحی به منظور فشرده سازی پیشرونده و بدون اتلاف تصاویر دو سطحی. تصاویر تن پیوسته با حداکثر ۶ بیت بر پیکسل را می توان براساس صفحه بیت کدگذاری کرد . از کدگذار حسابی حساس به فحوا استفاده می شود . همچنین نسخه اولیه و با رزولوشن پائین به مرور به کمک داده های فشرده سازی بیشتر بهبود داده می شود .
IBGIG2	ISO/IEC ITU-T/	ادامه استاندارد JBIGI برای تصاویر دو سطحی در کاربردهای نامبر، اینترنت و دسک تاپ (رومیزی) ، روش فشرده سازی استفاده شده مبتنی بر محتویات و استفاده از روشهای مبتنی بر دیکشنری در مورد نواحی متنی و نیم متنی تن می باشد . برای دیگر انواع محتویات تصویر ، از کدگذاری هافمن یا کدگذاری حسابی استفاده می شود . آن دارای هر دو قابلیت با اتلاف و بدون اتلاف می باشد .
JPEG	ISO/IEC ITU-T	استاندارد متعلق به "گروه متخصصین عکس" برای تصاویری با کیفیت عکاسی "سیستم کد گذاری پایه" آن با اتلاف بوده و از تبدیل گسسته کسینوسی DCT چندی سازی شده روی بلوک های ۸*۸ تصویر ، کدگذاری هافمن و کد گذاری طول تداوماستفاده می کنداین استاندارد یکی از عمومی ترین روش های فشرده سازی تصاویر در اینترنت می باشد.
JPEG-LS	ISO/IEC ITU-T	یک استاندارد دارای قابلیت بدون اتلاف و نیز تا هدی برای تصاویر تن – پیوسته مبتنی بر پیشبینی وقفی ،مدل کردن فحوا و کد گذاری گلمب
JPEG2000	ISO/IEC ITU-T	ادامه استاندارد JPEG برای افزایش میزان فشرده سازی تصاویر عکسبرداری با کیفیت می باشداز کد گذاری حسابی و تبدیل موجک گسسته (DWT) چندی سازی شده استفاده شده است.نوع فشرده سازی می تواند با/بدون اتلاف باشد.

۲-۲-۱ کدگذاری طول تداوم^۱ [۹۵]

تصاویر دارای شدت روشنایی‌های تکرار شونده در سطر یا ستون را، اغلب می‌توان با نمایش دنباله‌های متناوب از شدت‌های روشنایی یکسان به صورت "زوجهای طول تداوم" فشرده‌سازی نمود که در آن هر زوج طول تداوم مشخص‌کننده شروع یک شدت روشنایی جدید و تعداد پیکسل‌های متوالی شامل این شدت روشنایی می‌باشد. این روش معروف به کدگذاری طول تداوم (RLE) می‌باشد همچنین نوع دو بعدی توسعه یافته آن در استاندارد فشرده‌سازی مربوط به کدگذاری نمابر استفاده شده است. عمل فشرده‌سازی با حذف شکل ساده‌ای از افزونگی مکان (گروههای شدت روشنایی یکسان) انجام می‌شود.

۲-۲-۲ فشرده‌سازی CCITT یک بعدی [۹۵]

در استاندارد فشرده‌سازی گروه ۳ CCITT یک بعدی هر خط یک تصویر به صورت یک سری از کلمات کد هافمن با طول متغیر که نمایانگر طول تداوم‌های سیاه و سفید، به صورت یک در میان، در یک روبش از چپ به راست خط مربوطه می‌باشد. روش فشرده‌سازی به کار گرفته شده به طور رایج به عنوان کدگذاری هافمن اصلاح شده (MH) شناخته می‌شود. کلمات کد خودشان بر دو نوع هستند که استاندارد مربوطه به آنها با عنوان کدهای خاتمه‌دهنده و کدهای آرایشی، اشاره می‌کند.

1 Run Length Coding(RLE)

اگر بطور تداوم r کمتر از ۶۳ باشد ، در این صورت از یک کد خاتمه استفاده می شود . توجه کنید که استاندارد مربوطه کدهای خاتمه مختلفی را برای تداومهای سیاه و سفید مشخص می سازد . اگر $r > 63$ ، در این صورت از دو کد استفاده می شود : یک کد آرایش برای خارج قسمت $[r/63]$ و کد خاتمه برای باقیمانده r تقسیم بر ۶۴ و ممکن است که به شدت روشنایی (سیاه و سفید) دوره تداوم تحت کدگذاری بستگی داشته باشد یا نداشته باشد . اگر $[r/64] < 1792$ ، کدهای آرایش دوره تداوم سیاه و سفید به صورت جداگانه مشخص می شوند در غیر این صورت کدهای آرایش مستقل از شدت روشنایی تداوم هستند .

استاندارد ملزم می‌دارد که هر خط با یک کلمه کد طول تداوم سفید شروع شود، که می تواند در حقیقت کد 00110101 برای یک دوره تداوم سفید به طول صفر باشد . سرانجام، یک کلمه کد پایان خط (EOL) منحصر بفرد 000000000001 برای پایان دادن به هر خط و همچنین علامت دادن اولین خط هر پیام جدید به کار می‌رود . انتهای یک دنباله از تصاویر توسط شش EOL متوالی نشان داده می‌شود .

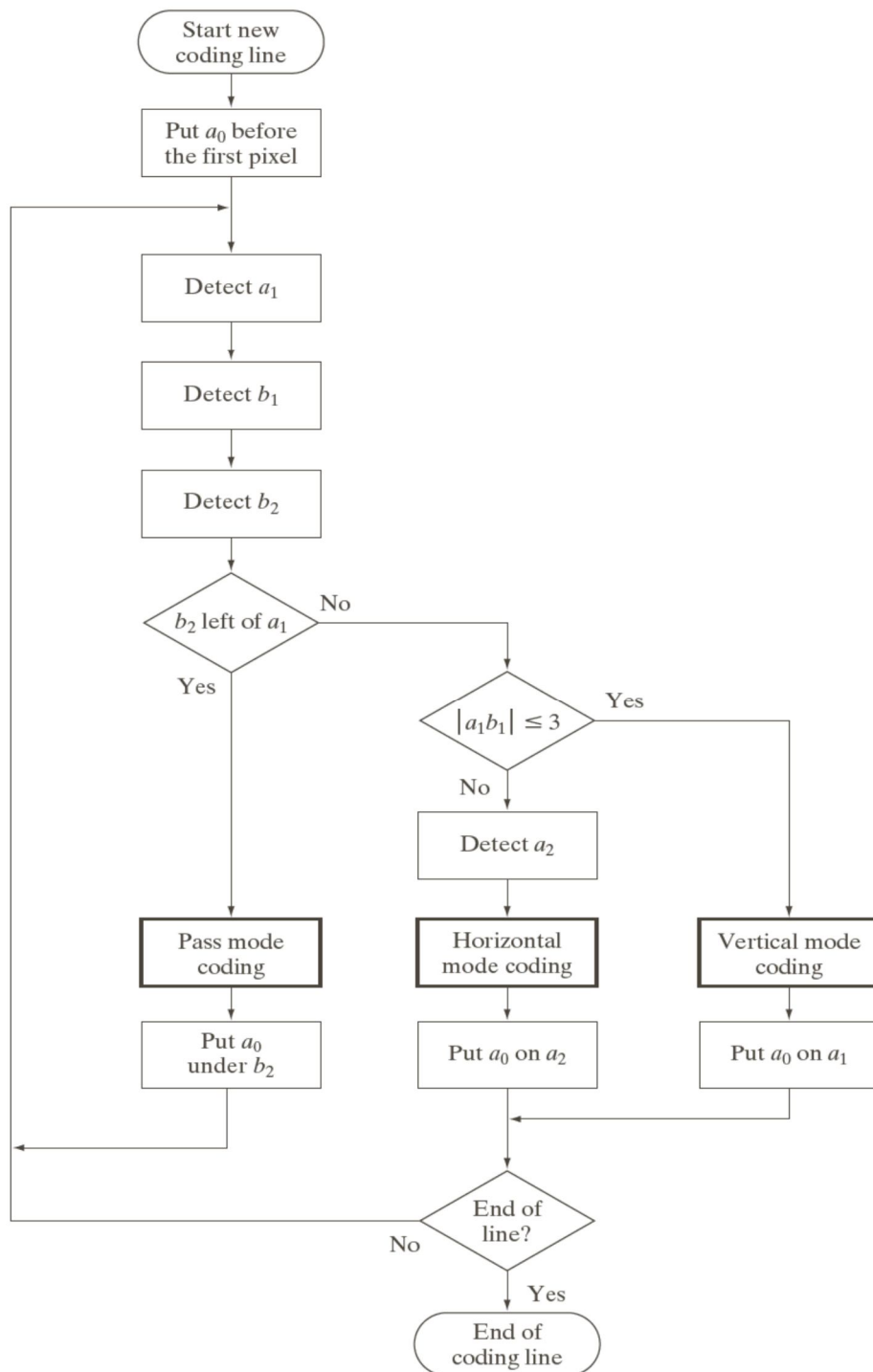
۲-۲-۳ فشرده سازی CCITT دو بعدی [۹۵]

رویکرد فشرده سازی دو بعدی اتخاذ شده برای هر دو استاندارد CCITT گروه ۳ و ۴ یک روش خط به خط است که در آن موقعیت هر انتقال دوره تداوم سیاه به سفید یا سفید به سیاه، با توجه به موقعیت یک عنصر مرجع a_0 که در خط کدگذاری فعلی واقع شده کدگذاری می شود . خط کد شده قبلی خط مرجع نامیده می شود . خط مرجع برای اولین خط هر تصویر جدید، یک خط سفید فرضی می‌باشد . تکنیک کدگذاری دو بعدی که به کار می‌رود ، کدگذاری نشان گذاری آدرس عنصر نسبی

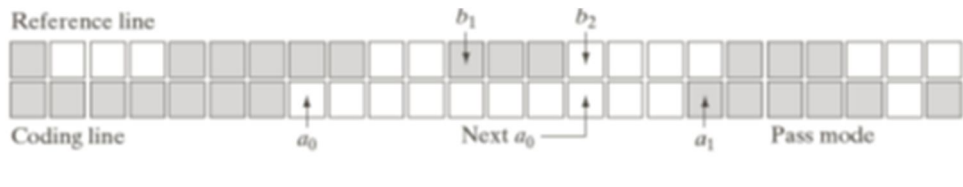
(READ) نامیده می‌شود. در استاندارد گروه ۳، به یک یا سه خط کد شده READ بین خطهای کد شده MH متوالی اجازه داده می‌شود و تکنیک مذکور موسوم به READ اصلاح شده (MR) است. در استاندارد گروه ۴، به تعداد بیشتری از خطهای کد شده READ اجازه داده شده و روش مربوطه، کدگذاری READ اصلاح شده بهبود یافته (MMR) نامیده می‌شود. همچنانکه قبلاً ذکر شد، کدگذاری فوق دو بعدی است؛ از این جهت که از اطلاعات خط قبلی برای کدگذاری خط فعلی استفاده میشود. در اینجا تبدیلات دو بعدی دخیل نیستند.

شکل ۲-۳ فرآیند کدگذاری دو بعدی پایه‌ای برای یک خط روبش تکی را نشان می‌دهد. توجه کنید که گامهای اولیه روال، معطوف به مکان یابی چندین عنصر تغییر کننده کلیدی هستند: b_2, b_1, a_2, a_1, a_0 . یک عنصر تغییر کننده، توسط استاندارد به صورت پیکسلی تعریف می‌شود که مقدارش از مقدار پیکسل قبلی روی همان خط متفاوت است. مهمترین عنصر تغییر کننده a_0 است (عنصر مرجع) که یا به مکان یک عنصر تغییر کننده سفید فرضی در سمت چپ اولین پیکسل هر خط کدگذاری جدیدی یا تعیین شده از مد کدگذاری قبلی تنظیم می‌شود. روشهای کدگذاری در پاراگراف بعدی بحث می‌شود.

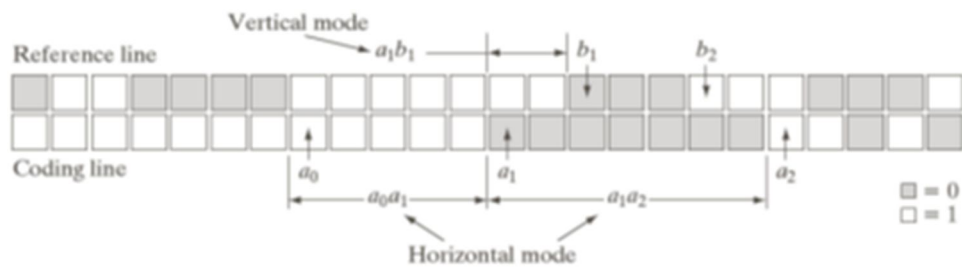
بعد از این که a_0 مکانیابی شد، a_1 بعنوان مکان عنصر تغییر یابنده بعدی به سمت راست a_0 بر روی کدگذاری فعلی a_2 با عنوان عنصر تغییر یابنده بعدی به سمت راست a_1 روی خط کدگذاری، b_1 مشخص می‌شود. اگر هر یک از این عناصر تغییر یابنده تشخیص داده نشود، آنها را به مکان پیکسل فرضی در سمت راست آخرین پیکسل روی خط مناسب تنظیم می‌کنیم. شکل ۲-۴ دو مثال روشن سازنده از روابط کلی بین عناصر تغییر یابنده گوناگون را ارائه می‌دهد.



شکل (۳-۲) روال کدگذاری READ دوبعدی CCITT. نماد گذاری $|a_1b_1|$ نشان دهنده مقدار مطلق فاصله بین عناصر تغییر یابنده a_1 و b_1 است. [۹۵]



الف



ب

شکل (۲-۴): پارامترهای کدگذاری CCITT مربوط به (الف) مد گذر و (ب) مد افقی و عمودی [۹۵]

بعد از شناسایی عنصر مرجع فعلی و عناصر تغییر یابنده مربوط به آن ، دو آزمون ساده جهت انتخاب یکی از سه مد کدگذاری ممکن است انجام می شود. مد گذر، مد عمودی، یا مد افقی. آزمون اولیه که متناظر با اولین نقطه انشعاب در روند نمای شکل ۲-۳ می باشد، مکان b_2 را با مکان a_1 مقایسه می کند. آزمون دوم متناظر با نقطه انشعاب دوم در شکل ۲-۳ می باشد که فاصله (برحسب پیکسل) بین مکانهای a_1 و b_1 آنرا محاسبه و با عدد ۳ مقایسه می کند. بسته به خروجی این آزمونها، به یکی از سه بلوک کدگذاری خلاصه شده در شکل ۲-۳ وارد شده و تکنیکهای کدگذاری مناسب اجرا می شود. سپس یک عنصر مرجع جدید، همانند یک واحد از روند نما در تدارک برای دور تکرار بعدی کدگذاری، برقرار می شود.

شکل ۲-۵ کدهای مشخص به کار رفته برای هر یک از سه مد کدگذاری ممکن را تعریف می کند. در مد گذر، که به طور مشخص موردی را که در آن b_2 مستقیماً در بالای a_1 است. مستثنی می کند،

تنها کلمه کد مد گذر یعنی 0001 لازم است. همچنان که شکل ۲-۴ (لف) نشان می دهد، این مد دوره‌های تداوم خط مرجع سفید یا سیاه را که با دوره‌های تداوم جاری خط کدگذاری سیاه و سفید همپوشانی ندارند، نشان می دهد.

در مد کدگذاری افقی، فاصله‌ها از a_0 به a_1 و از a_1 به a_2 باید در مطابقت با کدهای خاتمه و آرایشی، کدگذاری شده و سپس به کلمه کد افقی 001 اضافه شود. این مطلب در شکل ۲-۵ با نمادگذاری $001+M(a_0a_1)+M(a_1a_2)$ که در آن a_0a_1 به ترتیب معرف فاصله‌ها از a_0 به a_1 و از a_1 به a_2 هستند، نشان داده شده است، سرانجام در مد کدگذاری عمودی، یکی از شش کد باطول متغیر ویژه به فاصله میان b_1, a_1 تخصیص داده می شود. شکل ۲-۴(ب) پارامترهای دخیل در هر دو کدگذاری مد افقی و عمودی را نشان می دهد. کلمه مد تعمیم در پایین شکل ۲-۵ برای وارد کردن یک مد ارسال فشرده سازی نشده به کار می رود.

Mode	Code Word
Pass	0001
Horizontal	$001 + M(a_0a_1) + M(a_1a_2)$
Vertical	
a_1 below b_1	1
a_1 one to the right of b_1	011
a_1 two to the right of b_1	000011
a_1 three to the right of b_1	0000011
a_1 one to the left of b_1	010
a_1 two to the left of b_1	000010
a_1 three to the left of b_1	0000010
Extension	0000001xxx

شکل(۲-۵): کد دو بعدی CCITT [۹۵]

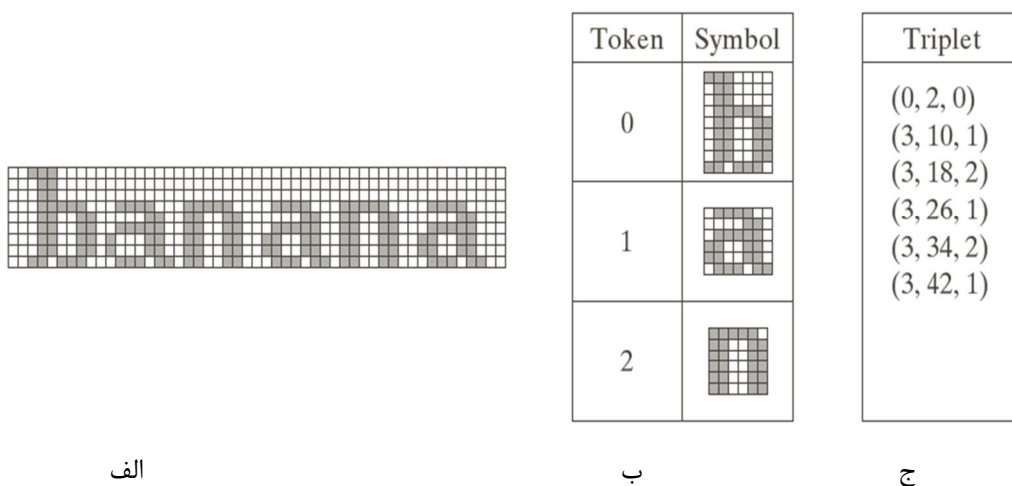
۲-۲-۴ کد گذاری مبتنی بر نماد

در کد گذاری مبتنی بر نماد یا نشانه، یک تصویر به صورت یک مجموعه از زیر تصاویر مکرراً رخ دهنده موسوم به نماد، نمایش داده می شود. هریک چنین نمادی در یک دیکشنری نماد ذخیره شده و تصویر را به صورت مجموعه‌ای از سه تایی ها $\{(x_1, y_1, t_1), (x_2, y_2, t_2), \dots\}$ کد گذاری می کند که در آن هر زوج (x_i, y_i) ، مکان یک نماد را در تصویر مشخص می کند و نشانه t_i آدرس نماد یا زیر تصویر در دیکشنری است.

به این معنی که هر سه تایی نمایانگر یک نمونه از یک نماد دیکشنری در تصویر است. ذخیره سازی نمادهای تکراری، تنها برای یکبار، می تواند تصویر را به طور قابل ملاحظه ای فشرده سازی کند به خصوص در کاربردهای ذخیره سازی و بازسازی اسناد که در آن نمادها غالباً نقشه‌های بیتی کاراکتری هستند که به دفعات زیاد تکرار می شوند.

تصویر دو سطحی ساده در شکل ۲-۶ (الف) را در نظر بگیرید. این تصویر حاوی تک کلمه banana است که متشکل از سه نماد منحصر بفرد می باشد: یک b، سه تا a و دو تا n به فرض اینکه b اولین نماد تشخیص داده شده در فرایند کد گذاری باشد، نقشه بیتی 9×7 در مکان صفر (۰) دیکشنری نماد ذخیره می شود همچنانکه شکل ۲-۶ (ب) نشان می دهد، نشانه مشخص کننده نقشه بیتی b عبارت از ۰ است از اینرو اولین سه تایی در نمایش تصویر کد گذاری شده [شکل ۲-۶ (ج) را ببینید]. به صورت $(0,2,0)$ می باشد. که نشان دهنده این است که گوشه بالای سمت چپ (یک قرارداد اختیاری) نقشه بیتی مستطیلی شکل نمایانگر نماد B قرار است که در مکان $(0,2)$ در تصویر کد گشایی شده قرار می گیرد. بعد از آنکه نقشه های بیتی برای نمادهای a, n مشخص شده و به دیکشنری اضافه شدند، بقیه تصویر را می توان با ۵ سه تایی اضافی کد گذاری کرد. مادامی که تعداد

۶ سه تایی مورد نیاز جهت مکان یابی مکانها در تصویر به همراه سه نقشه بیتی، مورد نیاز برای تعریف آنها، کوچکتر از تصویر اصلی هستند. فشرده سازی روی می دهد. در این حالت، تصویر آغازین تعداد $1 * 51 * 9$ یا ۴۵۹ بیت دارد و با فرض اینکه هر سه تایی متشکل از سه بایت است نمایش فشرده سازی شده دارای $[(6 \times 3 \times 8) + (9 \times 7) + (6 \times 7) + (6 \times 6)]$ یا ۲۸۵ بیت می باشد. نسبت فشرده سازی حاصله $C=1/61$. برای کدگشایی نمایش مبتنی بر نماد در شکل ۲-۶(ج)، شما بایستی صرفاً نقشه های بیتی نمادهای مشخص شده در سه تایی ها از دیکشنری نماد را خوانده و آنها را در مختصات مکانی مشخص شده در هر سه تایی قرار دهید.



شکل (۲-۶): الف) سند دوسطحی (ب) دیکشنری نماد (ج) سه تایی های استفاده شده برای یافتن مکان نمادها در سند [۹۵]

۲-۲-۵ فشرده سازی JBIG2

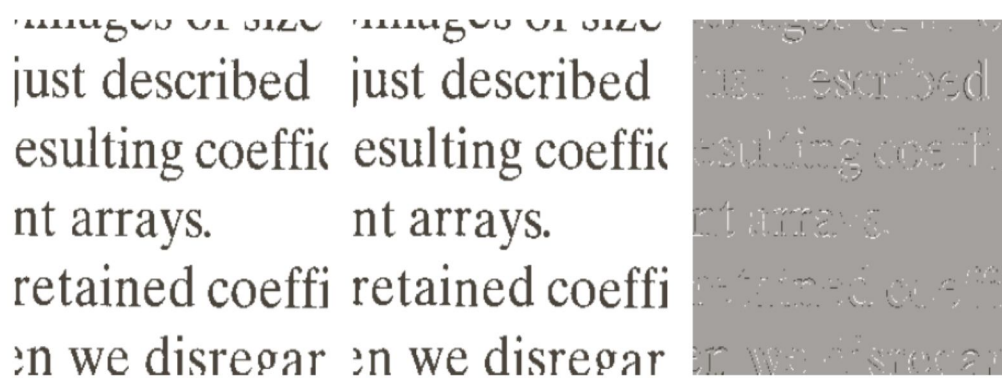
JBIG2 یک استاندارد بین المللی برای فشرده سازی تصاویر دو سطحی است. با تقطیع یک تصویر به نواحی همپوشان و یا غیر همپوشان متن، نیم تن و محتوای عام (ژنریک)، تکنیکهای فشرده سازی که به طور مشخص برای هر نوع محتوا بهینه سازی شده اند، از قرار زیر به کار گرفته می شوند:

۱- نواحی متنی متشکل از کاراکترهایی هستند که به طور ایده آل برای یک روش کدگذاری مبتنی بر نماد مناسبند. به طور مثال، هر نماد متناظر با یک نقشه بیتی کاراکتر است (یک زیر تصویر نمایانگر یک کاراکتر متن) خواهد بود. به طور معمول، تنها یک نقشه بیتی کاراکتر (یا زیرتصویر) در دیکشنری نماد برای هر کاراکتر حرف بزرگ و حرف کوچک که فونت مورد استفاده، وجود دارد. بعنوان مثال، در دیکشنری، یک نقشه بیتی a در دیکشنری، یک نقشه بیتی b یک نقشه بیتی c و به همین ترتیب الی آخر وجود دارد.

۲- در فشرده سازی JBIG2 با اتلاف که غالباً بدون اتلاف ادراکی یا بدون تلف دیداری خوانده می‌شود، ما از تفاوت‌های بین نگاشته‌های بیتی دیکشنری (یعنی نقشه های بیتی کاراکتر مرجع یا کلیشه‌های کاراکتری) و نمونه های مشخص کاراکترهای متناظر در تصویر صرف نظر می‌کنیم در فشرده سازی بدون اتلاف، تفاوتها ذخیره شده و همراه با سه تایی هایی که هر کاراکتر را کدگذاری می کنند. توسط کدگشا مورد استفاده قرار می‌گیرند تا نقشه های بیتی تصویر واقعی را تولید می کنند. تمام نقشه های بیتی به صورت حسابی یا با واسطه از MMR کدگذاری می شوند سه تایی هایی استفاده شده برای دستیابی به اقلام دیکشنری یا به صورت کد حسابی و یا کد هافمن کدگذاری می شوند.

۳- نواحی نیم تن، مشابه نواحی متنی هستند؛ از این جهت که از الگوهای تشکیل شده اند که در یک شبکه منظم شکل گرفته‌اند. نمادهایی که در دیکشنری ذخیره می شوند، نقشه های بیتی کاراکتر نیستند، بلکه الگوهای متناوبی هستند که نمایانگر شدتهای روشنایی هستند (مثلا یک عکس) که دیتراگذاری شده اند (یعنی با نویز دیترا جمع شده اند) تا تصاویر دو سطحی جهت چاپ عکس تولید کنند. نواحی عام (ژنریک) حاوی اطلاعات غیر متنی، غیر نیم تن مثل هنر خط^۱ و نویز بوده و با استفاده از کدگذاری حسابی یا MMR فشرده سازی می شوند.

1 Line art



الف

ب

ج

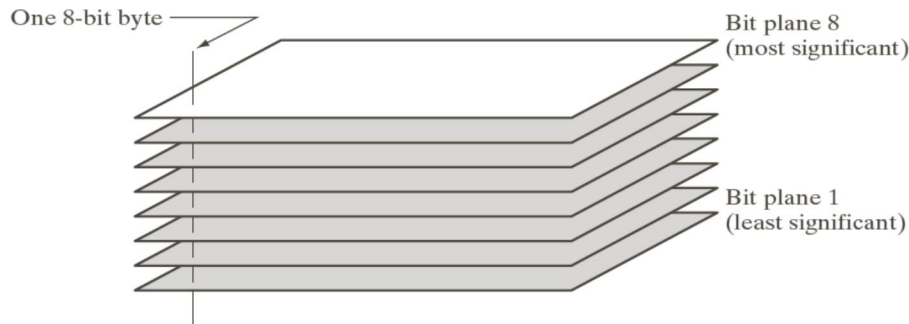
شکل (۲-۷): مقایسه فشرده سازی JBIG2 (الف) فشرده سازی بدون اتلاف و بازسازی (ب) بدون اتلاف از نظر ادراکی (ج) تفاضل مقیاس دهی شده بین آن دو

همچنان که در مورد بسیاری از استانداردهای فشرده سازی تصویر صادق است ، JBIG2 رفتار کدگشا را تعریف می کند؛ البته به طور صریح و روشن کدگشای استاندارد را تعریف نمی کند، بلکه به اندازه کافی انعطاف پذیر است تا طراحی کد گذاری های گوناگون را میسر سازد. اگرچه طراحی کدگذار، به صورت مشخص نشده، باقی گذاشته می شود یا رها می شود، با این وجود واجد اهمیت است؛ چرا که سطح فشرده سازی را که حاصل می شود، تعیین می نماید . بالاخره، کدگذار باید تصویر را به ناحیه هایی تقطیع کند، نمادهای متنی و نیم تن را که در دیکشنریها ذخیره می شوند، انتخاب نموده و تصمیم بگیرد که چه موقع این نمادها اساسا یکسان و یا متفاوت از نمونه های بالقوه نمادها موجود در تصویر می باشد. کدگشا صرفا از آن اطلاعات استفاده می کند تا تصویر اصلی را بازآفرینی کند .

۲-۲-۶ برش بندی صفحه بیتی

پیکسلها، اعدادی دیجیتالی هستند که از بیتها تشکیل شده اند به طور مثال ، هر پیکسل در یک تصویر خاکستری ۲۵۶ سطحی از ۸ بیت تشکیل شده است (یعنی یک بایت) . به جای مشخص کردن

بازه های سطح شدت روشنایی ، می توانیم اثر بیت های خاص را بر ظاهر کلی تصویر مشخص کنیم . همانطور که شکل ۲-۸ نشان می دهد ، یک تصویر ۸ بیتی را می توان متشکل از صفحات ۱ بیتی در نظر گرفت که صفحه ۱ حاوی کم ارزش ترین بین همه پیکسلها و صفحه ۸ حاوی پر ارزش ترین بیت است .



شکل (۲-۸): نمایش صفحه بیتی یک تصویر ۸ بیتی [۹۵]

شکل ۲-۹ (الف) یک تصویر ۸ بیتی خاکستری و شکل های ۲-۹ (ب) تا (ت) ۸ صفحه ۱ بیتی آنرا نشان می دهد . شکل ۲-۹ (ب) متناظر با پائین ترین بیت آن است. دقت کنید که ۴ تصویر مرتبه بالاتر به خصوص در تصویر آخر بخش مهمی از اطلاعات تصویر را در بر می گیرند. بیت های پایین تر مربوط به جزئیات ظریف تر تصویر هستند. تصویر اصلی یک مرز خاکستری با شدت روشنایی ۱۹۴ دارد. دقت کنید که مرزهای برخی تصاویر سیاه (۰) و برخی دیگر سفید (۱) هستند. جهت فهم علت این امر، یک پیکسل مثلا در میان مرز پائینی شکل ۲-۹ (الف) را در نظر بگیرید. پیکسل های متناظر در صفحات بیتی (که با صفحه با بالاترین مرتبه شروع می شود)، دارای مقادیر 11000010 می باشند که نمایش دودویی عدد ۱۹۴ است مقدار هر پیکسل در تصویر اصلی نیز به همین صورت از روی پیکسل های با مقدار دودویی متناظر آن در صفحات بیتی بازسازی می شود. برحسب توابع تبدیل شدت روشنایی، دشوار نیست که نشان دهیم که تصویر دودویی برای هشتمین صفحه بیتی یک تصویر ۸ بیتی را می توان با پردازش تصویر ورودی با یک تابع تبدیل شدن روشنایی آستانه گذار که مقادیر

شدت روشنایی بین ۰ و ۱۲۷ را به ۰ و تمام سطوح روشنایی بین ۱۲۸ و ۲۵۵ را به ۱ نگاشت می‌کند، بدست آورد. تصویر دودویی در شکل ۲-۹(ت)، نیز درست به همین شکل به دست آمده است.



الف

ب

ج



د

ه

و



ز

ح

ت

شکل (۲-۹): یک تصویر ۸ بیتی خاکستری به اندازه 500×1192 . (ب) تا (ت) صفحات بیتی ۱ تا ۸ که صفحه ۱ مربوط به کم اهمیت ترین بیت است. هر صفحه بیتی یک تصویر دودویی است. [۹۵]

تجزیه یک تصویر به صفحات بیتی آن جهت تحلیل نمودن اهمیت نسبی هر بیت در تصویر مفید بوده و می‌توان از آن جهت تعیین نمودن تعداد بیت مورد نیاز برای کوانتیزه سازی تصویر بهره گرفت. این عمل همچنین در فشرده سازی تصویر که در آن تعداد کمتری صفحه در بازسازی شکل به کارگیری می‌شوند، کاربرد دارد. بعنوان مثال، شکل ۲-۱۰ بازسازی تصویر را با صفحات بیتی ۷ و ۸ نشان می‌دهد. بازسازی بوسیله ضرب پیکسل‌های صفحه n م در ثابت 2^{n+1} انجام می‌شود. در حقیقت با این کار n امین بیت پر ارزش را به عدد مبنای ۱۰ تبدیل می‌کنیم. هر صفحه مورد استفاده در ثابت مربوطه ضرب می‌شود و همه صفحات با هم جمع شده تا تصویر خاکستری مورد نظر را بسازند از این رو جهت یافتن شکل ۲-۱۰ (الف) صفحه بیتی ۸ را در ۱۲۸ و صفحه بیتی ۷ را در ۶۴ ضرب می‌کنیم و

دو صفحه را با هم جمع کردیم. علیرغم بازیابی خصوصیات مهم تصویر اصلی، تصویر بازسازی شده به خصوص در پس زمینه هموار به نظر می رسد. این امر تعجب برانگیز نیست، چرا که دو صفحه می توانند تنها چهارسطح متمایز شدت روشنایی را ایجاد کنند. افزودن صفحه بیسی ۶ به بازسازی موقعیت را بهبود می بخشد. به شکل ۲-۱۰ (ب) دقت کنید که در پس زمینه این تصویر خطوط اضافی اشتباه ایجاد شده اند. این اثر با افزوده شدن صفحه پنجم به بازسازی تا حد زیادی کاهش یافته است.

استفاده از صفحات دیگر در بازسازی تاثیر زیادی روی ظاهر شکل نخواهد داشت. بنابراین نتیجه می گیریم که استفاده از صفحه ۴ بیسی با ارزش بالاتر ما را قادر می سازد که تصویر را با جزئیات قابل قبول بازسازی کنیم. ذخیره سازی این ۴ صفحه به جای تصویر اصلی نیاز به ۵۰٪ حافظه کمتر دارد (صرفنظر از مسائل معماری حافظه).



الف

ب

ج

شکل (۲-۱۰): تصویر بازسازی شده با استفاده از (الف) صفحات بیت ۷ و ۸ (ب) صفحات بیت ۶ و ۷ و (ج) صفحات بیت ۵ و ۶ و ۷ و ۸ [۹۵]

۲-۲-۷ کد گذاری صفحه بیسی

تکنیکهای طول تداوم و مبتنی بر نماد بخشهای قبلی را می توان به تصاویری با بیش از دو سطح شدت روشنایی، با پردازش هر یک از صفحات بیسی آنها به تنهایی اعمال نمود. تکنیک مزبور موسوم

به کدگذاری صفحه بیتی مبتنی بر مفهوم تجزیه یک تصویر چند سطحی (تکرنگ یا رنگی) به یک سری از تصاویر دودویی و فشرده‌سازی هر تصویر دودویی از طریق یکی از چندین روش شناخته شده فشرده سازی دودوی، می باشد.

شدت روشناییهای یک تصویر تکرنگ m بیتی را می توان به فرم چند جمله ای مبنای ۲ به صورت زیر نمایش داد :

$$a^{m-1}2^{m-1} + a^{m-2}2^{m-2} + \dots + a^1 2^1 + a_0 2^0 \quad (1)$$

بر مبنای این خاصیت، یک روش ساده تجزیه تصویر به مجموعه ای از تصاویر دودویی به این صورت است که تعداد m ضریب چند جمله‌ای را به تعداد m صفحه بیتی تک بیتی جداسازی کنیم. همچنان که ذکر شد، کم مرتبه‌ترین صفحه بیتی (صفحه متناظر با کم ارزش ترین بیت) با گردآوری a_0 بیت که از هر پیکسل تولید می شود، در حالی که صفحه بیتی با بالاترین مرتبه مشتمل بر بیتها یا ضرایب a^{m-1} می باشد. به طور کلی هر صفحه بیتی با مساوی قرار دادن پیسله‌های آن به مقادیر بیتها یا ضرایب چند جمله ای مناسب از هر پیکسل در تصویر اصلی اولیه ساخته می شود. عدم مزیت ذاتی این رویکرد تجزیه‌ای، این است که تغییرات اندک در شدن روشنایی می تواند تاثیر قابل توجهی بر روی پیچیدگی صفحات بیتی داشته باشد. اگر یک پیکسل با شدت روشنایی ۱۲۷ (01111111) به عنوان نمونه در مجاورت یک پیکسل با شدت روشنایی ۱۲۸ (10000000) باشد، در این صورت هر صفحه بیتی شامل یک گذار متناظر ۰ به ۱ (یا ۱ به ۰) خواهد بود. بعنوان مثال، از آنجایی که بیتهای با ارزش بیشتر کدهای دودویی برای ۱۲۷ و ۱۲۸ متفاوت خواهد بود. لذا بالاترین صفحه بیتی شامل یک پیکسل با مقدار صفر در کنار یک پیکسل با مقدار ۱ خواهد بود که یک گذر ۰ به ۱ (یا ۱ به ۰) در آن نقطه ایجاد می کند.

یک رویکرد بدیل (که اثر تغییرات اندک شدن روشنایی را تقلیل می دهد) این است که ابتدا تصویر را به توسط یک کدگری (Gray) m بیتی نمایش دهیم که گری m بیتی $g_{m-1} \dots g_2 g_1 g_0$ که متناظر با چند جمله ای در زیر می باشد را می توان از روابط زیر محاسبه نمود .

$$g_i = a_i \oplus a_{j-1} \quad 0 \leq i \leq m-2 \quad (2)$$

$$G_{m-1} = a_{m-1}$$

در اینجا \oplus نشانگر عمل یای انحصاری است. این کد دارای این خاصیت منحصر به فرد است که کلمات متوالی تنها در یک موقعیت (مکان) بیتی با هم تفاوت دارند. از اینرو، تغییرات اندک از شدت روشنایی احتمال کمتری دارد که بر تمامی صفحات m بیتی اثر بگذارند. بعنوان نمونه، چنانچه سطوح شدت روشنایی ۱۲۷ و ۱۲۸ در مجاورت هم باشند، فقط صفحه بیتی با بالاترین مرتبه، شامل یک گذار ۰ به ۱ خواهد بود، چرا که کدهای گری که متناظر با ۱۲۷ و ۱۲۸ هستند به ترتیب به صورت 01000000 , 11000000 می باشند.



الف



ب



ج



د



ه



و



ز

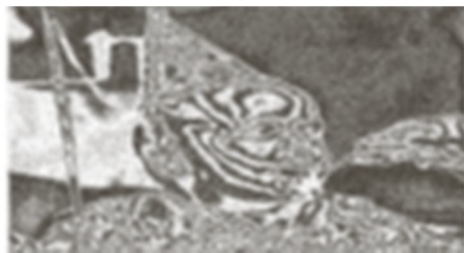


ح

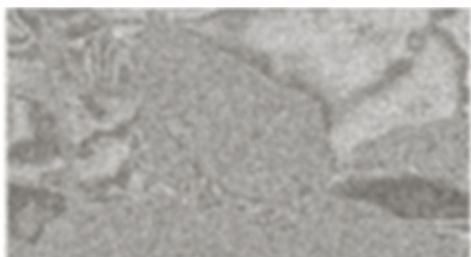
شکل (۱۱-۲) الف) تصویر تک رنگ ۲۵۶بیتی (ب) تا(ح) چهارصفحه بیتی مرتبه بالا با کد دودویی و کدگری از تصویرالف [۹۵]



الف



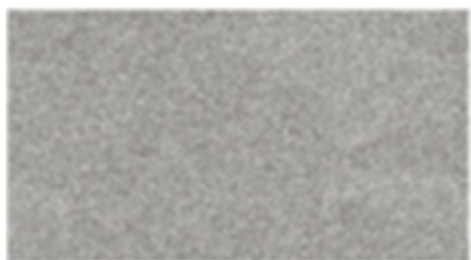
ب



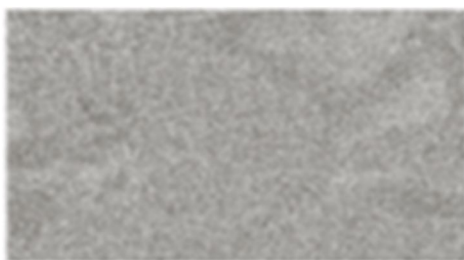
ج



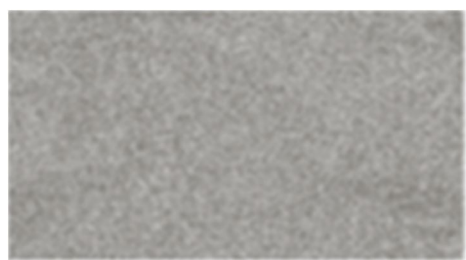
د



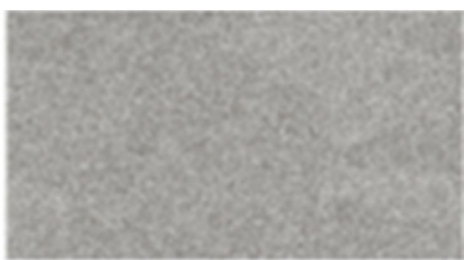
ه



و



ز



ح

شکل (۲-۱۲): (الف) تا (ح) چهار صفحه پایین مرتبه با کد دودویی (ستون چپ) و کد گری (ستون راست) از تصویر شکل ۲-۱۱ (الف) [۹۵]

۲-۳ مروری بر روشهای پایه فشرده‌سازی تصاویر متنی دودویی

انطباق الگو متداولترین روشی است که از آن برای فشرده‌سازی تصاویر دودویی متنی چاپی استفاده شده است. در واقع بیشتر روشها یا استانداردهای موجود از این روش استفاده کرده‌اند. به طور کلی روشهای فشرده‌سازی به دودسته بدون اتلاف و با اتلاف دسته‌بندی می‌شوند. اولین بار ایده استفاده از انطباق الگو در مرجع [۲۶] مطرح شد که الگوهای باقیمانده را کدگذاری نکرد و بنابراین روش پیشنهادی از نوع با اتلاف محسوب می‌شد. همچنین این مرجع، روش خاصی برای فشرده‌سازی و کدگذاری الگوهای نمونه پیشنهاد نداد. در مرجع [۲۸] روشی به نام انطباق ترکیبی الگو یا CSM^۱ برای بهبود روش قبلی پیشنهاد شد که الگوهای نمونه با تکرار بسیار کم را نادیده گرفت و نیز الگوهای باقیمانده را به کمک کدگذاری طول اجرای دوبعدی، مشابه با استانداردهای G3 و G4، فشرده و کدگذاری نمود. روشی که در مرجع [۲۹] پیشنهاد شد از یک کتابخانه از قبل بارگذاری شده استفاده نمود. این کتابخانه شامل حروف الفبا برای چند قلم متداول بود. روش انطباق و جایگذاری الگو یا PMS [۳۰] می‌توانست بخشهای گرافیکی را نیز پردازش کند. در این روش قسمتهای گرافیکی به اجزاء کوچکی شکسته شده و هر کدام به عنوان یک الگو در نظر گرفته می‌شدند. روش «عطف-یای وزن دار» یا WAN^۲ به عنوان بهبودی بر روش CSM پیشنهاد شد که واحد انطباق الگوی آن بهبود داده شده بود. یکی از مؤلفان این روش اعلام نمود که کارایی روشهای CSM، PMS و WAN به هنگامیکه اندازه الگوها کاهش می‌یابد یا این که نویز چندی سازی^۳ افزایش می‌یابد، تنزل می‌کند. بنابراین او روشی به نام راهبرد ترکیبی مستقل از اندازه یا CSIS^۴ پیشنهاد داد که مشابه روش PMS بوده، اما با نرمالیزه کردن اندازه الگوها سعی در رسیدن به عملکردی مستقل از اندازه الگوها داشت.

1 Combined Symbol Matching

2 Weighted AND-NOT

3 Quantization noise

4 Combined Size-Independent Strategy

در مرجع [۳۱] یک روش فشرده‌سازی با اتلاف چندمرحله‌ای پیشنهاد شد که ابتدا برخی پیش‌پردازشها مانند تصحیح خمش^۱ به کمک تحلیل هیستوگرام، روی تصویر انجام و سپس از یک روش انطباق الگو با مقایسه سخت‌گیرانه الگوها استفاده شد. به دلیل این مقایسه سخت‌گیرانه، برخی الگوهای مشابه، به اشتباه، نامشابه در نظر گرفته می‌شدند. در مرحله بعد به کمک یک روش خوشه‌یابی ساختاری چندمرحله‌ای چنین الگوهایی شناسایی و در هم ادغام شدند. این روش از کدگذار Q برای کدگذاری الگوهای نمونه و از کدگذار حسابی وفقی برای کدگذاری دنباله‌های اعداد استفاده کرد. در مرجع [۳۲] یک روش فشرده‌سازی با اتلاف سیگنالهای چندبعدی مبتنی بر الگوهای بازگشتی چندمقیاسی^۲ به نام تجزیه کننده چندمقیاسی چندبعدی یا MMP^۳ پیشنهاد شده است که قابل استفاده برای تمام انواع تصاویر دودویی از جمله تصاویر متنی است. اگرچه این روش مبتنی بر فن انطباق الگو نیست، به این دلیل که از فرهنگ لغت پویا در آن استفاده شده است، در این قسمت آورده شده است. در این روش، یک سیگنال چندبعدی به طور بازگشتی به بردارهایی با طول متغیر قطعه‌بندی شده و سپس هر قطعه به کمک بسط^۴ یا ادغام^۵ بردارهای متعلق به یک فرهنگ لغت کدگذاری می‌شود. این فرهنگ لغت به موازات ورود و کدگذاری داده‌های جدید، به روز می‌شود. در [۳۳] از نوعی فن انطباق الگو مبتنی بر آنتروپی استفاده شده است که در آن از میزان عدم قطعیت یا آنتروپی بین الگوها برای مقایسه و تعیین شباهت الگوها استفاده شده است. سیستم فشرده‌سازی برای تصاویر اسناد یا CDIS^۶ [۳۴] یک روش فشرده‌سازی با اتلاف مبتنی بر انطباق الگو است که در آن دنباله‌های الگوها و مکانها به روش کدگذاری سلسله مراتبی الگوها انجام می‌شود. روش mgtic [۳۶]، [۳۵] مبتنی بر انطباق الگو بوده و از کدگذاری حسابی وفقی مبتنی بر فحوا با فحوای ویژه‌ای برای

1 Skew

2 Multiscale recurrent patterns

3 Multi-dimensional Multiscale Parser (MMP)

4 Expansion

5 Contraction

6 Compression for Document Image System (CDIS)

کدگذاری الگوهای باقیمانده استفاده کرده است. در این روش از تصویر بازسازی شده به کمک الگوهای نمونه به عنوان فحوای مورد نیاز برای کدگذاری الگوهای باقیمانده استفاده شده است.

روش SPM برای کدگذاری الگوهای باقیمانده به طور مشابه با روش قبلی عمل می‌کند با این تفاوت که برای کدگذاری هر الگوی باقیمانده از فحوایی استفاده می‌کند که ترکیبی از الگوی فعلی و الگوی نمونه متناظر است. این تکنیک انطباق الگوی نرم نامیده شده است [۲۵]. روش SPM در حالت با اتلاف از چند نوع پردازش دیگر جهت افزایش نرخ فشردگی استفاده می‌کند که مهمترین آنها معکوس سازی انتخابی پیکسلها است. این تکنیک، رنگ پیکسلهایی را که در کدگذاری حسابی و فقی با دقت مناسبی پیش‌بینی نشده‌اند، معکوس می‌کند. در مجموع روش SPM نتایج نسبتاً بهتری نسبت به روش mgitic دارد [۲۵]. این روش بعنوان بستری برای پیاده‌سازی استاندارد JBIG2 توسط سازمان ISO در نظر گرفته شده است [۳۷]. در [۳۸] یک روش فشردگی با اتلاف برای تصاویر متنی چینی پیشنهاد شده که از یکی از ویژگیهای حروف الفبای چینی برای کاهش اندازه کتابخانه استفاده شده است. برخلاف حروف اسکری که از ۲۵۶ نماد تشکیل شده، مجموعه حروف الفبای چینی از تعداد بسیار زیادی حروف و اجزاء تشکیل شده که باعث پیچیده‌تر شدن آن نسبت بقیه رسم الخطها می‌شود. اجزائی که حروف چینی را تشکیل می‌دهند شکل ساده‌ای دارند بنابراین در این مرجع از روشی مبتنی بر مورفولوژی برای تجزیه و باز ترکیب حروف به منظور انطباق الگو استفاده شده است. همچنین در این مجموعه از حروف الفبا، می‌توان برخی مؤلفه‌های کوچک را درون فضای خالی مؤلفه‌های بزرگتر جای داده و به این ترتیب اندازه کتابخانه را کاهش داد. در مرجع [۳۹] یک روش فشردگی مبتنی بر انطباق الگو در حوزه موجک برای تصاویر متنی چاپی عثمانی پیشنهاد شده است. تمام پردازها در تصویر حاصل از زیرباند LL از تصویر تبدیل یافته انجام شده است. در قسمت انطباق الگوی این روش، هر الگوی استخراج شده با تمام مکانهای تصویر مقایسه شده و اجزاء تطابق یافته از تصویر حذف

می‌شوند. بنابراین این روش فقط قادر به حذف زیرمجموعه‌های کامل است اما قادر به شناسایی و حذف زیرمجموعه‌های جزئی که قسمت مهمی از تصاویر متنی فارسی و عربی را تشکیل می‌دهند نیست.

در روش انطباق الگو برای مقایسه الگوها و تعیین شباهت آنها از تکنیک انطباق کلیشه^۱ استفاده می‌شود. روشهای زیادی برای انطباق کلیشه تصاویر خاکستری وجود دارد. دسته‌ای از این روشها در حوزه تصاویر خاکستری و دسته دیگر در حوزه تصاویر دودویی عمل می‌کنند. در دسته اول از تکنیکهایی مانند تابع همبستگی متقابل نرمالیزه شده [۴۰، ۴۱] و شبکه‌های عصبی [۴۲] استفاده می‌شود. در این دسته برخی روشها قادر به تعیین مقدار درجه چرخش الگوها [۴۳] و برخی دیگر قادر به تعیین بهترین مقیاس برای رسیدن به بهترین انطباق می‌باشند [۴۴]. در دسته دوم، عمل انطباق کلیشه در حوزه تصاویر دودویی انجام می‌شود. معمولاً از تعداد نسبی پیکسلهای مشترک [۴۵] یا به طور معادل، تعداد پیکسلهای نقشه خطا [۳۵] استفاده می‌شود.

تمام روشهای فشرده سازی تصاویر دودویی متنی مبتنی بر انطباق الگو از نقشه خطای وزن دار برای انطباق کلیشه استفاده کرده‌اند [۳۶، ۳۵]. در روش CSM ابتدا نقشه خطا به کمک عملگر «یای انحصاری» بدست آمده و سپس مجموع وزن دار پیکسلهای آن محاسبه و با یک آستانه مقایسه می‌شود. نحوه محاسبه وزنها نیز به این صورت است که هرچه اندازه خوشه‌ای که پیکسلهای باقیمانده تشکیل می‌دهند بزرگتر باشد، وزن بزرگتری برای آنها در نظر گرفته می‌شود [۲۸]. روش WAN بین حالت‌های گذر از سفید به سیاه و سیاه به سفید تفاوت قائل شده و برای هرکدام وزن جداگانه‌ای در نظر می‌گیرد. همچنین این روش فقط الگوهایی را با هم مقایسه می‌کند که اندازه محیط آنها در شرایط مشخصی صدق کنند. در روش PMS هرگاه پیکسلی در نقشه خطا یافت شود که حداقل چهار همسایه آن برابر ۱ باشند اعلام عدم شباهت می‌کند. روشهای mgitic و SPM از معیار فاصله همینگ

1 Template Matching

برای تعیین میزان شباهت استفاده می‌کنند. این فاصله برابر تعداد نقاط اختلاف دو الگو، در صورتیکه مراکز هندسی مستطیل‌های محیطی آنها بر هم منطبق باشد، است [۲۵، ۳۵].

در فشرده‌سازی با اتلاف معمولاً همه یا تعدادی از الگوهای باقیمانده کدگذاری نمی‌شوند. در مراجع [۳۲، ۱] روشی برای اولویت دهی پیکسل‌های باقیمانده برای کدگذاری پیشنهاد شده که از فاصله این پیکسلها از نزدیکترین لبه‌ها برای تعیین اولویت کدگذاری استفاده می‌کند. در مراجع مذکور از اولویت دهی پیکسل‌های باقیمانده به منظور افزایش نرخ فشرده‌سازی و کاهش زمان انتظار لازم برای نمایش اطلاعات بر روی صفحه در فشرده‌سازی پیش‌رونده^۱ استفاده شده است.

استانداردهای اولیه فشرده‌سازی تصاویر دودویی متنی مانند G3^۲ و G4 از روش انطباق الگو استفاده نمی‌کردند؛ بلکه مبتنی بر انواع روشهای کدگذاری طول اجرا و کدهای هافمن بودند [۳۶، ۳۵، ۲۵]. در این استانداردها زمان اجرا بسیار مهمتر از نرخ فشرده‌سازی بود. بعد از آنها استاندارد JBIG1 معرفی شد که از کدگذاری حسابی و فقی مبتنی بر فحوا برای فشرده‌سازی و کدگذاری تصاویر دودویی متنی استفاده می‌شد. در این استاندارد از یک فحوا ۱۰ پیکسلی دوبعدی ویژه برای تعیین احتمال وقوع پیکسلها استفاده می‌شود [۴۶]. این استاندارد نیز از انطباق الگو استفاده نمی‌کند اما کارایی آن بهتر از استانداردهای قبلی است [۳۶، ۲۵]. پس از آن استاندارد JBIG2 در دهه اخیر معرفی شد که از انطباق الگو استفاده می‌کند [۲۶، ۲۴، ۲۳]. گرچه این استاندارد نوع کدگذار را دقیقاً مشخص نکرده است، بیشتر روشهای امروزی مبتنی بر این استاندارد، مانند روش j2b، از کدگذاری مشابه با روش SPM، یعنی انطباق الگوی نرم، استفاده می‌کنند. روش فشرده‌سازی j2b یک روش فشرده‌سازی تصاویر دودویی مبتنی بر استاندارد JBIG2 است که بخشی از روش فشرده‌سازی DjVu می‌باشد.

1 Progressive Compression
2 ITU Group 3

روش DjVu یکی از روشهای متداول امروزی برای فشرده‌سازی تصاویر اسناد، اعم از رنگی و خاکستری، است که محصول شرکت LizardTech می‌باشد.

کدگذاری الگوها بخش مهمی از روشهای مبتنی بر انطباق الگو محسوب می‌شود. روش mgitic از کدگذار دو مرحله‌ای ویژه‌ای [۳۶، ۳۵] برای کدگذاری الگوهای نمونه استفاده می‌کند. این کدگذار دو مرحله‌ای در حقیقت همان کدگذار حسابی و فقی مبتنی بر فحوا است اما این قابلیت را دارد که از فحواهای بزرگتری با زمان اجرای مناسب استفاده کند. هرچه اندازه فحوا بزرگتر شود، کارایی فشرده‌سازی کدگذاری حسابی مبتنی بر فحوا بیشتر، اما زمان اجرا و پیچیدگی محاسباتی آن به صورت تقریباً نمایی افزایش می‌یابد. روش mgitic برای کدگذاری الگوهای باقیمانده از کدگذاری تصویر اولیه به صورت شرطی و با استفاده از تصویر بازسازی شده به کمک الگوهای نمونه استفاده می‌کند. همچنین این روش برای کدگذاری دنباله‌های اعداد از تکنیک PPMC [۴۷] استفاده می‌کند. روش SPM برای کدگذاری الگوهای نمونه از روشی مشابه با استاندارد JBIG1 و برای کدگذاری الگوهای باقیمانده از تکنیک انطباق الگوی نرم استفاده می‌کند. این روش همچنین برای کدگذاری دنباله‌های اعداد از کدگذار QM چندنمادی [۲۵] استفاده کرده است.

بر طبق دانسته ما [۷-۳] هیچیک از روشهایی که تاکنون برای فشرده‌سازی تصاویر دودویی متنی معرفی شدند، از تکنیک کاهش تعداد اعضای کتابخانه استفاده نکرده‌اند. همچنین تا به حال از ترکیب روشهای توصیف مرز با روشهای کدگذاری طول اجرا و حسابی و فقی، برای کدگذاری انواع الگوها استفاده نشده است. به علاوه، ایده تبدیل سیگنال دوبعدی تصویر به سیگنال یک بعدی، که تأثیر زیادی روی افزایش سرعت محاسبات و نرخ فشرده‌سازی دارد، تاکنون در روشهای موجود استفاده نشده است. روشهای توصیف مرز، مانند توصیف کدزنجیره‌ای، سیگنال دوبعدی تصویر دودویی را به سیگنال یک بعدی معادل خود تبدیل کرده و در حقیقت نوعی روش فشرده‌سازی محسوب می‌شوند. روش توصیف کدزنجیره‌ای اولین بار در [۴۸] معرفی گردید. روش توصیف کدزنجیره‌ای

تفاضلی نیز که نوع دیگری از کدزنجیره‌ای است، اولین بار در مرجع [۴۹] معرفی و استفاده شد. این روشها چندین مزیت دارند. اول این که، یک نوع روش فشرده‌سازی محسوب می‌شوند [۵۰]. دوم این که، یک روش مستقل از انتقال برای توصیف تصاویر یا اشیاء دودویی محسوب می‌شوند. این ویژگی باعث سهولت مقایسه اشیاء می‌شود. سوم این که، یک توصیف کامل از اشیاء می‌باشند؛ بنابراین می‌توان هر ویژگی وابسته به شکل را به کمک آنها استخراج کرد [۵۰]. چهارم این که، محاسبه آنها نیاز به زمان اندکی دارد. یکی از معایب مهم روش کدزنجیره‌ای وابسته بودن آن به چرخش و تغییر مقیاس است؛ البته می‌توان از آن برای محاسبه ویژگیهای مستقل از چرخش مانند مساحت، استفاده کرد. کدزنجیره‌ای تفاضلی بر خلاف کدزنجیره‌ای، تا حد زیادی مستقل از چرخش است اما مشابه کدزنجیره‌ای، وابسته به تغییر مقیاس است.

مهمترین کاربردهای موجود کدزنجیره‌ای (تفاضلی) عبارتند از: نمایش شکل [۵۰، ۵۱، ۵۲]، آشکارسازی گوشه [۵۲، ۵۳، ۵۴، ۵۵، ۵۶]، محاسبه محیط و مساحت یک شیء و یا مستطیل محیطی آن [۵۷، ۵۲، ۵۰]، انطباق الگوی دوبعدی [۵۸]، تعیین میزان شباهت با علائم حرفی-عددی [۵۹]، استخراج ویژگیها یا توصیف کننده‌های شکل [۵۲، ۵۱]، تحلیل هارمونیک [۵۲]، هموارسازی مرز [۶۰، ۶۱]، محاسبه خمیدگی و راستی گسسته [۶۲، ۶۱] و تصحیح کجی در کلمات دستنویس [۶۳]. طبق دانسته ما [۳۶، ۶۴، ۶۵] تاکنون از کدزنجیره‌ای (تفاضلی) در کاربرد فشرده‌سازی تصاویر دودویی استفاده نشده است.

۲-۴ مروری بر روشهای جداسازی و فشرده‌سازی تصاویر اسناد

مبتنی بر مدل محتوای ترکیبی

در فشرده‌سازی به کمک مدل محتوای ترکیبی تصویر سند ورودی معمولاً به سه لایه پیش‌زمینه، پس‌زمینه و پوشش تجزیه شده و سپس هرکدام از این لایه‌ها جداگانه کدگذاری می‌شوند. لایه

پیش‌زمینه معمولاً زیرنمونه‌برداری شده و سپس به کمک یکی از روشهای متداول امروزی برای فشرده‌سازی تصاویر طبیعی مانند JPEG یا JPEG2000 فشرده می‌شود. لایه پس‌زمینه معمولاً مشابه با لایه پیش‌زمینه کدگذاری می‌شود. لایه پوشش یک تصویر متنی دودویی است که معمولاً توسط روشهای فشرده‌سازی مبتنی بر تکنیک انطباق الگو مانند روش Jb2 فشرده می‌شود. برای تشکیل لایه پوشش از روی تصویر سند از یکی از روشهای مناسب جداسازی متن از پس‌زمینه استفاده می‌شود. جداسازی متن تأثیر مهم و تعیین‌کننده‌ای روی کارایی فشرده‌سازی کل روش فشرده‌سازی دارد. از طرفی روشهای فشرده‌سازی امروزی مبتنی بر مدل محتوای ترکیبی عمدتاً در نوع روش جداسازی استفاده شده اختلاف دارند. به همین دلیل در این بخش تمرکز اصلی ما روی روشهای جداسازی متن از پس‌زمینه است که تاکنون در مدل محتوای ترکیبی استفاده شده‌اند.

ما می‌توانیم جداسازی متن را متشکل از فرآیندهای آستانه‌یابی (یا دودویی کردن) و اصلاح نتایج بدانیم. اغلب روشهای جداسازی موجود به طور مستقیم یا غیرمستقیم از این قالب استفاده کرده‌اند. روشهای بسیار زیادی برای دودویی کردن تصویر وجود دارد اما به نظر می‌رسد که فقط گروههای خاصی از آنها مناسب تصاویر اسناد می‌باشند [۶۶، ۶۷، ۶۸]. در [۶۷، ۶۸] روشهای دودویی کردن موجود امروزی ارزیابی شده و نهایتاً نتیجه‌گیری شده است که از بین هفت دسته مورد ارزیابی، روشهایی که متعلق به دسته‌های وفقی محلی و مبتنی بر شکل می‌باشند، برای تصاویر اسناد مناسب‌تر هستند. الگوریتم k-means به کرات به عنوان یک روش جداسازی (یا ناحیه‌بندی) در روشهای فشرده‌سازی مبتنی بر لایه مانند DjVu استفاده شده است. روشهای دودویی کردن وفقی و خوشه‌یابی به صورت ترکیبی نیز استفاده شده‌اند [۶۹، ۷۰].

در [۷۱، ۷۲] یک روش فشرده‌سازی چندلایه پیشنهاد شده است که ابتدا بلوکهای غیرهمپوشان 8×8 پیکسلها را به کلاسهای مختلفی مانند متن، تصویر و پس‌زمینه کلاس‌بندی می‌کند. سپس هر کلاس با استفاده از یک روش منحصر به فرد و مخصوص به آن کلاس فشرده می‌شود. از دو الگوریتم

ناحیه‌بندی برای فشرده‌سازی چندلایه استفاده شده است: یک روش ناحیه‌بندی مستقیم به نام الگوریتم MAP متوالی آموزش پذیر (TSMAP)¹ و یک الگوریتم ناحیه‌بندی نرخ-اعوجاج بهینه‌سازی شده (RDOS)². الگوریتم TSMAP به عنوان نماینده‌ای از اغلب الگوریتم‌های ناحیه‌بندی تصویرسند انتخاب شده است که در آن، ناحیه‌بندی فقط از روی تصویر سند ورودی انجام می‌شود. روش RDOS به عنوان یک الگوریتم ناحیه‌بندی که قادر به بهینه‌سازی نرخ-اعوجاج برای تصویر کدگذاری شده می‌باشد پیشنهاد شده است. روش RDOS به این صورت کار می‌کند که ابتدا هر یک از روشهای کدگذاری را (از بین چند روش موجود) به هر ناحیه از تصویر اعمال کرده و از آنجا، کلاس هر ناحیه را طوری تعیین می‌کند که کارایی نرخ-اعوجاج را تقریباً بیشینه کند. بهینه‌سازی RDOS مبتنی بر اعوجاج اندازه‌گیری شده و تخمینی از نرخ بیت هر روش کدگذاری است. مؤلفان اعلام کرده‌اند که روش RDOS نتایج ناحیه‌بندی مقاوتری تولید کرده است. همچنین این روش با تنظیم یک وزن، به کاربر اجازه کنترل مصالحه بین نرخ بیت و اعوجاج را می‌دهد. مؤلفان سپس روش پیشنهادی خود را به منظور انطباق با مدل محتوای ترکیبی اصلاح کردند [۷۳]. روش اصلاح شده پیشنهادی ابتدا یک تصویر متنی را به کلاسهای مختلفی ناحیه‌بندی می‌کند. سپس هر کلاس با توجه به ویژگیهای خود، به طرز متفاوتی به مدل محتوای ترکیبی سه‌لایه تبدیل می‌شود. عمل ناحیه‌بندی به کمک بهینه‌سازی نرخ-اعوجاج و به منظور نمایش در مدل محتوای ترکیبی سه‌لایه انجام می‌شود. این کار به صورت یک فرآیند حلقه بسته³ با اعمال هر تبدیل به هر ناحیه تصویر سند و سپس، انتخاب روشی که بهترین مصالحه نرخ-اعوجاج را نتیجه دهد، انجام می‌شود.

طرح پایین به بالا [۷۴] روشی است که مرتباً پیکسل‌های ناحیه‌بندی شده را طوری ادغام می‌کند که واریانس کلی پیش‌زمینه و پس‌زمینه کمینه شود. روش ناحیه‌بندی پیشنهادی در [۷۵]

1 Trainable Sequential MAP
2 Rate-Distortion Optimized Segmentation
3 Closed Loop

یک الگوریتم ناحیه‌بندی نسبتاً ساده برای پیاده‌سازی مدل محتوای ترکیبی سه‌لایه می‌باشد که در آن فرض می‌شود حجم حافظهٔ بافر موجود محدود است و الگوریتم ناحیه‌بندی باید تک-گذر^۱ باشد؛ بنابراین، عمل ناحیه‌بندی بلوک به بلوک انجام می‌شود. هر بلوک تصویر با توجه به ویژگی‌های محلی خود و همچنین فحوای اطراف خود که توسط مدل مارکوف تعیین می‌شود ناحیه‌بندی می‌شود. در [۷۶] یک روش ناحیه‌بندی برای جداسازی متن از پس‌زمینهٔ تصاویر اسناد پیشنهاد شده است که در تصاویر مورد بررسی، متن با پس‌زمینه همپوشانی دارد. از دو فاز (یا مرحله) برای انجام ناحیه‌بندی استفاده شده است. در فاز اول که شامل تبدیل رنگ و تحلیل خوشه‌یابی است تصویر متنی تک‌رنگ به سه صفحه بخش‌بندی می‌شود: صفحه تاریک، صفحه متوسط و صفحه روشن. در فاز دوم یک آستانه وفقی برای شناسایی متن به کمک دودویی کردن وفقی و استخراج بلوک تعیین می‌شود. در [۷۵] یک الگوریتم ناحیه‌بندی برای مدل محتوای ترکیبی با نواحی استخراج شده با رنگ ثابت (MECCA) پیشنهادی شده است که دارای مزایای سهولت تجزیه، قابلیت بهسازی متن و کاهش نویز می‌باشد. مدل سه‌لایه یک پیاده‌سازی ساده از مدل کلی MRC می‌باشد که یکی از عیوب آن این است که به دلیل منابع مورد نیاز در حین تجزیه برای تشکیل دو لایهٔ پیوسته-تون^۲، فایل‌های حاصل از این مدل وقتی به صورت PDF کدگذاری شوند، قابل چاپ روی چاپگرهای PDF و Postscript نمی‌باشند. این مشکل را به این صورت می‌توان حل نمود که عمل تجزیه را طوری انجام دهیم که شامل فقط یک لایه پیوسته-تون باشد. برای نیل به این مقصود مدل MECCA پیشنهاد شده است [۷۷]. مدل MECCA شامل یک لایه پس‌زمینه، N لایه پیش‌زمینه و N لایه پوشش می‌باشد. لایه پس‌زمینه می‌تواند یک لایه پیوسته-تون باشد اما، لایه‌های پیش‌زمینه باید رنگ ثابت داشته باشند.

1 One-Pass
2 Continuous-Tone

در [۶۶] یک روش ناحیه‌بندی مبتنی بر مدل محتوای ترکیبی استاندارد (ITU-T T.44) پیشنهاد شده است که شامل دو الگوریتم ناحیه‌بندی با هزینه بهینه‌شده (COS¹) و کلاس‌بندی مؤلفه‌های به هم متصل (CCC^۲) می‌باشد. الگوریتم COS برای به دست آوردن یک ناحیه‌بندی اولیه مورد استفاده قرار گرفته است. این الگوریتم از دو مؤلفه ناحیه‌بندی بلوکی و ناحیه‌بندی کلی تشکیل شده است. ناحیه‌بندی بلوکی تصویر ورودی را به بلوکهای همپوشان تقسیم کرده و یک ناحیه‌بندی اولیه نتیجه می‌دهد که در آن برای هر بلوک چهار کاندید ممکن در نظر گرفته شده است. ناحیه‌بندی نهایی برای کل تصویر به کمک فرآیند ناحیه‌بندی کلی انجام می‌شود که مبتنی بر کمینه‌سازی یک تابع هزینه کلی است. سپس روش CCC به نتایج ناحیه‌بندی حاصل از الگوریتم COS اعمال می‌شود. این روش شامل ابتدا استخراج هر مؤلفه به هم متصل در ناحیه‌بندی اولیه و سپس محاسبه یک بردار ویژگی برای هر مؤلفه و در نهایت، کلاس بندی هر ناحیه به هم متصل به یکی از دو کلاس متن و غیرمتن می‌باشد.

در [۹] روش فشرده‌سازی DjVu پیشنهاد شده است که مخصوص فشرده‌سازی با قدرت تفکیک مکانی بالا و کیفیت بالا برای تصاویر رنگی جاروب شده و به منظور انتقال آنها روی خطوط شبکه دارای سرعت پایین طراحی شده است. روش DjVu متن و ترسیمات خطی را (که نیاز به قدرت تفکیک مکانی بالایی دارند) از تصویر و پس‌زمینه (که هموارتر بوده و می‌توانند در قدرت تفکیک پایین‌تری کدگذاری شوند) جدا می‌کند. در روش ناحیه‌بندی پیشنهادی، تصویر به بلوکهای مربعی بخش‌بندی می‌شود. از یک الگوریتم خوشه‌یابی مبتنی بر الگوریتم k-means برای پیدا کردن دو رنگ غالب داخل هر بلوک استفاده می‌شود. سپس از یک الگوریتم همسان‌سازی^۳ برای اطمینان از این که بلوکهای همسایه رنگهای مشابه با پس‌زمینه و پیش‌زمینه اختیار کنند استفاده می‌شود. بعد از این

1 Cost-Optimized Segmentation
2 Connected Component Classification
3 Relaxation Algorithm

مرحله اگر رنگ هر پیکسل به نماینده خوشه مربوط به پیش‌زمینه نزدیکتر باشد (نسبت به نماینده خوشه مربوط به پس‌زمینه)، آن پیکسل به دسته پیش‌زمینه نسبت داده می‌شود. از مرحله بعدی برای پاکسازی و فیلتر کردن مؤلفه‌های پیش‌زمینه به کمک برخی معیارها استفاده می‌شود.

از بین روشهای فشرده‌سازی مبتنی بر محتوای مدل ترکیبی، روشهای DjVu [۹]، LuraDocument [۷۰، ۷۸] و DigiPaper [۶۹] به صورت محصولات تجاری درآمده‌اند. روش DjVu ناحیه بندی را همان طور که قبلاً نیز اشاره شد، به کمک یک روش خوشه‌یابی مبتنی بر روش k-means و به منظور تشکیل لایه پوشش انجام می‌دهد. در روش LuraDocument ابتدا از اعمال فرآیند دودویی کردنِ وقتی به تصویر اصلی و سپس تشکیل یک سند کوانتیزه شده استفاده می‌شود. در ادامه، از اعمال یک فرآیند آشکارسازی متن روی تصاویر اصلی و کوانتیزه شده برای تشکیل لایه پوشش نهایی استفاده می‌شود [۷۸، ۷۰]. فرآیند آشکارسازی متن مذکور مبتنی بر کلاس‌بندی نواحی به هم متصل که شامل مقادیر یکسانی در تصاویر کوانتیزه شده باشند، است. مراحل این فرآیند شامل شناسایی نواحی متصل به هم، فیلتر کردن نواحی متصل به هم به صورت مبتنی بر اندازه، اعمال یک فرآیند لبه‌یابی غیرخطی روی تمام پیکسل‌های مرزی نواحی متصل به هم و در نهایت، کلاس‌بندی متن مبتنی بر مقادیر قدرت لبه و واریانس داخلی نواحی متصل به هم می‌باشد. الگوریتم ناحیه‌بندی DigiPaper از نمایش نشان‌گذاری شده^۱ برای کمک به ناحیه‌بندی استفاده می‌کند. این روش از دودویی کردن وقتی برای تبدیل تصویر ورودی به تصویر دودویی استفاده می‌کند. همچنین، این روش از برخی ویژگیهای متنی برای اصلاح نتایج دودویی کردن و تعیین تعلق بخشهای مختلف تصویر سند به دسته‌های متن و پس‌زمینه استفاده می‌کند [۶۹].

1 Tokenized Representation

در [۱۰] یک روش به نام پردازش با قدرت تفکیک بهبود داده شده (RER) برای بهسازی همزمان کدگذار و کدگشا در مدل محتوای ترکیبی به منظور پردازش کم‌اعوجاج^۱ گذرهای نواحی لبه پیشنهاد شده است. روش RER اثر نامطلوب تخریب را تا حدی کاهش می‌دهد اما از طرف دیگر نرخ فشرده‌سازی را کاهش و زمان کدگشایی را افزایش می‌دهد. در [۷۹] از آستانه‌یابی بلوکی برای ناحیه‌بندی یک تصویر در مدل محتوای ترکیبی استفاده شده است. در این روش، سعی در بهینه‌سازی آستانه‌یابی بلوکی از دیدگاه نرخ-اعوجاج شده است. همچنین یک الگوریتم نسبتاً سریع برای تقریب روش بهینه ارائه شده است. در [۸۰] یک سیستم کدگذار/کدگشا پیشنهاد شده است که در آن، طرح تجزیه به لایه‌ها برای هماهنگی با کدگذار JPEG به همراه کدگذار حسابی و به منظور کدگذاری لایه‌های پس‌زمینه و پیش‌زمینه انطباق داده شده است. در این طرح از JPEG به همراه کدگذار حسابی برای کدگذاری لایه‌های پس‌زمینه و پیش‌زمینه و همچنین از JBIG برای کدگذاری لایه پوشش استفاده شده است. مؤلفان همچنین طرح مذکور را برای استاندارد JPEG2000 نیز انطباق داده‌اند [۸۱].

در [۸۲] یک تکنیک تقسیم و ادغام^۲ برای کلاس‌بندی یک ناحیه محلی به یکی از کلاسهای تک‌سطح^۳، دوسطح و چندسطح به منظور اعمال روشهای فشرده‌سازی مختلف به نواحی مناسب و در نتیجه بهبود نرخ فشرده‌سازی پیشنهاد شده است. بنابراین، در این طرح، روشهای فشرده‌سازی مناسب را می‌توان با توجه به کلاس هر ناحیه محلی انتخاب کرد. مؤلفان فرض کرده‌اند که نواحی محلی دوسطحی معمولاً شامل متن و جداول هستند. همچنین، نواحی چندسطحی شامل گرافیک می‌باشند. تکنیک پیشنهادی مبتنی بر اصل تقسیم و ادغام می‌باشد که در آن تصویر ورودی به بلوکهای غیرهمپوشان تقسیم و سپس، تعداد خوشه‌های معتبر در هر بلوک به کمک یک روش خوشه‌یابی

1 Low-Distortion Rendering
2 Divide-and-Conquer Technique
3 Uni-Level

مبتنی بر ترکیب فیلتر سیگما و خوشه‌یابی متوالی تخمین زده می‌شود. فیلتر رنگ سیگما برای تسهیل کار خوشه‌یابی استفاده می‌شود. سپس، بلوک‌هایی که مشخصات شبیه به یکدیگر داشته باشند، به یکدیگر پیوند زده می‌شوند. مؤلفان همچنین یک فرآیند برای تعیین اندازه بلوکها پیشنهاد داده‌اند. این فرآیند مبتنی بر لحاظ کردن دانه‌بندی ناحیه‌بندی مطلوب و اندازه مورد انتظار کاراکترهای متنی نسبت به ابعاد صفحه می‌باشد.

همان طور که در فصل ۱ نیز بیان شد یک محدودیت ذاتی لایه پوشش این است که فقط قادر به نمایش گذرهای سریع بین متن، ترسیمات خطی و گرافیک است حال آنکه در عمل، تصاویر اسناد در چنین نواحی‌ای گذرهای آهسته و نرم دارند. این گذرهای نرم اثرات مطلوبی در ضدهمپوشانی دارند و باعث می‌شوند که بتوان تصویر را در درجات تفکیک مکانی پایین‌تر کدگذاری کرده و در نتیجه، به نرخ فشرده‌سازی بالاتری دست یافت. تاکنون کار عمده‌ای برای حفظ این گذرهای آهسته و نرم در تصویر کدگشایی شده انجام نشده است. در [۸۳] سعی شده است که از مات کردن نواحی مرزی به کمک یک فیلتر گوسی برای حل این مشکل استفاده شود؛ اما این کار باعث افت کیفیت تصویر کدگشایی شده و تغییر مقادیر تصویر در نواحی گذر و نواحی متنی می‌شود. در تحقیق این پایان نامه، برای حل این مشکل، روشی مبتنی بر مدل کردن نواحی گذر (لبه‌ها) با یک تابع چندجمله‌ای مرتبه پایین پیشنهاد می‌شود.

۲-۵ جمع‌بندی و نتیجه‌گیری

در این فصل روشهای موجود برای فشرده‌سازی تصاویر متنی چاپی دودویی، جداسازی متن از پس‌زمینه و در نهایت فشرده‌سازی تصاویر اسناد به طور مختصر مرور و معرفی شدند.

همان طور که می‌توان ملاحظه نمود هیچ یک از روشهای فوق خوانایی، خوشایندی و قابلیت بازشناسی تصویر متنی دودویی حاصل از عمل جداسازی را در نظر نگرفته‌اند.

فصل سوم

مبانی نظری:

مقدمه:

در این فصل به بیان روش هایی می پردازیم که در طول بررسی و اجرای این پایان نامه مورد استفاده قرار گرفته و در جهت تسهیل در ادراک هر چه بیشتر و راحت تر مفاهیم موجود می باشد. در اینجا به ترتیب به بررسی روش ریاضی KLT، مبانی نظری موجک، استاندارد فشرده سازی JPEG، شبکه عصبی و در انتها روش های آستانه گذاری می پردازیم.

۳-۱ تبدیل KLT

۳-۱-۱ مبانی تبدیل KLT

هر پیش آمد یک آزمایش تصادفی را، می توان با مجموعه ای از N ($N > 1$) متغیر تصادفی $\{x_1, \dots, x_N\}$ نشان داد. به شکل برداری، این متغیرها را به صورت زیر می توان نوشت که به آن یک بردار تصادفی گفته می شود:

$$X = [x_1, \dots, x_N] \quad (1-3)$$

بردار میانگین بردار تصادفی X به صورت زیر تعریف می شود:

$$M = E(X) \triangleq [E(x_1), \dots, E(x_N)] \quad (2-3)$$

ماتریس واریانس این بردار تصادفی به صورت زیر تعریف می شود:

$$\Sigma = E[(X - M)(X - M)^T] \quad (3-3)$$

$$= E(XX^T) - MM^T$$

عناصر بر واقع بر قطر اصلی این ماتریس، σ_i^2 ، برابر واریانس متغیر تصادفی x_i بوده و می توان از آن به عنوان انرژی یا میزان اطلاعات موجود در i امین مولفه سیگنال X تعبیر کرد. بنابراین، کل اطلاعات (یا انرژی) موجود در سیگنال X با عبارت زیر قابل محاسبه و نمایش است:

$$\text{tr}(\Sigma) = \sum_{i=1}^N \sigma_i^2 \quad (4-3)$$

اگر تعداد زیادی از پیشامدهای $X_j, j = 1, \dots, K$ در دست باشد، می توان کمیت های M و Σ را به صورت زیر تخمین زد:

$$\hat{M} = \frac{1}{K} \sum_{j=1}^K X_j \quad (5-3)$$

$$\hat{\Sigma} = \frac{1}{K} \sum_{j=1}^K X_j X_j^T - \hat{M} \hat{M}^T \quad (6-3)$$

اگر در ماتریس کواریانس Σ ، بردارهای ویژه ϕ_k ($k=0, \dots, N-1$) متناظر با مقادیر ویژه λ_k باشند، این بردارها متعامد بوده و با کمک آنها می توان یک ماتریس متعامد Φ به ابعاد $N \times N$ به صورت زیر ساخت که معکوس آن با ترانپوز آن برابر می باشد:

$$\Phi \triangleq [\Phi_0, \dots, \Phi_{N-1}] \quad (7-3)$$

حال با در دست داشتن یک بردار سیگنال x می توان تبدیل متعامد KLT را چنین تعریف کرد:

$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \cdot \\ \cdot \\ y_{N-1} \end{bmatrix} = \mathbf{\Phi}^T \mathbf{x} = \begin{bmatrix} \Phi_0^T \\ \Phi_1^T \\ \cdot \\ \cdot \\ \Phi_{N-1}^T \end{bmatrix} \mathbf{x} \quad (۸-۳)$$

می‌توان نشان داد که در بین تمام تبدیلهای متعامد تنها تبدیل KLT است که اولاً سیگنال را به طور کامل ناهمبسته می‌کند. ثانیاً انرژی یا اطلاعات موجود در بردار سیگنال را تا بیشترین حد ممکن متراکم می‌کند. ویژگی اخیر در فشرده‌سازی سیگنال کاربرد مهمی دارد زیرا می‌توان تعدادی از ضرایب تبدیل را حذف کرده و با انجام عکس عمل تبدیل روی بردار کاهش بعد یافته، تقریبی از سیگنال اولیه به دست آورد.

۳-۱-۲ تبدیل KLT به طور کامل سیگنال را ناهمبسته می‌کند

در این قسمت به دنبال نشان دادن این مطلب هستیم که در بین تمام تبدیلهای متعامد، تنها تبدیل KLT است که به طور بهینه:

✓ سیگنال را به طور کامل ناهمبسته می‌کند.

✓ انرژی یا اطلاعات موجود در سیگنال را تا بیشترین حد ممکن متراکم می‌کند.

اولین ویژگی به سادگی از تعریف تبدیل KLT اثبات می‌شود. دومین ویژگی به این دلیل صادق و برقرار است که تبدیل KLT انرژی موجود در N مولفه‌ی سیگنال را طوری بازتوزیع می‌کند که بیشترین مقدار انرژی در تعداد کمی از مولفه‌های خروجی تبدیل $\mathbf{y} = \mathbf{\Phi}^T \mathbf{x}$ قرار بگیرد (این مطلب بعداً نشان داده خواهد شد).

برای اثبات اولین ویژگی، در خروجی تبدیل $y = \Phi^T X$ بردار میانگین m_y و ماتریس کواریانس \sum_y را در نظر بگیرید:

$$m_y = E(y) = E(\Phi^T x) = \Phi^T E(x) = \Phi^T m_x \quad (9-3)$$

$$\begin{aligned} \sum_y &= E(yy^T) - m_y m_y^T = E[(\Phi^T x)(\Phi^T x)^T] - (\Phi^T m_x)(\Phi^T m_x)^T \\ &= E(\Phi^T (xx^T) \Phi) - \Phi^T m_x m_x^T \Phi = \Phi^T [E(xx^T) - m_x m_x^T] \Phi \\ &= \Phi^T \sum_x \Phi = \Lambda = \text{diag}[\lambda_0, \lambda_1, \dots, \lambda_{N-1}] \end{aligned} \quad (10-3)$$

رابطه‌ی فوق را به شکل ماتریسی زیر نیز می‌توان نوشت:

$$\sum_y = \begin{bmatrix} \dots & \dots & \dots \\ \dots & \sigma_{ij} & \dots \\ \dots & \dots & \dots \end{bmatrix} = \Phi^T \sum_x \Phi = \Lambda = \begin{bmatrix} \lambda_0 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_{N-1} \end{bmatrix} = \begin{bmatrix} \sigma_0^2 & 0 & \dots & 0 \\ 0 & \sigma_1^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{N-1}^2 \end{bmatrix}$$

با توجه به روابط فوق، به دو مشاهده می‌توان دست یافت:

✓ ماتریس کواریانس خروجی تبدیل KLT، $y = \Phi^T x$ ، قطری می‌شود یعنی کواریانس بین هر

دو مولفه مختلف y_i و y_j همواره برابر با صفر است: $\sigma_{ij} = 0$ ؛ به عبارت دیگر، سیگنال به طور

کامل ناهمبسته می‌شود.

✓ واریانس y_i همان i امین مقدار ویژه ماتریس کواریانس بردار سیگنال x است؛ یعنی، $\sigma_i^2 = \lambda_i$

۳-۱-۳ تبدیل KLT به طور بهینه انرژی را متراکم می‌کند

در این قسمت نشان داده می‌شود که تبدیل KLT انرژی سیگنال را به گونه‌ای بازتوزیع^۱ می‌کند که تا بیشترین حد ممکن در تعداد کمی از مولفه‌های (یا ضرایب) پس از اعمال تبدیل KLT قرار داده شود. فرض کنید A یک ماتریس متعامد دلخواه باشد که در $A^{-1}=A^T$ صدق کند. این ماتریس را بر حسب بردارهای ستونی آن (a_i ها که $i=0,1,\dots,N$ است) به صورت زیر می‌توانیم توصیف کنیم:

$$A=[a_0,\dots,a_{N-1}] \quad (۱۱-۳)$$

یا:

$$A^T = \begin{bmatrix} a_0^T \\ \vdots \\ a_{N-1}^T \end{bmatrix} \quad (۱۲-۳)$$

حال بر پایه‌ی این ماتریس می‌توانیم روی هر بردار سیگنال دلخواه مانند x تبدیلی متعامد به صورت زیر انجام دهیم:

$$y = \begin{bmatrix} y_0 \\ \vdots \\ y_{N-1} \end{bmatrix} = A^T x = \begin{bmatrix} a_0^T \\ \vdots \\ a_{N-1}^T \end{bmatrix} x \quad (۱۳-۳)$$

که i آمین مولفه‌ی بردار y به صورت $y_i = a_i^T x$ قابل محاسبه است.

تبدیل معکوس به صورت زیر قابل محاسبه است:

$$x = Ay = [a_0, \dots, a_{N-1}] \begin{bmatrix} y_0 \\ \vdots \\ y_{N-1} \end{bmatrix} = \sum_{i=0}^{N-1} y_i a_i \quad (۱۴-۳)$$

^۱ Redistribute

واریانس مولفه‌های سیگنال قبل و بعد از تبدیل KLT به صورت زیر قابل محاسبه است:

$$\sigma_{x_i}^2 = E[(x_i - \mu_{x_i})^2] \underline{\underline{\Delta}} E(e_{x_i}) \quad (15-3)$$

$$\sigma_{y_i}^2 = E[(y_i - \mu_{y_i})^2] \underline{\underline{\Delta}} E(e_{y_i})$$

در این روابط مولفه‌ی $e_{x_i} \underline{\underline{\Delta}} (x_i - \mu_{x_i})^2$ را می‌توان به عنوان انرژی دینامیکی (یا اطلاعات موجود در) مولفه‌ی i ام سیگنال در نظر گرفت. بنابراین، تریس ماتریس کواریانس (یعنی $tr \sum_x$) برابر با کل انرژی یا اطلاعات موجود در سیگنال است و کل انرژی موجود در X به صورت زیر قابل محاسبه می‌باشد:

$$tr \sum_x = \sum_{i=0}^{N-1} \sigma_{x_i}^2 = \sum_{i=0}^{N-1} E[(x_i - \mu_{x_i})^2] = \sum_{i=0}^{N-1} E(e_{x_i}) \quad (16-3)$$

با توجه به خاصیت جابجایی موجود در عملگر تریس (یعنی $tr(AB) = tr(BA)$) داریم:

$$tr \sum_y = tr(\Phi^T \sum_l \Phi) = tr(\Phi^T \Phi \sum_r) = tr \sum_x \quad (17-3)$$

نتیجه‌ی رابطه‌ی فوق این است که انرژی یا اطلاعات کل موجود در سیگنال بعد از اعمال تبدیل KLT نیز محفوظ باقی می‌ماند؛ گرچه توزیع این مقدار انرژی بین N مولفه‌ی سیگنال قبل و بعد از تبدیل ممکن است بسیار تغییر کند.

۳-۱-۴ کاربرد KLT در فشرده سازی

برای فشرده‌سازی این مراحل را طی می‌کنیم: برای بردار سیگنال داده شده به ابعاد $N \times 1$ ابتدا ماتریس کواریانس و بردارها و مقادیر ویژه آن را محاسبه می‌کنیم. سپس عددی مانند $m < N$ انتخاب

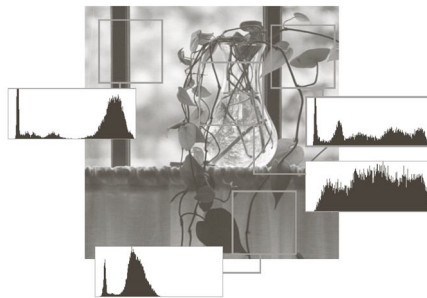
می‌کنیم که تعداد مولفه‌های مورد نیاز در بردار تبدیل بوده و طوری انتخاب می‌شود که انرژی مربوط به این تعداد مولفه، $\sum_{i=0}^{m-1} \lambda_i / \sum_{i=0}^{N-1} \lambda_i$ ، بزرگتر یا مساوی یک مقدار آستانه‌ی از قبل مشخص باشد. حال برای تشکیل ماتریس تبدیل $N \times m$ از m بردار ویژه ماتریس کواریانس متناظر با بزرگترین m مقدار ویژه آن به عنوان ستونهای ماتریس تبدیل Φ_m مورد نظر استفاده می‌کنیم. بردار تبدیل یافته y یک بردار $m \times 1$ است که نسبت به بردار سیگنال اولیه x ابعاد آن کاهش یافته و بنابراین فشرده شده است. برای به دست آوردن تقریبی از سیگنال اولیه x کافی است تبدیل معکوس KLT را به صورت زیر استفاده کنیم:

$$\hat{x} = \Phi_m y \quad (18-3)$$

۲-۳ آنالیز موجک [۹۶]

در سال ۱۹۰۹ هار اولین کسی بود که به موجک‌ها اشاره کرد. در سال‌های ۱۹۳۰ ریاضیدانان به قصد تحلیل ساختارهای تکین موضوعی به فکر اصلاح پایه‌های فوریه افتادند.

تبدیلات موجک مبتنی بر فرکانس متغییر و طول زمانی محدود می‌باشند که این امر به آنها اجازه می‌دهد تا معادل یک درجه بندی فرکانس را برای یک تصویر فراهم سازند. که آشکار می‌سازد چه فرکانس‌هایی و در چه مواقعی باید نواخته شود. همان طور که در شکل ۱-۳ مشاهده می‌کنید هیستوگرام محلی می‌تواند از یک محل به محل دیگر تصویر به طور قابل ملاحظه‌ای تغییر کند.



شکل (۱-۳): یک تصویر و تغییرات هیستوگرام محلی آن

در سال ۱۹۸۷ نشان داده شد که موجک‌ها اساس راهکار جدیدی موسوم به نظریه چند درجه تفکیکی^۱ می باشد. تبدیل هار از قدیمی ترین و ساده ترین موجک‌های متعامد می باشد.

۳-۲-۱ تبدیل موجک (WAVELET TRANSFORM)

تبدیل موجک، سیگنال را به سیگنالهای شیفت یافته و تغییر مقیاس یافته یک موجک اصلی و پایه می شکند. می توان رابطه ای برای تبدیل موجک بصورت زیر تعریف کرد:

$$h_0(Z) = h_{00}(Z) = \frac{1}{\sqrt{N}}, Z \in [0,1] \quad (۱۹-۳)$$

$$h_k = h_{pq}(z) = \frac{1}{\sqrt{N}} \begin{cases} 2^{\frac{p}{2}} & (q-1)/2^p \leq z < (q-0.5)/2^p \\ -2^{\frac{p}{2}} & (q-0.5)/2^p \leq z < q/2^p \\ 0 & z \in [0,1] \text{ و غیر این صورت} \end{cases} \quad (۲۰-۳)$$

می توان رابطه ای برای تکمیل تبدیل موجک به صورت زیر تعریف کرد.

$$C(\text{scale}, \text{position}) = \int_{-\infty}^{\infty} f(t) \psi(\text{scale}, \text{position}, t) dt \quad (۲۱-۳)$$

در حالت گسسته، تبدیل موجک براساس دوفیلتر تعریف می شود، فیلتر پایین گذر برای جدا کردن اطلاعات فرکانس پایین و دیگری فیلتر بالاگذر جهت تفکیک اطلاعات فرکانس بالا. ضرایب مربوط به

1 multiresolution

فیلتر پایین گذر با نام تقریب^۱ شناخته می شوند و نشان دهنده اطلاعات فرکانس پایین اند . ضرایب مربوط به فیلتر بالا گذر با نام جزئیات^۲ شناخته می شوند .

در تجزیه موجک موارد زیر قابل توجه است:

۱- به استثنای زیر تصویر واقع در گوشه بالا سمت چپ (تصویر تقریب) هیستوگرام های محلی بسیار مشابه می باشند و بسیاری از پیکسل ها به صفر نزدیک است. که این تعداد زیاد صفرها به عنوان کاندید های خوبی برای فشرده سازی تصاویر تبدیل می کند.

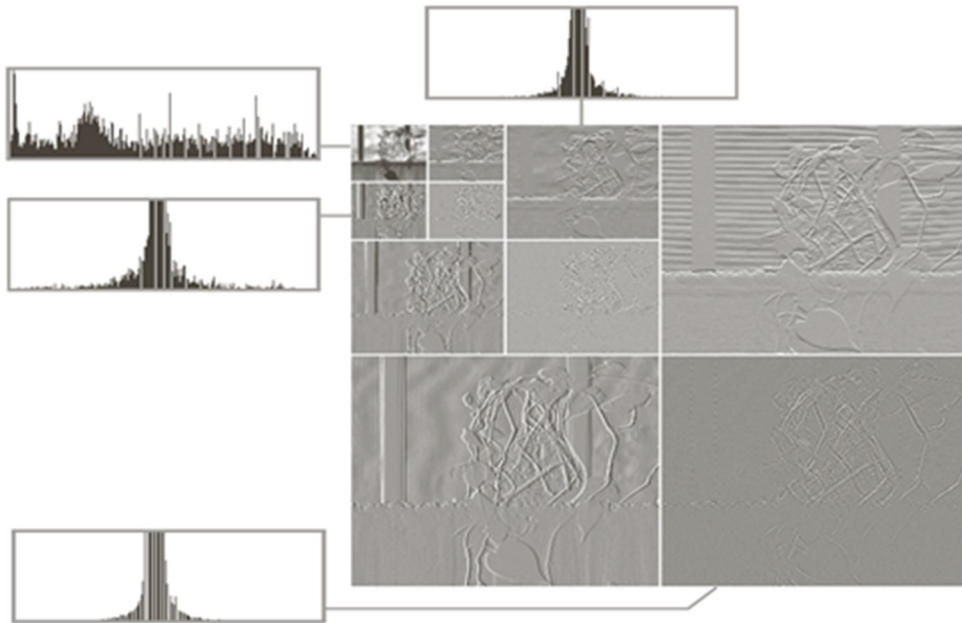
۲- زیر تصاویر ناشی از تجزیه تصویر اصلی را می توان برای بازسازی تصویر اصلی بدست آورد. شکل های ۲-۳ (ب) تا (د) که دارای اندازه های $64*64$ و $128*128$ و $256*256$ هستند از تصاویر اصلی ۲-۳ (الف) شکل حاصل شده اند که قابلیت بازسازی کامل شکل از تصویر اولیه را دارد .

۳- زیر تصاویر باند LL ناشی از این تجزیه همچنان که از گوشه پایین سمت راست به گوشه بالا سمت چپ حرکت می کنیم، از نظر اندازه کوچک می شود .

۴- هر یک از زیر تصویرهای به وجود آمده از این تجزیه از خود حساسیت های جهت دار نشان می دهند به این صورت که زیر تصویر بالا سمت راست لبه افقی در تصویر اصلی را ضبط می کند و پایین سمت چپ لبه عمودی را و پایین سمت راست نسبت به تغییرات قطری حساس است.

1 Approximation

2detail



الف



ب

ج

د

شکل (۳-۲): (الف) تبدیل موجک گسسته با استفاده از توابع پایه هارکه تغییرات هیستوگرام محلی آن نشان داده شده. (ب)-(د) چند تغریب مختلف $64 * 128$ ، $64 * 256$ ، $128 * 256$ که می توانند از الف به دست آیند.

می توان به صورت بسیار کوتاه خصوصیات تبدیل موجک را چنین شرح داد:

خصوصیت ۱: تفکیک پذیری، مقیاس پذیری، انتقال پذیری

خصوصیت ۲: قابلیت دقت چندگانه

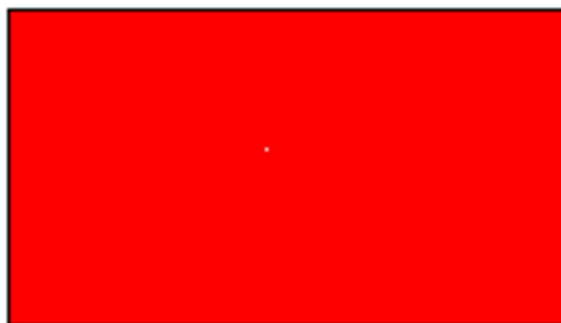
خصوصیت ۳: تعامد

۳-۳ JPEG

JPEG یک استاندارد فشرده سازی تصویر است که توسط گروه "Joint Photographic Experts Group" تولید شده است. JPEG به صورت رسمی به عنوان استاندارد بین المللی در سال ۱۹۹۲ پذیرفته شد. تا به امروز به عنوان یک روش رایج برای فشرده سازی تصاویر مورد استفاده است. JPEG یک روش فشرده سازی تصویر با اتلاف می باشد. این استاندارد، روش کد گذاری را با استفاده از DCT^۱ به کار می گیرد. استفاده از تبدیل DCT در JPEG بر پایه ۳ نظریه می باشد:

نظریه ۱: محتوای مفید تصویر به طور نسبتاً آهسته‌ای در طول تصویر تغییر می کند، برای مثال در یک بلوک تصویر ۸*۸، شدت تغییر زیاد در فضای کوچک و در زمان های متعدد غیر معمول است.

نظریه ۲: آزمایش های روانشناسی، نشان داده است که انسانها در صورت حذف مولفه های فرکانس بالای تصویر متوجه نمی شوند و به مولفه های فرکانس پایین حساس تر هستند و چشم انسان وجود یک تغییر کوچک در یک زمینه ثابت را خیلی زود درک می کند (مولفه فرکانس پایین) ولی وجود تغییرات در یک زمینه پر از تغییرات را براحتی درک نمی کند (شکل ۳-۳).

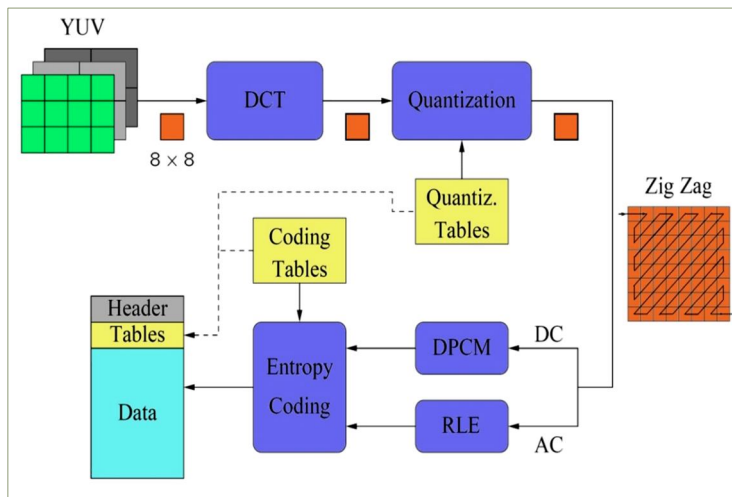


شکل (۳-۳) یک تغییر کوچک در یک زمینه ثابت

تبدیل گسسته کسینوسی^۱

نظریه ۳: چشم انسان به شدت روشنایی بسیار حساس تر است تا تغییرات رنگ در یک تصویر.

در زیر بلوک دیاگرام تبدیل jpeg را ملاحظه می فرمایید:



شکل (۳-۴): بلوک دیاگرام تبدیل JPEG

همانطور که در بلوک دیاگرام قبل مشخص است می توان مراحل الگوریتم فشرده سازی JPEG را

به صورت زیر شرح داد:

(۱) تبدیل مدل RGB به YIQ یا YUV

(۲) کاهش نمونه های I, Q (اطلاعات رنگ)

(۳) تبدیل تصویر به بلوکهای 8x8

(۴) محاسبه تبدیل DCT روی بلوکهای 8x8

در این حالت اگر تصویر ورودی را با $f(i,j)$ نمایش دهیم و تصویر تبدیل یافته را با $F(u,v)$ نمایش دهیم در مرحله کوانتیزه کردن تصویر تبدیل یافته F به صورت \hat{F} در می آید یعنی هر یک از ۶۴ جز فرکانسی بدست آمده برای هر بلوک بر ضرایب کوانتیزه کردن تقسیم شده و سپس به نزدیکترین عدد صحیح گرد می شود.

$$\hat{F}(u, v) = \text{round}\left(\frac{F(u, v)}{Q(u, v)}\right) \quad (22-3)$$

ماتریس Q، مقادیر استاندارد دارد. البته قابل تغییر نیز می‌باشد، برای ماتریس خروجی ناشی از تبدیل DCT طراحی شده، ماتریس Q(u, v) در پایین ترین گوشه راست، مقادیر بزرگتری دارد. این امر، باعث وارد کردن اتلاف بیشتر در فرکانس های بالاتر خواهد شد. مقادیر Q(u, v) پیش فرض از مطالعات روانشناسی و با هدف افزایش نرخ فشرده سازی و کاهش اتلاف های ادراکی تصاویر JPEG بدست آمده اند. مقادیر استفاده شده برای ماتریس Q، برای شدت روشنایی و اطلاعات رنگ متفاوت است که در زیر نمایش داده شده است.

16	11	10	16	24	40	51	61	17	18	24	47	99	99	99	99
12	12	14	19	26	58	60	55	18	21	26	66	99	99	99	99
14	13	16	24	40	57	69	56	24	26	56	99	99	99	99	99
14	17	22	29	51	87	80	62	47	66	99	99	99	99	99	99
18	22	37	56	68	109	103	77	99	99	99	99	99	99	99	99
24	35	55	64	81	104	113	92	99	99	99	99	99	99	99	99
49	64	78	87	103	121	120	101	99	99	99	99	99	99	99	99
72	92	95	98	112	100	103	99	99	99	99	99	99	99	99	99

Luminance

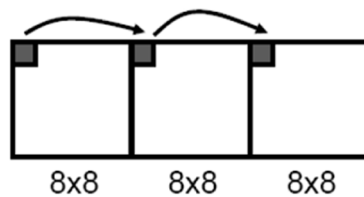
Chrominance

الف

ب

شکل (3-5): (الف) ماتریس مربوط به شدت روشنایی و (ب) ماتریس مربوط به اطلاعات رنگ

برای کد کردن اعداد از گوشه بالای سمت چپ بصورت مورب و زیگ زاگ جاروب شده و 64 ضریب پشت سر هم قرار می‌گیرند. اولین ضریب این رشته مؤلفه DC بلوک است و 63 ضریب بعدی مؤلفه های AC هستند به خاطر کوانتیزاسیون تعداد زیادی از ضرایب انتهایی صفر هستند به همین دلیل مقادیر AC به روش RLE کد می‌شوند مقادیر DC به روش DPCM کد می‌شوند.



شکل (۳-۶): محل قرار گرفتن ضرایب DC در تبدیل که به روش DPCM کد گذاری می‌شوند.

هر ضریب DC کد گذاری شده توسط DPCM با (AMPLITUDE, SIZE) نشان داده می‌شود.

SIZE	AMPLITUDE
1	-1, 1
2	-3, -2, 2, 3
3	-7, -4, 4, 7
4	-15, -8, 8, 15
.	.
.	.
.	.
10	-1023, -512, 512, 1023

شکل (۳-۷): نحوه نمایش ضرایب DC پس از کد گذاری توسط DPCM

در این حالت اولین مقدار DC را نگه می‌داریم سپس بقیه مقادیر را به صورت اختلاف آن با مقدار DC قبلی ثبت می‌کنیم. چون اعداد به دست آمده از این روش اعداد کوچکی هستند ذخیره سازی آن با یکی از روش‌های کد کردن آنتروپی امکان پذیر و بسیار موثر می‌شود.

۳-۴ شبکه عصبی [۹۷]

۳-۴-۱ پرسپترون

شبکه پرسپترون جزو معروف ترین شبکه‌های عصبی بوده و حالت چند لایه آنها (البته در این حالت از قانون یادگیری پرسپترون استفاده نمی‌کنند!) از پر کاربردترین شبکه‌های عصبی هستند .

خروجی پرسپترون $y = f(y_{in})$ است، که تابع فعال سازی آن به صورت زیر است:

$$f(y_{in}) = \begin{cases} 1 & \text{if } y_{in} > \theta \\ 0 & \text{if } -\theta \leq y_{in} \leq \theta \\ -1 & \text{if } y_{in} < -\theta \end{cases} \quad (23-3)$$

این تابع یک خروجی شبکه را تعیین می کند. علاوه بر دو خروجی معمول $+1$ (تعلق به دسته یا گروه) و -1 (عدم تعلق به دسته یا گروه) شامل ناحیه صفر بین θ و $-\theta$ نیز است که به بیانگر ناحیه‌ای است که روی آن تصمیم گیری نمی‌شود. طبق قانون آموزش پرسپترون، اگر خطایی برای الگوی ورودی آموزش رخ بدهد، وزن ها طبق این فرمول تغییر می کنند:

$$w_i(new) = w_i(old) + ax_it \quad (24-3)$$

در این رابطه، t بیانگر مقدار هدف بوده و مقدار آن $+1$ یا -1 است. a یا نرخ یادگیری¹ سرعت تغییر وزن‌ها را تعیین می کند و در ادامه بیشتر به آن پرداخته می‌شود. در این شبکه، اگر خطایی پیش نیاید وزن ها تغییر نمی کند و آموزش تا زمانی ادامه پیدا می کند که هیچ خطایی رخ ندهد.

قضیه ی همگرایی قانون یادگیری پرسپترون بیان می کند که اگر وزن هایی وجود داشته باشند که به شبکه امکان تولید پاسخ صحیح برای تمام الگوهای آموزش را بدهد، آنگاه روش یادگیری پرسپترون در هنگام تنظیم وزن ها این مقادیر را پیدا می کند.

1 Learning Rate

۳-۴-۲ شبکه ی پس انتشار استاندارد

در دهه ی ۱۹۷۰ توجه به شبکه های عصبی به میزان زیادی کاهش یافت که علت اصلی آن وجود محدودیت های مختلف شبکه های عصبی یک لایه در حل مسائل بود. اما کشف شبکه های عصبی چند لایه توسط چند محقق به طور جداگانه، به ویژه انتشار گسترده روش موثر آموزش این شبکه ها [۹۱] نقش عمده ای در پیدایش مجدد شبکه های عصبی به عنوان ابزاری برای حل مسائل گوناگون داشت. این روش آموزش پس انتشار خطا^۱ یا قانون دلتای تعمیم یافته^۲ نامیده می شود.

آموزش یک شبکه با الگوریتم پس انتشار شامل سه مرحله ی پیش خور^۳ کردن الگوی آموزش ورودی، محاسبه و پس انتشار کردن خطای مربوط و تنظیم وزن ها است. بعد از آموزش، شبکه فقط محاسبات مرحله ی پیش خور را لازم دارد.

در ادامه این بخش، الگوریتم پس انتشار استاندارد را به همراه ذکر برخی از نکات در طراحی این شبکه ها توضیح خواهیم داد. در پایان، روابط ریاضی الگوریتم آموزش و چکیده ی مختصری از قضیه های مربوط به قابلیت های شبکه های چند لایه ارائه شده است.

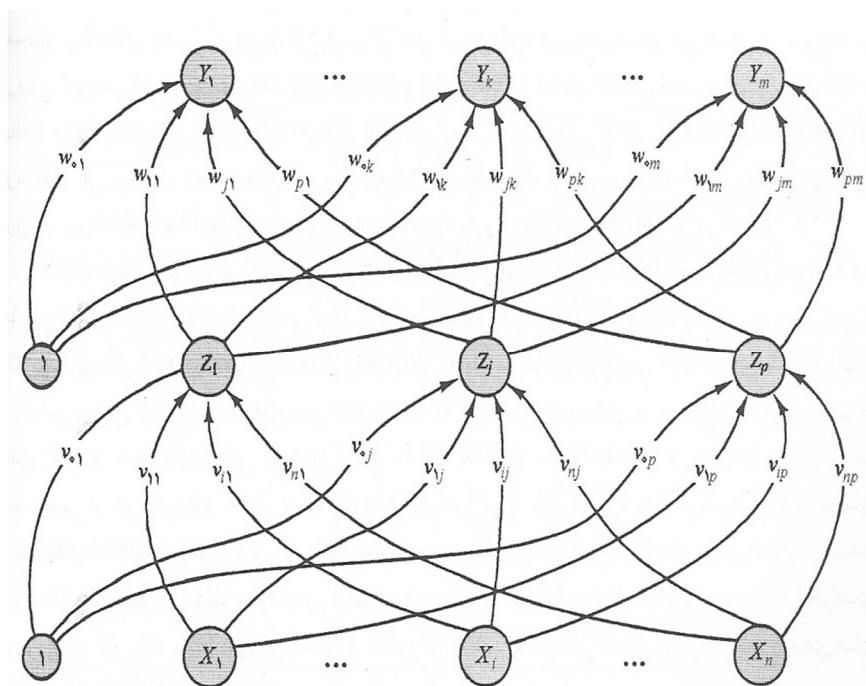
۳-۴-۲ ساختار

ساختار کلی یک شبکه چند لایه پس انتشار در شکل ۳-۱۰ نشان داده شده است.

1 Error BackPropagation

2 Delta Rule Generalized

3 Feedforward



شکل (۳-۸) ساختار یک شبکه ی عصبی پس انتشار با یک لایه ی مخفی [۹۷]

در شبکه ی عصبی نشان داده شده در این شکل؛ یک لایه ورودی (واحدهای X) ، یک لایه شامل واحدهای مخفی (واحدهای Z) و یک لایه خروجی (واحدهای Y) نشان داده شده است. همان طور که در این شکل دیده می شود. واحدهای خروجی و واحدهای مخفی می توانند بایاس نیز داشته باشند. بایاس روی واحد خروجی Y_k با w_{0k} نشان داده شده است و بایاس روی واحد مخفی Z_j با v_{0j} مشخص شده است. این بایاس ها همانند وزن های روی اتصالات از واحدهایی که خروجی آنها همیشه ۱ است، عمل می کنند.

۳-۴-۲ الگوریتم

همان طور که قبلاً گفته شد، آموزش یک شبکه با پس انتشار شامل سه مرحله است: پیش خور کردن الگوی آموزش ورودی، پس انتشار خطای مربوط و تنظیم وزن ها. در مرحله ی پیش خور، هر واحد ورودی x_i یک سیگنال ورودی دریافت می کند و این سیگنال را به هر یک از واحدهای Z_1, \dots, Z_p می فرستد. سپس هر واحد مخفی ، فعال سازی خود را محاسبه می کند و سیگنال خود، z_i را به

همه واحدهای خروجی می فرستند. هر واحد خروجی y_k فعال سازی خود را (برابر با y_k) محاسبه می کند تا پاسخ شبکه را برای الگوی ورودی ارائه شده تشکیل دهد.

همان طور که گفته شد، در آموزش با نظارت برای هر الگوی ورودی یک مقدار هدف نیز در دسترس است. در حین آموزش با پس انتشار، هر واحد خروجی فعال سازی محاسبه شده y_k خود را با مقدار هدف خود t_k ، مقایسه می کند تا خطای مربوط آن الگو برای آن واحد را مشخص کند. بر اساس این خطا، عامل δ_k که $k = (1, \dots, m)$ ، محاسبه می شود. عامل δ_k که در ادامه تشریح می شود، برای توزیع کردن مقدار خطای واحد خروجی y_k به تمام واحدهای لایه ی قبل (واحدهای مخفی که به y_k متصل هستند) مورد استفاده قرار می گیرد. در مرحله تنظیم وزن ها نیز از این عامل برای به روز کردن وزن های بین لایه ی خروجی و لایه ی مخفی استفاده می شود. به همین ترتیب، عامل δ_j ($j = 1, \dots, p$) برای هر واحد مخفی Z_j محاسبه می شود. از δ_j فقط برای به روز کردن وزن های بین لایه ی مخفی و لایه ی خروجی استفاده می شود. و لازم نیست خطای لایه مخفی را به لایه ی ورودی انتقال دهیم.

پس از این که تمامی عامل های δ معین شدند، وزن های شبکه برای تمام لایه ها به طور هم زمان تنظیم می شوند. تنظیم وزن w_{jk} (وزن واحد مخفی Z_j به واحد خروجی y_k) بر اساس عامل δ_k و فعال سازی واحد مخفی Z_j (مقدار Z_j) صورت می گیرد. تنظیم وزن v_{ij} (وزن واحد ورودی x_i ، به واحد مخفی Z_j) بر اساس عامل δ_j و فعال سازی واحد ورودی x_i صورت می گیرد.

علائم و نام هایی که در الگوریتم آموزش پس انتشار به کار می بریم، عبارتند از:

$$X \quad \text{بردار آموزش ورودی : } x = (x_1, \dots, x_i, \dots, x_n)$$

$$T \quad \text{بردار هدف خروجی : } t = (t_1, \dots, t_i, \dots, t_n)$$

δ_k بخشی از تنظیم وزن برای w_{jk} که مقدار آن با توجه به خطا در واحد خروجی y_k (اختلاف خروجی شبکه و خروجی هدف) حساب می شود. همچنین، این عاملی است که اطلاعات خطای رخ داده در واحد y_k را به واحدهای مخفی وارد شده به واحد y_k منتشر می کند.

δ_i بخش تنظیم وزن تصحیح خطا برای v_{ij} که بعد از پس انتشار اطلاعات خطا از لایه خروجی به واحد مخفی Z_j محاسبه می شود:

a نرخ یادگیری

x_i واحد ورودی i ، برای یک واحد ورودی، سیگنال ورودی و سیگنال خروجی آن واحد یکی بوده و برابر با x_i هستند.

v_{0j} بایاس در واحد مخفی j .

Z_j واحد مخفی j . ورودی واحد Z_j با z_in_j مشخص می شود و برابر است با :

$$(25-3) z_in_j = v_{0j} + \sum_j x_i v_{ij}$$

سیگنال خروجی (فعال سازی) Z_j ، z_j نام دارد و داریم: $Z_j = f(z_in_j)$

w_{0k} بایاس در واحد خروجی k .

y_k واحد خروجی k . ورودی شبکه به y_k با z_in_k نشان داده می شود و برابر است با :

$$(26-3) z_in_k = w_{0k} + \sum_j z_i w_{ik}$$

سیگنال خروجی (فعال سازی) Y_k با y_k نشان داده می شود و داریم: $y_k = f(z_in_k)$

۳-۴-۲-۳ تابع فعال سازی

برخلاف برخی از شبکه های دیگر، تابع فعال سازی مورد استفاده در شبکه های پس انتشار نمی تواند هر تابعی باشد و باید چند ویژگی مهم را داشته باشد. این تابع باید پیوسته، مشتق پذیر و به صورت یکنواخت نزولی باشد. علاوه بر این، برای داشتن کارایی محاسباتی، مشتق این تابع باید به راحتی قابل محاسبه باشد. مقدار مشتق توابع فعال سازی مورد استفاده در اینجا را می توان برحسب مقدار خود تابع بیان کرد. همچنین انتظار می رود که تابع فعال سازی قابلیت اشباع^۱ را داشته باشد، یعنی به صورت مجانبی به مقادیر بیشینه و کمینه خود نزدیک شود.

یکی از متداول ترین توابع فعال سازی که شرایط بیان شده را دارد، تابع سیگموئید دودویی است. دامنه ی این تابع (0,1) است و به صورت زیر تعریف می شود:

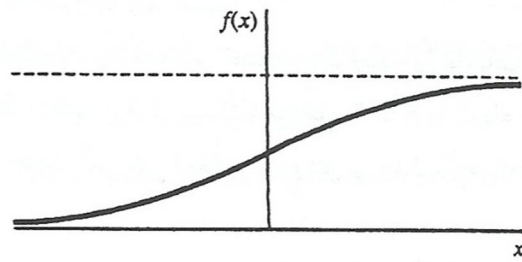
$$f_1(x) = \frac{1}{1+exp(-x)} \quad (27-3)$$

که مشتق آن برابر است با :

$$f_1^j(x) = f_1(x)[1 - f_1(x)] \quad (28-3)$$

نمودار این تابع در شکل ۳-۱۱ نشان داده شده است.

1 saturate



شکل (۹-۳) تابع سیگموئید دودویی با دامنه $(0,1)$

یکی دیگر از توابع فعال سازی رایج، تابع سیگموئید دو قطبی با دامنه $(-1,1)$ است که به صورت زیر تعریف می شود:

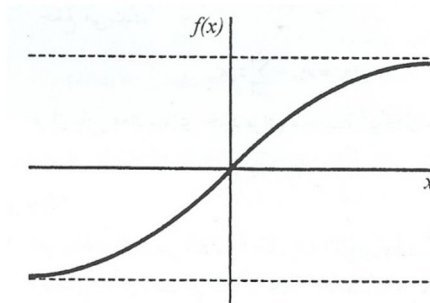
$$f_2(x) = \frac{2}{1+\exp(-x)} - 1 \quad (۲۹-۳)$$

و مشتق آن برابر است با :

$$f_2'(x) = \frac{1}{2} [1 + f_2(x)][1 - f_2(x)] \quad (۳۰-۳)$$

نمودار این تابع در شکل ۱۱-۳ نشان داده شده است. توجه کنید که تابع سیگموئید دو قطبی ارتباط نزدیکی با تابع تانژانت هیپربولیک زیر دارد (بخش ۱-۴-۳ را ببینید):

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (۳۱-۳)$$



شکل (۱۰-۳) تابع سیگموئید دو قطبی، دامنه $(-1,1)$

۳-۴-۲-۴ الگوریتم آموزش

هر یک از توابع فعال سازی تعریف شده در بخش قبل را می‌توان در الگوریتم پس انتشار استاندارد که در اینجا ارائه می‌شود، مورد استفاده قرار داد. توجه داشته باشید که به خاطر وجود رابطه‌ی ساده بین مقدار تابع و مشتق آن، نیاز به محاسبه‌ی بخش نمایی مشتق‌ها در مرحله‌ی پس انتشار الگوریتم نیست.

الگوریتم آموزش پس انتشار به صورت زیر است:

مرحله ی ۰: به وزن‌ها مقدار اولیه بدهید (مقادیر تصادفی کوچک را انتخاب کنید)

مرحله ی ۱: تا زمانی که شرایط توقف برقرار نیست. مراحل ۲ تا ۹ را انجام دهید.

مرحله ی ۲: برای هر جفت آموزش (مقادیر ورودی و هدف) مراحل ۳ تا ۸ را انجام دهید.

پیش‌خور:

مرحله ی ۳: هر واحد ورودی $(x_i, i = 1, \dots, n)$ سیگنال ورودی x_i را دریافت می‌کند و آن را در تمام واحدها در لایه ی بعدی (واحدهای مخفی) پخش می‌کند.

مرحله ی ۴: هر واحد مخفی $(z_j, j = 1, \dots, p)$ سیگنال‌های ورودی وزن دار خود را جمع می‌بندد.

$$z_in_j = v_{0j} + \sum_{i=1}^n x_i v_{ij} \quad (۳۲-۳)$$

و از تابع فعال سازی خود برای محاسبه ی سیگنال خروجی استفاده می‌کند، یعنی:

$$z_j = f(z_in_j) \quad (۳۳-۳)$$

و این سیگنال را به تمام واحدهای لایه ی بعدی (واحدهای خروجی) می فرستد.

مرحله ی ۵: هر واحد خروجی $(y_k, k = 1, \dots, m)$ سیگنال های ورودی وزن دار خود را جمع می بندد.

$$z_{in_k} = w_{0k} + \sum_{j=1}^p z_j w_{jk} \quad (34-3)$$

و از تابع فعال سازی خود برای محاسبه ی سیگنال خروجی استفاده می کند.

$$y_k = f(y_{in_k}) \quad (35-3)$$

پس انتشار خطا:

مرحله ۶: هر واحد خروجی $(y_k, k = 1, \dots, m)$ الگوی هدف متناظر با الگوی آموزش ورودی را دریافت می کند و خطا را محاسبه می کند.

$$\delta_k = (t_k - y_k) f'(y_{in_k}) \quad (36-3)$$

پارامتر تصحیح وزن را که بعداً در به روز کردن w_{jk} به کار می رود، محاسبه می کند:

$$\Delta w_{jk} = a \delta_k z_j \quad (37-3)$$

پارامتر تصحیح بایاس را که بعداً در به روز کردن w_{0k} به کار می رود، محاسبه می کند،

$$\Delta w_{0k} = a \delta_k \quad (38-3)$$

و δ_k (مقادیر دلتا) را به واحدهای لایه ی قبلی (لایه مخفی) می فرستد.

مرحله ی ۷: هر واحد مخفی ($Z_j, j = 1, \dots, p$) ورودی های دلتای خود را جمع می بندد (ورودی های دریافتی از واحدهای لایه ی خروجی).

$$\delta_{in_j} = \sum_{k=1}^m \delta_k W_{jk} \quad (39-3)$$

و آن را در مشتق تابع فعال سازی خود ضرب می کند تا پارامتر مربوط به اطلاعات خطا را محاسبه کند.

$$\delta_j = \delta_{in_j} f'(z_{in_j}) \quad (40-3)$$

تصحیح وزن خود را که بعداً برای به روز کردن v_{ij} به کار می رود، محاسبه می کند.

$$\Delta v_{ij} = a \delta_j x_i \quad (41-3)$$

و تصحیح بایاس خود را که بعداً برای به روز کردن v_{0j} به کار می رود، محاسبه می کند.

$$\Delta v_{0j} = a \delta_j \quad (42-3)$$

به روز کردن وزن ها و بایاس ها:

مرحله ی ۸: هر واحد خروجی ($y_k, k = 1, \dots, m$) وزن ها و بایاس های خود را به روز می کند:

هر واحد مخفی بایاس ها و وزن های خود را به روز می کند ($j = 0, \dots, p$):

$$w_{jk}(new) = v_{ij}(old) + \Delta v_{ij} \quad (43-3)$$

مرحله ی ۹: شرایط توقف را بررسی کنید.

دقت کنید که در پیاده سازی این الگوریتم ، برای دلتاهای واحدهای خروجی (δ_k در مرحله ی ۶) و دلتاهای واحدهای مخفی (δ_j در مرحله ی ۷) باید از آرایه های جداگانه استفاده کرد.

در اینجا یک دور^۱ معادل یک بار ارائه کل بردارهای آموزش است

مبنای ریاضی الگوریتم پس انتشار تکنیک بهینه سازی کاهش گرادیان^۲ است. بر مبنای این روش، گرادیان (شیب) تابع جهتی را که تابع در آن سریع تر افزایش می یابد، نشان می دهد و شیب با علامت منفی جهتی را نشان می دهد که در آن تابع سریع تر کاهش می یابد. در اینجا تابع مورد نظر تابع خطای شبکه است و متغیرهای مورد نظر که مشتق برحسب آنها گرفته می شود. وزن های شبکه هستند. بر اساس این اصل، روابط به روز کردن وزن های شبکه پس انتشار به دست می آیند.

۳-۵ آستانه گذاری [۹۵]

به علت ویژگیهای شهودی، سادگی پیاده سازی و سرعت محاسباتی، آستانه گذاری تصویر یک نقش مهمی در کاربردهایی چون تقطیع تصویر دارد. در این قسمت، ما آستانه گذاری را در یک مسیر مرسوم توضیح داده و روشهای موجود را بررسی می کنیم.

1 Epoch

2 Gradient Descent

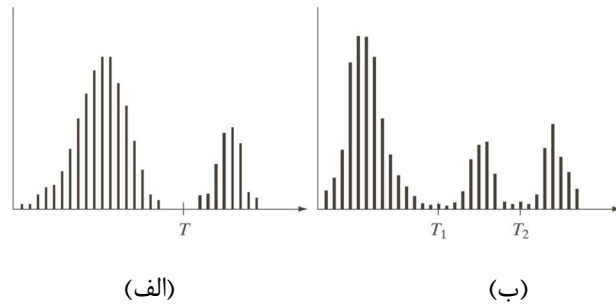
۳-۵-۱ مبانی آستانه گذاری شدت روشنایی

در این قسمت روشهایی برای تقسیم بندی تصویر به طور مستقیم بر مبنای مقادیر شدت روشنایی و خصوصیات این مقادیر توضیح می دهیم.

فرض کنید هیستوگرام شدت روشنایی شکل ۳-۱۳ (الف) متناظر با یک تصویر $f(x,y)$ است که متشکل از اشیای روشن روی زمینه تیره می باشد طوری که شدت روشنایی پیکسهای شیء و زمینه در دو حالت غالب گروه بندی شده اند. یک راه واضح برای استخراج اشیا از زمینه، انتخاب یک مقدار آستانه T است، طوری که این دو حالت را جدا کند. آنگاه هر پیکسل (x,y) که برای آن $f(x,y) > T$ یک پیکسل شیء و در غیر این صورت یک پیکسل زمینه خوانده می شود. به عبارت دیگر تصویر تقطیع شده $g(x,y)$ به صورت زیر است :

$$g(x,y) = \begin{cases} 1 & \text{اگر } f(x,y) > T \\ 0 & \text{اگر } f(x,y) \leq T \end{cases} \quad (۳-۴۴)$$

که T یک ثابت برای کل تصویر است، فرآیند که در این معادله بیان شد را آستانه گذاری سراسری گویند. اگر T در تصویر تغییر کند، عبارت آستانه گذاری متغیر استفاده می شود. گاهی عبارت آستانه گذاری محلی یا ناحیه ای نیز برای آن استفاده می شود. در این حالت مقدار T در هر نقطه (x,y) تصویر به خصوصیات همسایگی آن نقطه بستگی دارد (به عنوان مثال متوسط شدت روشنایی پیکسلها در آن همسایگی) اگر T به مختصات مکانی (x,y) بستگی داشته باشد، آستانه گذاری متغیر غالباً آستانه گذاری دینامیکی یا تطبیقی خوانده می شود. استفاده از این نامها عمومی نیست بلکه در مقالات پردازش تصویر با نام های متفاوت آنها را خواهیم دید.



شکل (۳-۱۱): هیستوگرام‌های شدت روشنایی که قابل جداسازی هستند (الف) توسط یک آستانه جدا کرد. (ب) به وسیله دو آستانه جدا کرد.

شکل ۳-۱۳ (ب) سه حالت غالب هیستوگرام برای توصیف تصویر نشان داده شده است (مثلاً دو نوع شیئی روشن روی یک زمینه تیره). اینجا آستانه گذاری چند سطحی برای دسته بندی یک پیکسل (x,y) استفاده می شود. اگر $f(x,y) \leq T_1$ پیکسل متعلق به پس زمینه است، اگر $T_1 < f(x,y) \leq T_2$ پیکسل در یک گروه از اشیاء قرار می گیرد اگر $f(x,y) > T_2$ پیکسل در گروه دیگر قرار می گیرد. در این صورت تصویر تقطیع شده به صورت زیر حاصل می شود:

$$g(x,y) = \begin{cases} a, & \text{اگر } f(x,y) > T_2 \\ b, & \text{اگر } T_1 < f(x,y) \leq T_2 \\ c, & \text{اگر } f(x,y) \leq T_1 \end{cases} \quad (۴۵-۳)$$

که در آن a, b, c سه مقدار شدت روشنایی مشخص هستند. مسائل تقطیع که به بیش از دو آستانه نیاز دارند، به دشواری حل می شوند (غالباً غیرممکن است). در این موارد نتایج بهتر با استفاده از دیگر روشها مثل آستانه گذاری متغیر، یا آستانه گذاری ناحیه ای به دست می آید.

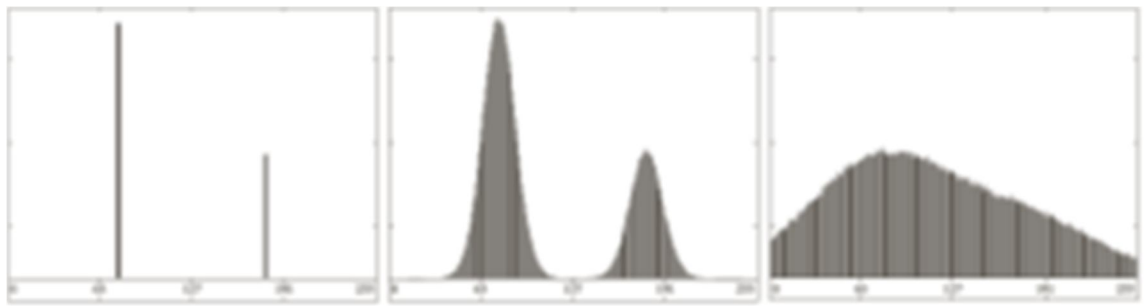
۳-۵-۲ نقش نویز در، آستانه گذاری تصویر



الف

ب

ج



د

ه

و

شکل (۳-۱۲): الف: تصویر ساختگی بدون نویز، تصویر اصلی با نویز گوسی میانگین صفر و انحراف معیار ۱۰، تصویر اصلی با نویز گوسی میانگین صفر و انحراف معیار ۵۰

برای توضیح اثرات نویز روی هیستوگرام تصویر شکل ۳-۱۴ (الف) را ببینید. این تصویر ساختگی ساده، بدون نویز است، در نتیجه هیستوگرام آن همان طور که شکل ۳-۱۴ (د) نشان می دهد، شامل دو ضربه است. تقطیع این تصویر به دو ناحیه کار ساده‌ای است که شامل قرار دادن یک آستانه هر کجای بین این دو ضربه است. شکل ۳-۱۴ (ب) تصویر اصلی که با نویز گوسی با میانگین صفر و انحراف معیار ۱۰ سطح شدت روشنایی خراب شده است، را نمایش می دهد. اگرچه حالت‌های هیستوگرام پهن‌تر شده‌اند (شکل ۳-۱۴ (ه))، جداسازی آنها به علت وجود دره عمیق بین حالت‌ها، ساده است. یک آستانه وسط دو قله قرار داده می‌شود تا جداسازی خوبی انجام شود. (شکل ۳-۱۴ (ج))

تصویر اصلی که با نویز گوسی با میانگین صفر و انحراف معیار ۵۰ سطح شدت روشنایی خراب شده است، را نمایش می دهد. همان طور که هیستوگرام شکل ۳ - ۱۴ (و) نشان می دهد، شرایط بسیار بحرانی است به طوری که هیچ راهی برای جدا کردن دو حالت وجود ندارد. بدون پردازش های اضافی ما امید کمی برای یافتن یک آستانه مناسب برای تقطیع کردن این تصویر داریم.

۳-۵-۳ نقش شدت روشنایی و انعکاس

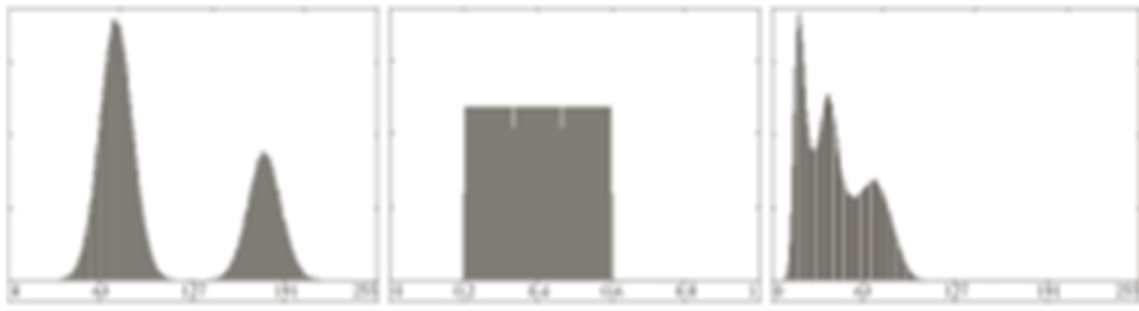
شکل ۳-۱۵ تاثیر شدت روشنایی روی هیستوگرام را نمایش می دهد. شکل ۳-۱۵ (الف) تصویر نویزی شکل ۳-۱۴ (ب) است و شکل ۳-۱۵ (د) هیستوگرام متناظر آن است. همان طور که گفته شد این تصویر با یک آستانه به راحتی تقطیع می شود. ما اثرات غیر یکنواختی شدت روشنایی را با ضرب تصویر شکل ۳-۱۵ (الف) در یک تابع شدت روشنایی متغیر مانند تابع شیب شدت روشنایی شکل ۳-۱۰ (ب) نشان می دهیم. هیستوگرام متناظر آن در شکل ۳-۱۵ (و) نشان داده شده است. شکل ۳-۱۵ (ج) حاصل ضرب تصویر و این نمونه سایه دار را نشان می دهد. همان طور که شکل ۳-۱۵ (و) نشان می دهد، عمق دره بین قله ها خراب شده است. جداسازی حالتها بدون پردازش اضافی غیر ممکن است. اگر شدت روشنایی کاملاً یکنواخت بود، باز هم نتایج مشابه به دست می آمد، اما برای انعکاس تصویر به علت تغییرات انعکاسی سطح اشیا و پس زمینه، اینچنین نیست .



الف

ب

ج



د

ه

و

شکل (۳-۱۳): (الف) تصویر نویزی (ب) شیب شدت روشنایی در بازه $[0.2-0.6]$ (ج) حاصل ضرب (الف) و (ب).

(د) - (و) هیستوگرام‌های متناظر

نکته کلیدی در پاراگراف گذشته، نقش یکسانی است که شدت روشنایی و انعکاس در تقطیع تصویر فوق با استفاده از آستانه گذاری یا هر روش دیگر بازی می کنند

۳-۵-۴ آستانه گذاری سراسری ساده

همان طور که گفتیم زمانی که توزیعهای شدت روشنایی اشیا و پیکسلهای پس زمینه به اندازه کافی متمایز باشند، استفاده از یک آستانه (سراسری) برای کل تصویر قابل استفاده است. الگوریتم تکرار زیر برای تعیین آستانه سراری استفاده می شود:

۱. انتخاب یک تخمین اولیه برای مقدار آستانه سراسری T .

۲. تقطیع تصویر با استفاده از T و معادله زیر بدین ترتیب دو گروه پیکسل تولید می شود $G1$ شامل همه پیکسلها با مقادیر شدت روشنایی بزرگتر از T ، و $G2$ شامل پیکسلهای با شدت روشنایی کمتر یا مساوی T .

$$g(x, y) = \begin{cases} 1 & f(x, y) > T \quad \text{اگر} \\ 0 & f(x, y) \leq T \quad \text{اگر} \end{cases} \quad (46-3)$$

۳. محاسبه متوسط مقادیر شدت روشنایی M_1 و M_2 به ترتیب برای پیکسلهای گروه $G1$ و $G2$

۴. محاسبه یک آستانه جدید:

$$T = 0.5(M_1 + M_2) \quad (47-3)$$

۵. تکرار مرحله ۲ تا ۴ این که اختلاف بین مقادیر آستانه T در دو تکرار پشت سر هم کمتر از پارامتر از پیش تعریف شده ΔT باشد.

۳-۵-۵ آستانه چند گانه

رجوع شود به [۹۵]

۳-۵-۶ آستانه گذاری متغیر

یکی از ساده ترین روش های آستانه گذاری متغیر، تقسیم بندی تصویر به مستطیل های ناهمپوشان است. این روش برای جبران غیر یک نواختی شدت روشنایی و یا انعکاس تصویر استفاده می شود. این مستطیل ها به اندازه کافی کوچک انتخاب می شوند. تا شدت روشنایی در هر مستطیل تقریباً یکنواخت باشد.

۳-۵-۶-۱ آستانه گذاری متغیر بر مبنای خصوصیات محلی تصویر

یک روش کلی تر از تقسیم بندی تصویر، محاسبه یک آستانه در هر نقطه (x,y) ، تصویر بر مبنای یک یا چند ویژگی خاص که در همسایگی پیکسل (x,y) محاسبه شده است می باشد. اگر چه ممکن است فرایند دشواری به نظر بیاید، ولی الگوریتم ها و سخت افزار های مدرن، پردازش سریع همسایگی را ممکن ساخته.

ما روش ساده آستانه گذاری محلی را با استفاده از انحراف معیار و میانگین پیکسل ها در یک منطقه همسایگی هر نقطه تصویر، توضیح می دهیم این دو کمیت برای تعیین آستانه های محلی بسیار مناسب هستند، زیرا این دو کمیت توصیف گرهای متوسط شدت روشنایی و تباین محلی هستند. فرض کنید در یک تصویر δ_{xy} و m_{xy} به ترتیب انحراف معیار و میانگین مجموعه پیکسلهای منطقه همسایگی (x,y) باشند مقدار آستانه محلی از روی فرم های رایج متغیرها، به صورت زیر بیان می شود:

$$T_{xy} = a\delta_{xy} + bm_{xy} \quad (62-3)$$

که a و b ، ثابت های غیر منفی هستند و

$$T_{xy} = a\delta_{xy} + bm_G \quad (63-3)$$

که m_G میانگین سراسری تصویر است. تصویر تقطیع شده به صورت زیر محاسبه می شود:

$$g(x,y) = \begin{cases} 1, & f(x,y) > T_{XY} \\ 0, & f(x,y) < T_{XY} \end{cases} \quad (64-3)$$

که $f(x,y)$ تصویر ورودی است. این معادله برای همه پیکسل های تصویر ارزیابی می شود، و برای هر موقعیت (x,y) با استفاده از پیکسل های همسایگی s_{xy} ، یک آستانه متفاوت محاسبه می شود.

قدرت کافی می تواند به آستانه گذاری محلی اضافه شود در صورتی که گذاره هایی بر مبنای پارامترهای محاسبه شده از روی همسایگی (x,y) استفاده شود:

$$g(x, y) \begin{cases} 1, & \text{شرط صحیح} \\ 0, & \text{شرط غلط} \end{cases} \quad (۶۵-۳)$$

که در آن شرط Q گزاره‌های است که بر مبنای عاملیت‌های محاسبه شده با استفاده از پیکسل‌های همسایگی S_{xy} به دست می‌آید.

۳-۵-۶-۲ استفاده از جابجایی میانگینها

یک روش خاص روش آستانه گذاری محلی بر اساس محاسبه میانگین متحرک^۱ در طول خط روبش تصویر است. این پیاده سازی برای پردازش متن بسیار مناسب است، چرا که در آن کاربرد سرعت ملزومات مهم است. معمولاً روبش تصویر به صورت خط به خط و زیگزاگ انجام می‌شود و بدین ترتیب بایاس شدت روشنایی کاهش میابد. اگر Z_{k+1} شدت روشنایی نقطه باشد که روبشگر در گام $k+1$ با آن مواجه شده است، شدت روشنایی در آن نقطه به صورت زیر داده می‌شود:

$$\begin{aligned} m(k+1) &= \frac{1}{n} \sum_{i=k+2-n}^{k+1} Z_i \\ &= m(k) + \frac{1}{n} (Z_{k+1} - Z_{k-n}) \end{aligned} \quad (۶۶-۳)$$

که در آن n تعداد نقاطی است که برای محاسبه متوسط استفاده شده است از آنجا که میانگین متحرک برای هر نقطه از تصویر محاسبه می‌شود، تقطیع با استفاده از معادله ۶۴ با $T_{xy} = b m_{xy}$ انجام می‌شود، که b عدد ثابت و m_{xy} مقدار میانگین متحرک از معادله ۶۶ در نقطه (x, y) تصویر ورودی است.

¹ moving average

فصل چهارم

روش های پیشنهادی جهت ایجاد تصویر دودویی و فشرده سازی آنها

۴-۱ مقدمه

دودویی سازی تصاویر متنی یکی از مراحل مهم در برخی کاربردها مانند بازشناسی و فشرده سازی تصاویر اسناد است. امروزه روشهای بسیار زیادی برای انجام این کار پیشنهاد شده است. از این میان، برخی روشها به سرعت انجام محاسبات، برخی به وفقی بودن انتخاب آستانه، برخی به استفاده از تحلیل چند مقیاسی برای انتخاب مقدار آستانه، و برخی نیز به ترکیب چند روش دودویی سازی توجه نشان داده اند. در هر صورت، طبق دانسته های ما، تاکنون روشی که همه این ویژگیها را همزمان در نظر گرفته باشد معرفی نشده است.

در این فصل، روشی برای دودویی سازی تصاویر متنی پیشنهاد می شود که در چارچوب مدل محتوی ترکیبی به منظور تولید لایه های مختلف قابل استفاده بوده و هر چهار ویژگی نام برده شده را دربر دارد. در این روش برای ترکیب دو روش دودویی سازی، از شبکه عصبی و برای تحلیل چندمقیاسی از فیلتر بانک استفاده شده است.

معمولاً روشهای متعلق به تجزیه تصویر به چند لایه، تصویر متنی را به سه تصویر یا لایه شامل لایه ی ماسک، لایه ی پیش زمینه و لایه ی پس زمینه تجزیه و هر کدام را جداگانه فشرده سازی می کنند. در هر صورت، در گروههای دوم و سوم نیاز به انجام تبدیلی برای فشرده سازی است.

طبق دانسته های ما، تاکنون از تبدیل KLT برای فشرده سازی تصاویر متنی استفاده ای نشده است. یکی از مهمترین ویژگیهای این تبدیل این است که انرژی سیگنال را به طور بهینه در ضرایب تبدیل متراکم می کند؛ به بیان دیگر با در دست داشتن تنها قسمت کوچکی از ضرایب تبدیل می توان

عمده‌ی انرژی اولیه سیگنال را بازسازی کرد. گرچه عمده‌ی تبدیلهای موجود به ویژه تبدیل DCT و تبدیل موجک چنین قابلیت‌هایی دارند اما این قابلیت به طور بهینه فقط در تبدیل KLT وجود دارد. با وجود برخورداری تبدیل KLT از قابلیت بهینه‌سازی تراکم انرژی در کمترین تعداد ضریب، امروزه از این تبدیل در فشرده‌سازی تصاویر به ویژه تصاویر متنی استفاده‌ای نشده است. به نظر می‌رسد دو دلیل عمده برای این امر وجود دارد: یکی حجم محاسباتی نسبتاً بالا برای محاسبه مقادیر و بردارهای ویژه و نیز ماتریس کواریانس و دیگری نیاز به در دست داشتن تعداد نمونه‌های نسبتاً زیاد برای محاسبه‌ی دقیقتر ماتریس کواریانس. البته دلیل اول تنها برای کاربردهایی مهم است که نیاز به پیاده‌سازی سخت‌افزاری وجود داشته باشد. امروزه در کاربردهای صرفاً نرم‌افزاری با وجود کامپیوترهای شخصی پرسرعت، دیگر حجم محاسبات بالا به عنوان مشکل محسوب نمی‌شود.

دو روش پیشنهادی برای فشرده‌سازی تصاویر متنی عبارتند از:

روش فشرده‌سازی پیشنهادی اول در این پایان‌نامه از تبدیل KLT برای فشرده‌سازی تصاویر متنی استفاده می‌کند. در این روش پیشنهادی نشان داده می‌شود، که چگونه با در دست داشتن تنها یک تصویر می‌توان نمونه‌های مورد نیاز برای محاسبه‌ی تبدیل را تهیه نمود.

در روش پیشنهادی دوم در این پایان‌نامه روشی جهت پیش‌پردازش تصاویر متنی قبل از اعمال استاندارد JPEG، پیشنهاد می‌شود که باعث حفظ تقریبی کیفیت تصویر متنی می‌شود. در واقع در این روش هدف این است که از استانداردهای موجود (به طور خاص استاندارد JPEG) به طور بهینه‌تری در جهت فشرده‌سازی تصاویر متنی استفاده شود. البته، این روش پیشنهادی قابل اعمال به حالت استاندارد JPEG2000 نیز می‌باشد. روش پیشنهادی دوم مبتنی بر وزن دادن به ضرایب فرکانس میانی و دوباره‌چینی ضرایب ماتریس تبدیل DCT است. نتایج آزمایشها روی دسته‌ای از تصاویر متنی نشان می‌دهد که در SNR یکسان، روش پیشنهادی باعث افزایش میزان فشرده‌سازی

استاندارد JPEG به طور متوسط تا ۱,۵ برابر می‌شود. طبق دانسته‌ی ما، تاکنون کار چندانی برای استفاده موثر از استاندارد JPEG در فشرده‌سازی تصاویر متنی انجام نشده است.

۲-۴ روش پیشنهادی جهت دودویی سازی تصاویر متنی

در کاربرد فشرده‌سازی تصاویر متنی از چند جنبه نیاز به دودویی‌سازی است: از آن جا که متن موجود در یک تصویر متنی اهمیت بسیار بیشتری نسبت به بقیه اجزای تصویر دارد، در فشرده‌سازی با اتلاف تصاویر متنی یک روش این است که به جای ذخیره‌ی تصویر متنی رنگی یا سطح خاکستری، آن را دودویی کرده و تصویر حاصل را فشرده کنیم. این روش باعث افزایش قابل توجه میزان فشرده‌سازی می‌شود.

در این پایان‌نامه برای فشرده‌سازی تصاویر اسناد از مدل محتوای ترکیبی یا MRC^۱ استفاده می‌شود که در این مدل نیاز به استفاده از یک روش دودویی‌سازی جهت تشکیل لایه‌ی پوشش می‌باشند [۹۲].

روشهایی که امروزه برای دودویی‌سازی تصاویر متنی پیشنهاد شده‌اند، یا تنها به انجام دودویی کردن بسنده کرده‌اند و یا این که علاوه بر عمل مذکور به یکی از پارامترهای سرعت انجام محاسبات، وفقی بودن آستانه‌ی انتخاب شده، استفاده از تحلیل چند مقیاسی برای انتخاب آستانه، در کار خود توجه داشته‌اند.

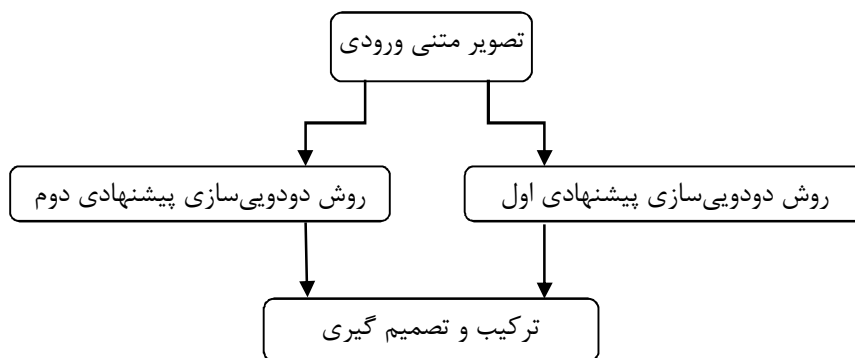
بالا بودن سرعت انجام محاسبات در پیاده‌سازی زمان- حقیقی و سخت‌افزاری اهمیت دارد. اهمیت وفقی بودن آستانه‌ی انتخاب شده در تصاویری معلوم می‌شود که برخی ویژگیها مانند شدت روشنایی پس‌زمینه به طور یکنواخت در تمام نواحی تصویر متنی وجود نداشته باشند. در تحلیل چند مقیاسی میتوان تصویر متنی را در چند مقیاس مختلف تجزیه و بررسی کرد تا آستانه‌ی مناسبتری انتخاب کرد. اهمیت ترکیب روشهای دودویی‌سازی در این است که با بهره‌گیری از محاسن چندین روش

1 Mixed Raster Content

دودویی‌سازی، کارایی کلی آن بهتر از کارایی هر یک از روشهای استفاده شده به تنهایی خواهد شد. طبق دانسته‌ی ما، تاکنون روشی که توانسته باشد همزمان به همه‌ی این ویژگیها دست پیدا کند، معرفی نشده است [۹۲].

در این فصل، یک روش دودویی‌سازی پیشنهاد می‌شود که در آن علاوه بر انجام دودویی کردن تصویر متنی ورودی، به کاهش حجم محاسبات، انتخاب وفقی آستانه، استفاده از تحلیل چند مقیاسی برای انتخاب دقیقتر آستانه، و ترکیب مناسب دو روش دودویی‌سازی نیز توجه نشان داده شده است. همه‌ی این اعمال شامل حجم نسبتاً کمی از اعمال جمع و ضرب (گاهاً موازی هم) هستند که این امر باعث دست‌یافتن به دو مزیت پایین بودن حجم محاسبات و قابلیت پیاده‌سازی در بسترهای سخت‌افزاری مناسب مانند مدارات FPGA و یا پردازشگرهای سیگنال DSP می‌شوند. که به ذات دارای توانایی انجام محاسبات به صورت موازی و مجزا از هم می‌باشند.

دیاگرام بلوکی روش دودویی‌سازی پیشنهادی در شکل (۴-۱) نشان داده شده است. همان طور که ملاحظه می‌شود ابتدا دو روش دودویی‌سازی مختلف به موازات یکدیگر روی تصویر متنی ورودی اعمال می‌شوند. سپس، نتیجه‌های این دو روش با هم ترکیب و تصویر دودویی نهایی به دست آورده می‌شود. هر یک از زیربخشهای که در ادامه می‌آیند، یکی از بلوکهای دیاگرام مذکور را توضیح می‌دهند.



شکل (۴-۱): دیاگرام بلوکی روش دودویی‌سازی پیشنهادی

۴-۲-۱ روش دودویی سازی پیشنهادی اول

اولین روش دودویی سازی پیشنهادی مبتنی بر استخراج یک رابطه برای محاسبه‌ی وفقی یک آستانه‌ی محلی است. در این روش، برای هر پیکسل متعلق به تصویر متنی واقع در مختصات (X, Y) ، در قدم اول، یک بلوک مربعی به ابعاد $N \times N$ در حالت بهینه باید ۵ برابر عرض قلم باشد ولی با توجه به اینکه در این پایان‌نامه از تصاویر استاندارد متنی استفاده می‌شود می‌توان عرض N را به طور تجربی و بر حسب آزمایش و خطا مقدار ثابتی فرض کرد، که در بیان نتایج لحاظ شده است) و به مرکزیت آن پیکسل در نظر گرفته می‌شود. سپس از ویژگیهای آماری میانگین و انحراف معیار به کمک رابطه‌ی زیر برای محاسبه‌ی آستانه‌ی مربوط به آن پیکسل استفاده می‌شود.

$$T_{x,y} = \mu + \frac{\mu + P \cdot \delta}{\sigma \cdot M - 1} \quad (1-4)$$

در این رابطه، پارامترهای μ و σ به ترتیب میانگین و انحراف معیار پیکسل‌های واقع در بلوک مربعی مذکور، پارامتر M یک عدد ثابت، و پارامتر P ضریبی وابسته به پارامترهای میانگین و انحراف معیار می‌باشند. رابطه‌ی پیشنهادی (۱-۴) از این واقعیت به دست آمده است که مقدار آستانه معمولاً نسبت به میانگین فاصله‌ای دارد که این فاصله تابعی از انحراف معیار است. برای مثال در روش دودویی سازی آتسو، مقدار آستانه از رابطه‌ی زیر به دست می‌آید:

$$T = \mu + k \cdot \sigma \quad (2-4)$$

که پارامتر k عدد ثابتی است.

در رابطه‌ی (۱-۴) برای به دست آوردن رابطه‌ی پارامتر P از دو فرض استفاده می‌کنیم: فرض اول این است که اگر انحراف معیار پیکسل‌های موجود در بلوک مربعی بسیار کوچک (یعنی، بلوک مذکور یک ناحیه‌ی یکنواخت) باشد، مقدار آستانه‌ی T باید به صفر میل کند تا در نتیجه تمام پیکسل‌های مذکور جزو ناحیه‌ی غیرمتنی طبقه‌بندی شوند. فرض دوم این است که برای انحراف معیارهای بزرگ (و نزدیک به مقدار میانگین، $\sigma \rightarrow \mu$) مقدار آستانه باید برابر مجموع مقدار میانگین و $\frac{\sigma}{M}$ شود، در این

حالت به این نتیجه میرسیم که ناحیه مذکور در ناحیه متنی واقع شده و چون پیکسل های ناحیه متنی در اقلیت قرار دارند ما اقدام به تقویت آنها می کنیم. با توجه به دو فرض فوق، رابطه ی P به صورت زیر در می آید:

$$p = (\mu - 1) - \frac{1}{M} \quad (3-4)$$

در این حالت با افزایش آستانه باعث می شویم بیشتر نواحی به سمت سیاه (ناحیه متنی) تغییر یابند در این صورت ناحیه متنی را که از لحاظ تعداد پیکسل نسبت به پس زمینه از تعداد بسیار کمتری برخوردار است تقویت کرده و احتمال ایجاد پیکسل سیاه را در این نواحی افزایش می دهیم.

بعد از این که مقدار آستانه ی مربوط به پیکسل فعلی در بلوک مورد نظر محاسبه شد، در قدم بعدی به کمک فیلتر بانک مبتنی بر موجک هار، دو تقریب از تصویرمتنی اصلی به دست آورده و سپس با صفر کردن ضرایب جزئیات تصاویر را دوباره بازسازی می کنیم تا به اندازه اصلی برگردند. با این تفاوت که تصاویر بدست آمده ناشی از این تبدیلات حالتی همچون عبور از فیلترهای پایین گذر داشته و نرم تر و یک نواخت تر شده اند. و بسته به عمق تجزیه، در آنها از بلوکهای مربعی با ابعاد به ترتیب $2 \times N$ و $4 \times N$ استفاده کرده و فرآیند توضیح داده شده برای محاسبه ی آستانه را برای این بلوکها تکرار می کنیم. بنابراین، برای هر پیکسل از تصویر متنی اصلی، سه مقدار آستانه به دست می آوریم که این مقادیر در هنگام ترکیب دو روش دودویی سازی پیشنهادی (زیربخش ۴-۲-۳) مورد استفاده قرار می گیرند.

حداقل به دو علت از موجک هار در ساختار فیلتر بانک استفاده کرده ایم: یکی سادگی و سرعت محاسبات و پیاده سازی آن است و دیگری این است که نواحی لبه شامل تغییرات ناگهانی می باشند و موجک هار نیز این رفتار را به خوبی از خود نشان می دهد.

۴-۲-۲ روش دودویی سازی پیشنهادی دوم

دومین روش پیشنهادی برای دودویی سازی تصویر متنی استفاده از شبکه عصبی پرسپترون است. در این روش، ابتدا تصویر متنی را با بلوکهای مربعی به ابعاد N که با یکدیگر در راستای افقی و عمودی ۵۰ درصد همپوشانی داشته باشند، تقسیم می‌کنیم. بنابراین، هر پیکسل از تصویر اصلی، در حالت کلی، متعلق به چهار بلوک از بلوکهای مذکور است. برای هر یک از این چهار بلوک، یک بردار ویژگی شامل دو ویژگی میانگین و واریانس پیکسلهای مربوطه محاسبه می‌کنیم. بنابراین، برای هر پیکسل از تصویر متنی اصلی، هشت ویژگی محاسبه می‌شود که این ویژگیها ورودی یک شبکه‌ی عصبی پرسپترون سه لایه با قاعده‌ی یادگیری پس انتشار خطا را تشکیل می‌دهند. این شبکه عصبی در لایه‌ی اول خود دارای هشت گره، در لایه‌ی دوم (یا میانی) دارای ده گره و در لایه‌ی خروجی دارای یک گره می‌باشد. مقدار خروجی شبکه، در حقیقت همان آستانه‌ی انتخاب شده متناظر با بلوکهای به ابعاد N است، که برای مقایسه با شدت روشنایی پیکسل مورد نظر بدست آمده است. در ادامه، از همان دو تصویر تقریب به دست آمده در زیربخش قبل و بلوکهای با همان ابعاد گفته شده (یعنی ابعاد $2 \times N$ و $4 \times N$) استفاده کرده و برای هر بلوک، فرآیند گفته شده در ابتدای این زیربخش را تکرار می‌کنیم تا برای هر بلوک یک آستانه به دست آید. بنابراین، در روش پیشنهادی دوم، برای هر پیکسل متعلق به تصویر متنی اصلی، سه مقدار آستانه به دست می‌آید که در زیربخش بعدی از آنها استفاده می‌شود.

در ضمن برای آموزش شبکه عصبی و تعیین مقدار آستانه مورد نظر از نرم افزار ACDSEE استفاده شد که در آن مقدار آستانه مناسب برای تصویر به صورت دستی در این قسمت با تغییر دادن‌های زیاد به دست آمد. بدین صورت که مقدار آستانه را آنقدر تغییر می‌دهیم تا از لحاظ دیداری به بهترین آستانه دست پیدا کنیم، این مقدار آستانه به عنوان تارگت در شبکه عصبی مورد استفاده قرار می‌گیرد.

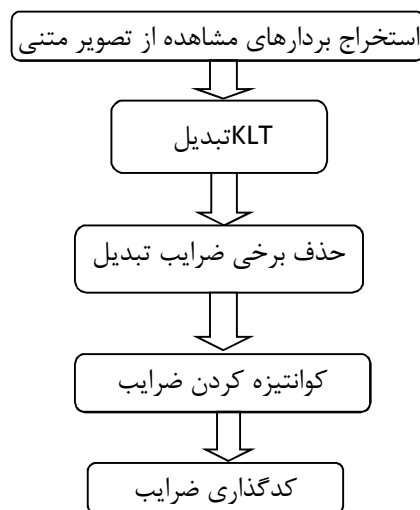
۳-۲-۴ ترکیب دو روش پیشنهادی و تصمیم‌گیری

هر روش دودویی‌سازی در حقیقت یک طبقه‌بند محسوب می‌شود که هر پیکسل از تصویر متنی را به یکی از دو طبقه‌ی متنی و غیرمتنی نسبت می‌دهد. در این مرحله، برای هر پیکسل از تصویر متنی ورودی، هر یک از شش آستانه‌ی محاسبه شده در دو زیربخش قبل، ورودی یک شبکه‌ی عصبی پرسپترون سه لایه با قاعده‌ی یادگیری پس انتشار خطا را تشکیل می‌دهند. این شبکه عصبی در لایه‌ی اول خود دارای شش گره، در لایه‌ی دوم (یا میانی) دارای هشت گره و در لایه‌ی خروجی دارای یک گره می‌باشد. مقدار خروجی شبکه، در حقیقت همان آستانه‌ی انتخاب شده نهایی برای دودویی‌سازی است که پیکسل مورد نظر با آن مقایسه شده و نوع آن پیکسل که به کدام گروه تعلق دارد را مشخص می‌کند.

۳-۴ روشهای پیشنهادی جهت فشرده سازی تصاویر متنی

۱-۳-۴ روش اول

این روش مبتنی بر تبدیل KLT می‌باشد و بلوک دیاگرام این روش فشرده‌سازی پیشنهادی، برای تصاویر متنی، در شکل (۲-۴) نشان داده شده است.



شکل (۲-۴) بلوک دیاگرام روش فشرده‌سازی پیشنهادی اول

در بلوک اول هر سطر از تصویر متنی به تعدادی زیرسطر با طول متساوی NL تقسیم می‌شود. اگر طول یک سطر مضرب صحیحی از NL نباشد، به تعداد کافی صفر به آن افزوده می‌شود. بنابراین از هر سطر تصویر متنی تعدادی زیر سطر به دست می‌آید که هر کدام به عنوان یک بردار مشاهده از یک بردار تصادفی مفروض در نظر گرفته می‌شوند. با این کار تعداد قابل ملاحظه‌ای بردار مشاهدات به دست می‌آید که در تخمین نسبتاً دقیق ماتریس کواریانس مفید می‌باشند. بنابراین، مشکلی که در ابتدای این فصل برای استفاده از تبدیل KLT ذکر شد به این طریق قابل حل است. در بلوک دوم، تبدیل KLT به بردارهای مشاهدات اعمال می‌شود. در بلوک سوم، عددی مانند $m < N$ انتخاب می‌کنیم که تعداد مولفه‌های مورد نیاز در بردار تبدیل بوده و طوری انتخاب می‌شود که انرژی مربوط به این تعداد مولفه، $\sum_{i=0}^{m-1} \lambda_i / \sum_{i=0}^{N-1} \lambda_i$ بزرگتر یا مساوی یک مقدار آستانه‌ی از قبل مشخص باشد. حال برای تشکیل ماتریس تبدیل $N \times m$ از m بردار ویژه ماتریس کواریانس متناظر با بزرگترین m مقدار ویژه آن به عنوان ستونهای ماتریس تبدیل Φ_m مورد نظر استفاده می‌کنیم. بردار تبدیل یافته y یک بردار $m \times 1$ است که نسبت به بردار سیگنال اولیه x ابعاد آن کاهش یافته و بنابراین فشرده شده است. تعداد قابل ملاحظه‌ای از ضرایب تبدیل محاسبه شده حذف می‌شوند. ضرایب باقیمانده در بلوک چهارم کوانتیزه می‌شوند. برای کوانتیزه کردن از بردار کوانتیزاسیونی استفاده می‌شود که برگرفته از ماتریس کوانتیزاسیون مورد استفاده در استاندارد JPEG می‌باشد. برای این کار، ضرایب ماتریس کوانتیزاسیون مذکور به ترتیب زیگزاگ انتخاب شده و در بردار کوانتیزاسیون قرار داده می‌شوند. در آخرین بلوک، یعنی بلوک پنجم، از کدگذاری حسابی وفقی برای کدگذاری ضرایب کوانتیزه شده و تبدیل آنها به دنباله‌ی بیتی استفاده می‌شود.

۴-۳-۲ روش دوم

روش پیشنهادی دوم مبتنی بر پردازش تصاویر متنی به منظور استفاده موثر از استاندارد JPEG در کاربرد فشرده‌سازی تصاویر متنی است. در حقیقت به دنبال استفاده از استانداردهای عمومی موجود

در کاربردهای خاص (اینجا، کاربرد فشرده‌سازی تصاویر متنی) می‌باشیم. استاندارد فشرده‌سازی JPEG مبتنی بر تبدیل DCT بوده و ویژه تصاویر طبیعی طراحی شده است. اثر این استاندارد روی تصاویر متنی، به دلیل ویژگی متراکم‌سازی تبدیل DCT و انجام کوانتیزاسیون ضرایب تبدیل، مانند اعمال یک فیلتر پایین گذر است. از طرفی اطلاعات مهم یک تصویر متنی در لبه‌های آن قرار دارند که عمدتاً متناظر با فرکانسهای میانی هستند. بنابراین، اعمال فشرده‌سازی JPEG باعث افت کیفیت دیداری تصویر متنی می‌شود. با توجه به مقبولیت و عمومیت استاندارد JPEG لزوم تصحیح یا تعمیم آن به تصاویر متنی ضروری به نظر می‌رسد.

در این بخش روشی جهت پیش‌پردازش تصاویر متنی قبل از اعمال روش JPEG پیشنهاد می‌شود که باعث حفظ تقریبی کیفیت تصویر متنی می‌شود. البته، روش پیشنهادی قابل اعمال به حالت استاندارد JPEG2000 نیز می‌باشد. روش پیشنهادی مبتنی بر وزن دادن به ضرایب فرکانس میانی و دوباره‌چینی ضرایب ماتریس تبدیل DCT است. نتایج آزمایشها روی دسته‌ای از تصاویر متنی نشان می‌دهد که در SNR یکسان، روش پیشنهادی باعث افزایش میزان فشرده‌سازی استاندارد JPEG می‌شود.

طبق دانسته‌ی ما، تاکنون کار چندانی برای استفاده موثر از استاندارد JPEG در فشرده‌سازی تصاویر متنی انجام نشده است.

هدف روش پیشنهادی این است که از روی تصویر متنی ورودی، تصویر متنی دیگری ایجاد شود که منطبق بر رفتار فرکانسی استاندارد JPEG باشد. در روش JPEG به دلیل استفاده از تبدیل DCT ضرایب بزرگ عمدتاً در فرکانسهای پایین انباشته شده و هر چه به سمت فرکانسهای بزرگتر پیش برویم، اندازه ضرایب کوچکتر می‌شود. به این ویژگی، خاصیت متراکم‌سازی انرژی گفته می‌شود. پس از تبدیل DCT، از کوانتیزاسیون ضرایب استفاده می‌شود که باعث صفر شدن ضرایب کوچک می‌شود؛ بنابراین، قسمت عمده ضرایب فرکانسهای میانی و فرکانسهای بالا صفر می‌شوند که این امر باعث

افزایش فشرده‌سازی تصویر ورودی می‌گردد. صفر شدن عمده‌ی فرکانسهای میانی به بالا معادل با اعمال یک فیلتر پایین گذر روی تصویر ورودی است.

مهمترین اطلاعات یک تصویر متنی، بر خلاف تصاویر طبیعی، در لبه‌های تصویر و متناظر با مرز بین کلمات با پس‌زمینه قرار دارند. لبه‌های یک تصویر متنی معادل با فرکانسهای میانی و گاهاً فرکانسهای بالا می‌باشند. رفتار پایین‌گذری استاندارد JPEG این فرکانسهای را حذف و در نتیجه، لبه‌های یک تصویر را هموار می‌کند. بنابراین، عمده‌ی اطلاعات متنی از دست رفته و به عبارت معادل، کیفیت تصویر متنی افت پیدا می‌کند.

در روش پیشنهادی، پس از اعمال تبدیل DCT به تصویر ورودی، فرکانسهای میانی نسبت به فرکانسهای اطراف بیشتر تقویت می‌شوند. برای تعیین میزان تقویت، در این پایان نامه یک روش پیشنهاد شده است. این کار باعث می‌شود تا فرآیند کوانتیزاسیون ضرایب مذکور را حذف نکند و در نتیجه، اطلاعات متنی حفظ شوند. البته اثر دیگر این کار این است که کارایی فشرده‌سازی را تا حدی پایین می‌آورد زیرا تعداد ضرایب صفر کاهش یافته است. برای جبران این اثر منفی، در روش پیشنهادی، ضرایب دوباره چینی می‌شوند تا طول دنباله‌ی ضرایب صفر بزرگتر شده و کارایی کدگذاری دوره تداوم مورد استفاده در استاندارد JPEG بیشتر شود. که در مراحل بعد جهت جابه‌جایی فرکانس استفاده می‌شود.

۴-۳-۲-۱ بررسی اثر تبدیل DCT

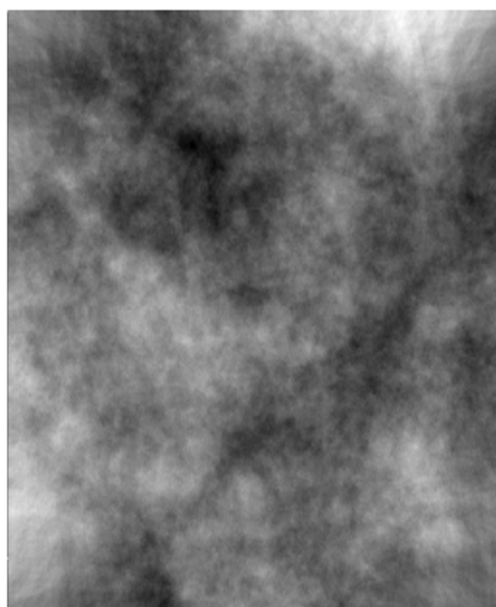
ابتدا از تصویر تبدیل DCT می‌گیریم. سپس نواحی فرکانس بالا را صفر می‌کنیم. تاثیر این کار را روی کیفیت دو نوع از تصاویر یکی متنی و دیگری طبیعی بررسی می‌کنیم. اگر در طیف DCT، کل طیف را در هر یک از راستاهای سطری و ستونی به چهار قسمت تقسیم کنیم، شانزده ناحیه طیفی به دست می‌آید:

۱	۲	۳	۴
۸	۷	۶	۵
۹	۱۰	۱۱	۱۲
۱۶	۱۵	۱۴	۱۳

شکل (۳-۴) محل های قرار گیری ۱۶ طیف تبدیل DCT

در تصاویر متنی فرکانسهای میانی و بالا اهمیت بسیاری در کیفیت متن دارند. برای مثال در دو تصویر

زیر:



In July 1845 Texas formally accepted an American proposal to be annexed to the United States. Already strained relations between the United States and Mexico rapidly worsened. President James K. Polk ordered General Zachary Taylor and his troops to Corpus Christi. In March 1846, under instructions, Taylor took up positions on the Rio Grande. On April 26 an American squadron of dragoons surrounded by Mexicans and unable to fight its way out surrendered. Several Americans lost their lives in the defeat.

On May 8 the Mexicans intercepted Taylor at Palo Alto but were driven back. The next day Mexicans again gave way in battle at Resaca de la Palma. In June Taylor began a march toward Monterrey, taking that city on September 25. Two months later the Americans took Saltillo and with little effort filed into Tampico.

Santa Anna now took the field against the American forces in northern Mexico, finally engaging the Americans at Buena Vista in February. Upon learning of the Mexican's design, General John E. Wool, marched from San Antonio to join Taylor's forces, abandoning his prior plans to take Chihuahua. After extremely bitter fighting Santa Anna pulled out his army, leaving Taylor in control of northern Mexico.

While Taylor pursued the enemy, Colonel Stephen W. Kearny took the "Army of the West" into New Mexico, capturing Santa Fe on August 16, 1846. Kearny then divided his forces, taking part to California and sending the remainder under Alexander W. Doniphan against Chihuahua. After General Winfield Scott captured Mexico City the two countries finally reached a settlement.

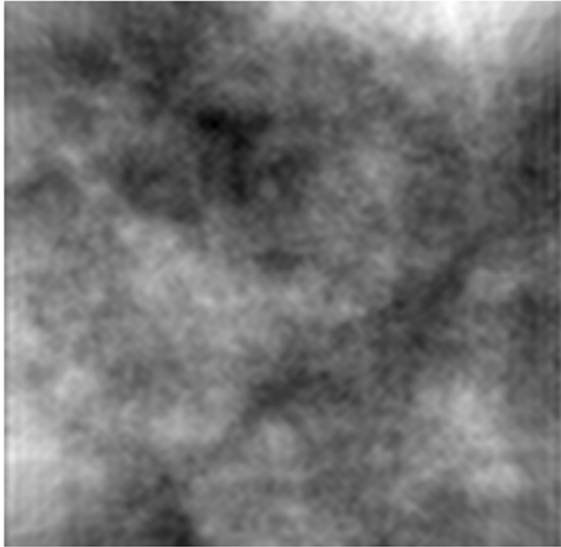
Source: Maps adapted from *The West Point Atlas of American Wars*, Volume 1, Frederick A. Praeger, 1959.

Copyright 1976. Board of Regents, The University of Texas System. All rights reserved.

تصویر دوم: یک تصویر طبیعی (تصویر ابر)

شکل (۴-۴): تصویر اول: یک تصویر متنی

اگر تمام شانزده ناحیه طیفی به جز ناحیه ۱ را در طیف این تصاویر صفر کنیم، اثر آن به صورت زیر خواهد شد:



In July 1845 Texas formally accepted an American proposal to be annexed to the United States. Already strained relations between the United States and Mexico rapidly worsened. President James K. Polk ordered General Zachary Taylor and his troops to Corpus Christi. In March 1845, under instructions, Taylor took up positions on the Rio Grande. On April 25 an American squadron of dragoons surrounded by Mexicans and unable to fight its way out surrendered. Several Americans lost their lives in the defeat. On May 5 the Mexicans intercepted Taylor at Palo Alto but were driven back. The next day Mexicans again gave way in battle at Resaca de la Palma. In June Taylor began a march toward Monterrey, taking that city on September 25. Two months later the Americans took Saltillo and with little effort filed into Tampico. Santa Anna now took the field against the American forces in northern Mexico, finally engaging the Americans at Buena Vista in February. Upon learning of the Mexican's design, General John E. Wool, marched from San Antonio to join Taylor's forces, abandoning his prior plans to take Chihuahua. After extremely bitter fighting Santa Anna pulled out his army, leaving Taylor in control of northern Mexico. While Taylor pursued the enemy, Colonel Stephen W. Kearny took the "Army of the West" into New Mexico, capturing Santa Fe on August 15, 1845. Kearny then divided his forces, taking part to California and sending the remainder under Alexander W. Doniphan against Chihuahua. After General Winfield Scott captured Mexico City the two countries finally reached a settlement.

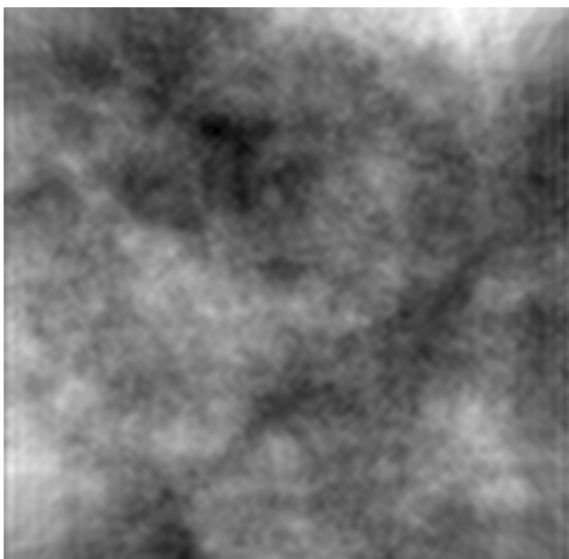
Source: Maps adapted from *The West Point Atlas of American Wars*, Volume 1, Frederick A. Praeger, 1959.

Copyright 1978, Board of Regents, The University of Texas System. All rights reserved.

شکل (۴-۵): اثر حذف نواحی طیفی ۲ تا ۱۶ در تصویر اول

اثر حذف نواحی طیفی ۲ تا ۱۶ در تصویر دوم همانطور که مشاهده می‌شود نگهداری فقط یک ضریب (ضریب DC) باعث می‌شود که گویی هر دو تصویر از یک فیلتر پایین گذر عبور کرده و اثر نرم شدن لبه‌ها به وضوح خود را نشان می‌دهد.

اگر فقط نواحی میان گذر (۲ و ۳ و ۶ و ۷ و ۸ و ۹ و ۱۰ و ۱۱) را حذف کنیم:



In July 1845 Texas formally accepted an American proposal to be annexed to the United States. Already strained relations between the United States and Mexico rapidly worsened. President James K. Polk ordered General Zachary Taylor and his troops to Corpus Christi. In March 1845, under instructions, Taylor took up positions on the Rio Grande. On April 25 an American squadron of dragoons surrounded by Mexicans and unable to fight its way out surrendered. Several Americans lost their lives in the defeat. On May 5 the Mexicans intercepted Taylor at Palo Alto but were driven back. The next day Mexicans again gave way in battle at Resaca de la Palma. In June Taylor began a march toward Monterrey, taking that city on September 25. Two months later the Americans took Saltillo and with little effort filed into Tampico. Santa Anna now took the field against the American forces in northern Mexico, finally engaging the Americans at Buena Vista in February. Upon learning of the Mexican's design, General John E. Wool, marched from San Antonio to join Taylor's forces, abandoning his prior plans to take Chihuahua. After extremely bitter fighting Santa Anna pulled out his army, leaving Taylor in control of northern Mexico. While Taylor pursued the enemy, Colonel Stephen W. Kearny took the "Army of the West" into New Mexico, capturing Santa Fe on August 15, 1845. Kearny then divided his forces, taking part to California and sending the remainder under Alexander W. Doniphan against Chihuahua. After General Winfield Scott captured Mexico City the two countries finally reached a settlement.

Source: Maps adapted from *The West Point Atlas of American Wars*, Volume 1, Frederick A. Praeger, 1959.

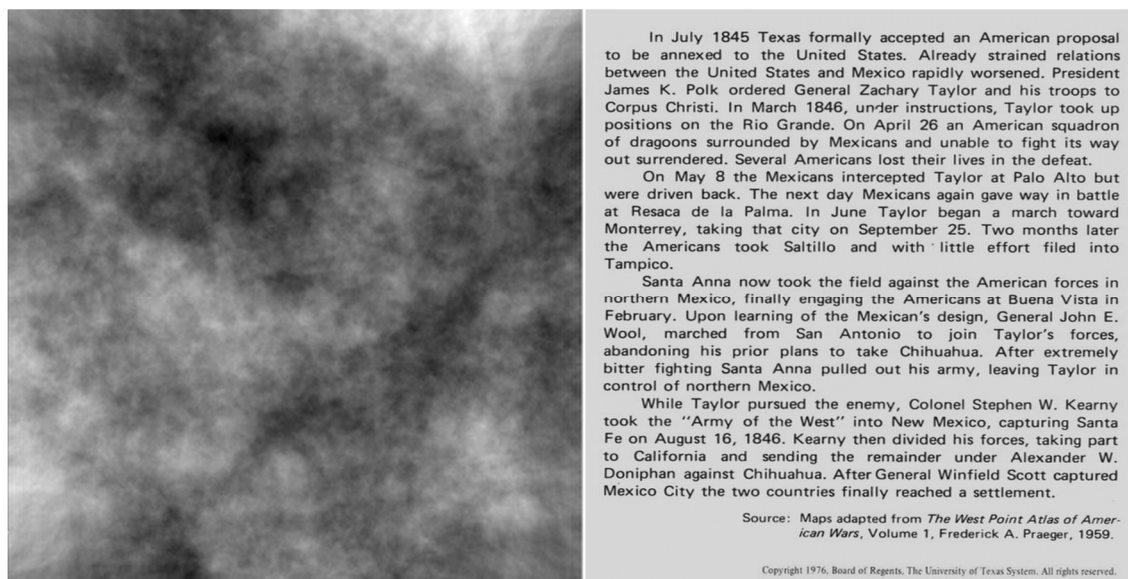
Copyright 1978, Board of Regents, The University of Texas System. All rights reserved.

تصویر دوم مربوط به حذف فرکانسهای میانی

شکل (۴-۶): تصویر اول مربوط به حذف فرکانسهای میانی

در این حالت تصویر متن همانند حالت قبل کیفیت دیداری خود را به وضوح از دست داده ولی تصویر طبیعی آسیب دیداری کمتری را تجربه می‌کند.

اگر فقط فرکانسهای بالا یعنی نواحی ۴ و ۵ و ۱۲ و ۱۳ و ۱۴ و ۱۵ و ۱۶ را صفر کنیم:

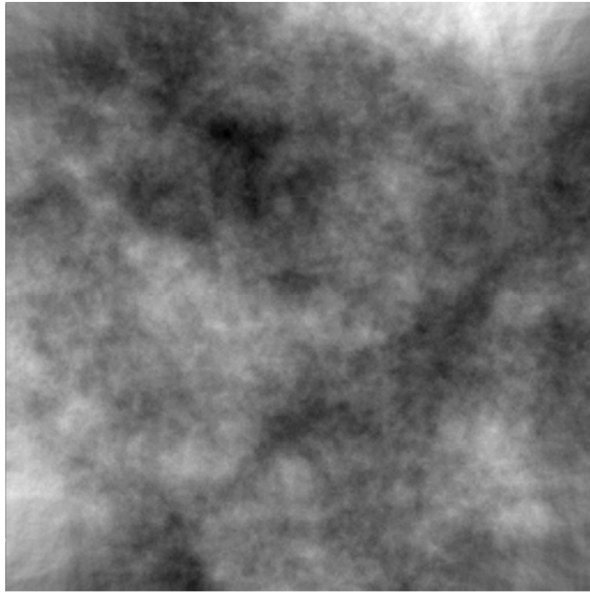


تصویر دوم مربوط به حذف فرکانسهای بالا

شکل (۴-۷): تصویر اول مربوط به حذف فرکانسهای بالا

مشاهده می‌شود که با باقی ماندن فرکانسهای میانی و پایین وضوح تصویر متنی به خوبی حفظ شده است و از این تجربه می‌توان نتیجه گرفت که فرکانسهای مهم در یک تصویر متنی فرکانسهای میانی و پایین می‌باشد و در فرایند فشرده سازی با اتلاف از دست دادن فرکانسهای بالا تاثیر چندانی در خوش‌آیندی متن از لحاظ دیداری نخواهد داشت.

در بین نواحی فرکانس بالای ذکر شده اگر بخواهیم تاثیر نواحی ۱۲ و ۱۳ و ۱۴ را جداگانه بررسی کنیم:



In July 1845 Texas formally accepted an American proposal to be annexed to the United States. Already strained relations between the United States and Mexico rapidly worsened. President James K. Polk ordered General Zachary Taylor and his troops to Corpus Christi. In March 1846, under instructions, Taylor took up positions on the Rio Grande. On April 26 an American squadron of dragoons surrounded by Mexicans and unable to fight its way out surrendered. Several Americans lost their lives in the defeat.

On May 8 the Mexicans intercepted Taylor at Palo Alto but were driven back. The next day Mexicans again gave way in battle at Resaca de la Palma. In June Taylor began a march toward Monterrey, taking that city on September 25. Two months later the Americans took Saltillo and with little effort filed into Tampico.

Santa Anna now took the field against the American forces in northern Mexico, finally engaging the Americans at Buena Vista in February. Upon learning of the Mexican's design, General John E. Wool, marched from San Antonio to join Taylor's forces, abandoning his prior plans to take Chihuahua. After extremely bitter fighting Santa Anna pulled out his army, leaving Taylor in control of northern Mexico.

While Taylor pursued the enemy, Colonel Stephen W. Kearny took the "Army of the West" into New Mexico, capturing Santa Fe on August 16, 1846. Kearny then divided his forces, taking part to California and sending the remainder under Alexander W. Doniphan against Chihuahua. After General Winfield Scott captured Mexico City the two countries finally reached a settlement.

Source: Maps adapted from *The West Point Atlas of American Wars*, Volume 1, Frederick A. Praeger, 1959.

Copyright 1976. Board of Regents, The University of Texas System. All rights reserved.

شکل (۴-۸): تصویر اول مربوط به صفرکردن نواحی ۱۲ و ۱۳ و ۱۴ تصویر دوم مربوط به صفرکردن نواحی ۱۲ و ۱۳ و ۱۴

مشاهده می شود حذف این فرکانس ها تاثیری در هیچ یک از تصاویر ندارد.

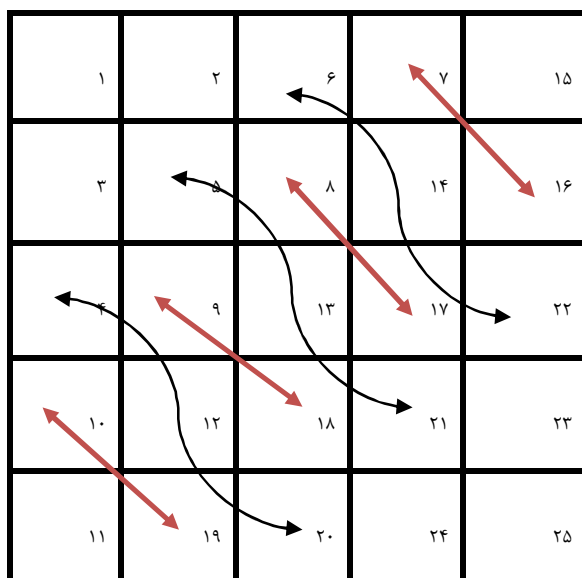
نحوه جابجایی فرکانس های تبدیل DCT :

با حذف نواحی مختلف فرکانسی، پس از بررسی های لازم به صورت شهودی و با استفاده از روشی که

در زیر توضیح داده خواهد شد (۴-۴-۱)، نواحی موثر و تاثیر گذار بر روی تصاویر متنی به دست آمد. با

توجه به جدول کوانتیزاسیون تبدیل JPEG، با انتقال این نواحی به محل هایی از جدول که دچار

کمترین تضعیف می شدند، جدول ضرایب به صورت زیر اصلاح شد:



شکل (۴-۹) جابجایی محل فرکانس های تصویر متنی پس از تبدیل DCT

نحوه وزن دهی ضرایب تبدیل DCT :

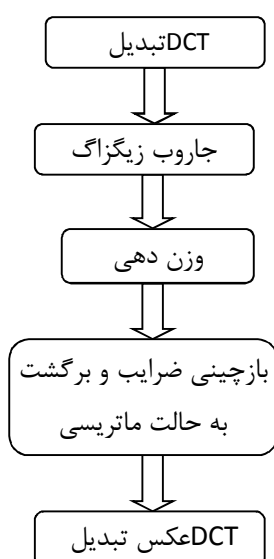
پس از جابجایی فرکانسی آن دسته از طیف های تبدیل DCT که دارای اهمیت بیشتری بودند، طبق جدول زیر اقدام به تقویت این دسته از ضرایب می کنیم. در جدول زیر محل و مقدار تقویت هر طیف نشان داده شده است.

1	1	10	10	5
1	10	10	5	2
10	10	5	2	2
10	5	2	2	1
5	2	2	1	1

شکل (۴-۱۰) میزان تقویت هر یک از بازه های فرکانسی

۴-۴ روش پیش پردازش پیشنهادی

در این بخش روش پیشنهادی توضیح داده می‌شود. بلوک دیاگرام روش پیشنهادی در شکل ۴-۱۱ نشان داده شده است.



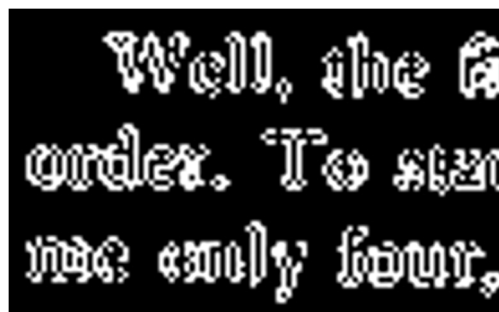
شکل (۴-۱۱) بلوک دیاگرام روش پیشنهادی

در مرحله اول، مشابه با روش JPEG، تصویر به بلوکهای 8×8 تقسیم و از هر بلوک تبدیل DCT گرفته می‌شود. سپس ضرایب ماتریس تبدیل به کمک جاروب زیگزاگ به دنباله یا برداری یک بعدی به نام دنباله‌ی فرکانسی تبدیل می‌شود. در ادامه، ضرایب وزن دهی می‌شوند تا ضرایب فرکانسهای میانی بیشتر از بقیه تقویت شوند. وزن ضرایب از قبل تعیین شده و ثابت می‌باشند. در ادامه، در اولین زیربخش نحوه تعیین وزن ضرایب بیان شده است. سپس ضرایب تقویت شده، به صورت گروهی دسته بندی و سپس به طور نزولی بازچینی می‌شوند تا در نتیجه‌ی این کار و پس از کوانتیزاسیون انجام شده در JPEG دنباله‌ی بزرگتری از صفرها تولید شود. در دومین زیربخش، روش پیشنهادی برای بازچینی ضرایب توضیح داده می‌شود. در پایان دنباله‌ی ضرایب تقویت شده و بازچینی شده با توجه به ترتیب جاروب زیگزاگ اولیه، به حالت ماتریسی برگردانده شده و عکس تبدیل DCT روی آن اعمال

می‌گردد. به این ترتیب در پایان، تصویری خواهیم داشت که مناسب برای اعمال فشرده‌سازی JPEG بوده و کارایی فشرده‌سازی آن را افزایش می‌دهد.

۴-۴-۱ نحوه‌ی تعیین وزن ضرایب تبدیل

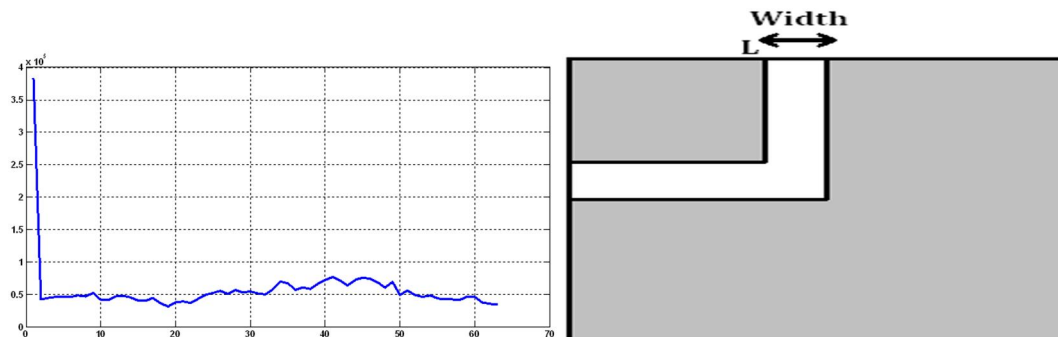
برای تعیین وزن هر ضریب، میزان اهمیت آن در بازسازی لبه‌ها را تعیین می‌کنیم. برای تعیین نواحی لبه‌ها، گرادیان تصویر ورودی را محاسبه کرده و آن را بین ۰ تا ۱ نرمالیزه می‌کنیم. سپس مقادیری از گرادیان که کمتر از $Th1$ باشند را برابر صفر و بقیه مقادیر را برابر ۱ قرار می‌دهیم. به این ترتیب یک تصویر دودویی ایجاد می‌شود که بیانگر نواحی لبه‌ها است. نمونه‌ای از چنین تصویری در شکل ۴-۱۲ نشان داده شده است. مقدار آستانه‌ی $Th1$ تعیین کننده عرض نواحی ایجاد شده است. یعنی هر چه میزان $Th1$ بزرگتر باشد، لبه‌ها نازکتر و هر چه $Th1$ کوچکتر باشد نواحی لبه ضخامت بیشتری خواهند داشت.



شکل (۴-۱۲): نمونه‌ای از تصویر لبه‌ها

سپس از تصویر ورودی تبدیل DCT گرفته و فرآیندی به نام فرآیند صفر کردن نوار را از ضرایب پایین تا بالا تکرار می‌کنیم. در این فرآیند، مطابق شکل ۴-۱۳ ضرایب متعلق به نواری به عرض Width که از ضریب L ام شروع شده باشد را برابر صفر قرار داده و عکس تبدیل DCT را محاسبه می‌کنیم. سپس تفاضل تصویر اصلی با تصویر بازسازی شده را محاسبه و انرژی آن نواحی که متناظر با نواحی لبه در تصویر دودویی حاصل از گرادیان باشند را محاسبه می‌کنیم. به این ترتیب با تغییر مقدار L می‌توان میزان اهمیت هر یک از بازه‌های فرکانسی (به عرض Width) را در بازسازی لبه‌های تصویر متنی

تعیین کرد. یک نمونه نمودار مربوط به انرژی نواحی لبه بر حسب پارامتر L در حالت $Width=10$ در شکل ۴-۱۳ الف نشان داده شده است.



شکل (۴-۱۳): (الف) نواحی نواری شکل در ماتریس تبدیل DCT (ب) یک نمونه نمودار انرژی نواحی لبه بر حسب پارامتر L

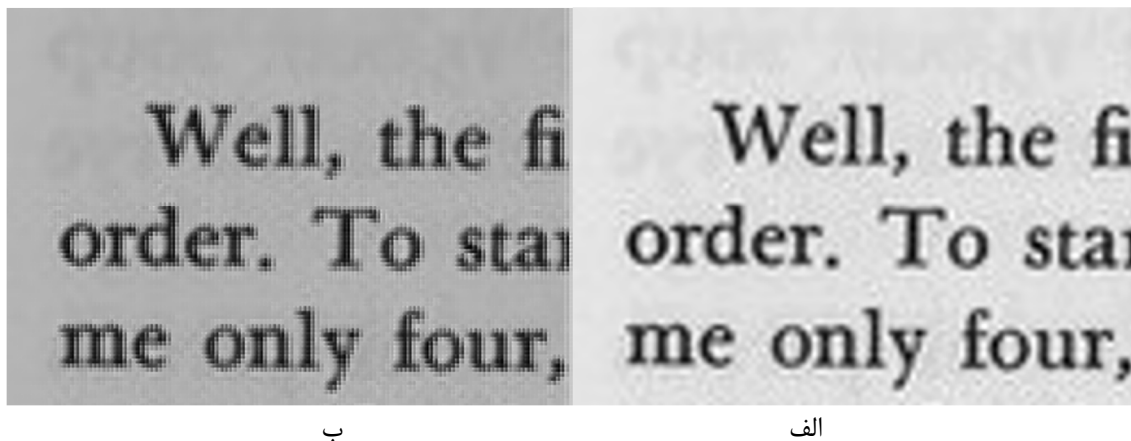
مطابق شکل ۴-۱۳ ب ملاحظه می شود که فرکانسهای میانی نسبت به فرکانسهای بالا تاثیر بیشتری در بازسازی و حفظ کیفیت لبهها دارند. با توجه به نمودار فوق می توان برای هر ضریب تبدیل DCT وزنی بین ۱ تا ۱۰ تعیین و استفاده کرد.

۴-۴-۲ بازچینی ضرایب تبدیل

پس از انجام مرحله ی وزن دهی دنباله ی فرکانسی، این دنباله را بازچینی می کنیم. برای این کار کل دنباله فرکانسی را به N زیربازه تقسیم کرده و به هر یک شماره ی ۱ تا N نسبت می دهیم. سپس برای هر زیربازه میانگین قدر مطلق ضرایب مربوطه را محاسبه می کنیم. در ادامه، دنباله ی مقادیر میانگین حاصل شده را به ترتیب نزولی مرتب می کنیم تا ترتیب جدیدی از شمارهها به وجود آید. حال، زیربازه های فرکانسی را به همین ترتیب جدید به دست آمده کنار هم قرار می دهیم تا دنباله ی فرکانسی جدید و مرتب شده ای به دست آوریم. قابل ذکر است که دنباله ی شماره های مرتب شده باید در کنار داده های تصویر متنی پیش پردازش شده ذخیره گردند تا در حین بازسازی تصویر متنی نهایی مورد استفاده قرار بگیرند. سربار این دنباله چندان قابل توجه نمی باشد زیرا این دنباله شامل N عدد می باشد که معمولاً $N=8$ انتخاب می شود.

هدف از بازچینی ضرایب تبدیل همان طور که قبلاً نیز بیان شد، کمک به تولید دنباله‌های صفر بزرگتر در حین فرآیند فشرده‌سازی JPEG پس از مرحله‌ی کوانتیزاسیون و در نتیجه، افزایش کارایی فشرده‌سازی کدگذاری طول دوره تداوم است.

در پایان روش پیش‌پردازش پیشنهادی در شکل ۴-۱۱، برای هر تصویر متنی ورودی، یک تصویر متنی حاصل می‌شود که آماده برای اعمال فرآیند فشرده‌سازی JPEG است. شکل ۴-۱۴ نمونه‌ای از چنین تصویر پردازش شده‌ای را نشان می‌دهد. همان طور که ملاحظه می‌شود، روش پیش‌پردازش پیشنهادی باعث نرم‌تر کردن لبه‌های متنی می‌شود. اما اگر همین تصویر با روش JPEG فشرده‌سازی و بازسازی شود و سپس ضرایب تبدیل DCT به محل اولیه خود (قبل از مرحله بازچینی) منتقل شوند، تصویر متنی نهایی، کیفیت بهتری نسبت به اعمال مستقیم روش JPEG خواهد داشت.



شکل (۴-۱۴): (الف) تصویر اصلی، (ب) نتیجه اعمال پیش‌پردازش پیشنهادی روی تصویر (الف)

در قسمت نتایج شبیه‌سازی، حالت استفاده از روش پیش‌پردازش پیشنهادی قبل از اعمال روش JPEG با حالت اعمال مستقیم روش JPEG مقایسه می‌گردد.

روش پیش‌پردازش پیشنهادی باعث نرم‌تر کردن لبه‌های متنی می‌شود. اما اگر تصویر با روش JPEG فشرده‌سازی و بازسازی شود و سپس ضرایب تبدیل DCT به محل اولیه خود (قبل از مرحله بازچینی) منتقل شوند، تصویر متنی نهایی، کیفیت بهتری نسبت به اعمال مستقیم روش JPEG خواهد داشت.

۴-۵ جمع بندی و نتیجه گیری

در این فصل به بیان روش های استفاده شده در این پایان نامه پرداخته شد، که در آن یک روش جهت دودویی سازی تصاویر متنی و دو روش جهت فشرده سازی تصاویر سطح خاکستری متنی، ارائه شد که این روشها در قالب مدل محتوای ترکیبی استفاده می شوند. در روش اول، از تبدیل KLT استفاده شد و در روش دوم از یک پیش پردازش روی تصاویر متنی جهت استفاده از استاندارد JPEG برای فشرده سازی تصاویر متنی استفاده شد. در فصل بعد، به بررسی نتایج حاصل شده از روش های پیشنهادی می پردازیم و با بعضی از روش های مرسوم مقایسه می کنیم.

فصل پنجم

پایگاه داده و نتایج آزمایش ها

۵-۱ پایگاه داده

نظر به پیشرفت روز افزون علم پردازش تصویر و وجود وجوه تخصصی آن در بررسی خطوط فارسی، از قبیل استخراج متن، شناسایی نوری حروف (OCR) و فشرده‌سازی تصاویر متنی، نیاز به یک پایگاه داده تخصصی از متون فارسی احساس می‌شود. یافت نشدن پایگاه جامع که به راحتی و رایگان در اختیار بوده و بتواند پاسخ گوی این نیاز روز افزون باشد بسیار محسوس بود. لذا در راستای این پایان‌نامه و برطرف کردن این کاستی اقدام به ایجاد یک پایگاه داده متنی شده و سعی بر این بوده که انواع مختلفی از تصاویر متنی را در بر بگیرد. در زیر به توصیف این پایگاه می پردازیم.

۵-۱-۱ سخت افزار استفاده شده

سخت افزارهایی که عمل روبش (اسکن کردن) توسط آنها صورت گرفت شامل دو اسکنر AV210D2+ شرکت AVISION و F2400 شرکت HP است و به جزاسناد چاپی موجود، چهار نوع چاپ گر جهت ایجاد اسناد کاغذی استفاده شد. دو چاپگر سوزنی به نام های LQ300 و LQ100 با علامت تجاری EPSON و چاپ گر های لیزری w1102 و F2400 با علامت تجاری HP مورد استفاده قرار گرفت.

۵-۱-۲ دسته بندی تصاویر

بیشتر این تصاویر به صورت رنگی روبش شده‌اند و تعداد محدودی هم به صورت دودویی (سیاه و سفید) جهت بعضی از مصارف خاص مورد روبش قرار گرفته است که تصاویر رنگی این دسته از تصاویر سیاه و سفید نیز موجود می باشد. از هر تصویر چهار درجه تفکیک ۱۰۰dpi و ۲۰۰dpi و ۳۰۰dpi و ۶۰۰dpi تهیه شده است. دسته بندی این تصاویر به شرح می باشد.

در کلی ترین حالت تصاویر به ۳ دسته دست نویس، چاپی و ترکیب دست نویس و چاپی تقسیم شده اند، هر دسته شامل زبان های خاصی می باشد که آنها نیز طبق شرایط زیر دسته بندی می شوند:

۱-عمدتاً فارسی

۲-عمدتاً عربی

۳-عمدتاً انگلیسی

۴-ترکیب فارسی و عربی

۵-ترکیب فارسی و انگلیسی

۶-ترکیب عربی و فارسی

هر دسته فوق الذکر نیز به سه دسته

۱-عمدتاً متن

۲-عمدتاً گرافیک

۳-متن و گرافیک

تقسیم شده اند. لازم به ذکر است منظور از گرافیک در متن فوق الذکر وجود جداول و منحنی ها و عکس و تصاویر طبیعت و... می باشد. این تصاویر همچنین شامل چیدمان هایی به صورت متن و گرافیک جدا و متن و گرافیک همپوشان هستند.

۵-۱-۳ درصد فراوانی انواع تصاویر در پایگاه داده

در این پایگاه سعی بر این بوده که تصاویر بدون هر گونه نقصی و در بهترین شرایط روبش شود، ولی در تعداد محدودی از تصاویر جهت آن دسته از افراد که قصد دارند بر روی تصاویر آسیب دیده، پردازش داشته باشند، مانند خمش در محل شیرازه و کجی کل تصویر (چرخش) وجود تصاویر کم رنگ و نویزدار و تصاویری که در حاشیه آنها، دارای نواقصی هستند نیز موجود است.

تعداد کل تصاویر موجود در این پایگاه ۱۴۴۹ می باشد. نکته دیگری که در این پایگاه داده مورد توجه قرار گرفته، استفاده از قلم‌های مختلف و دارای اندازه‌های متنوع می باشد. همچنین تراکم متن موجود در هر صفحه، از حالتی که کل صفحه توسط متن پوشیده شده تا حالتی که تنها چند کلمه محدود در یک صفحه قرار دارند، متغیر می باشد؛ بنابراین، این کار تنوع موجود در این پایگاه داده را به نوعی دیگر افزایش داده است.

در میان تصاویر موجود در این پایگاه داده، اکثریت تصاویر دارای متن های سطح خاکستری می باشد ولی تعداد معدودی تصویر دارای متن رنگی نیز موجود است. سعی بر این بوده که این پایگاه داده با تاکید بر متن فارسی تشکیل شود ولی جهت ایجاد یک پایگاه داده جامع، از تصاویر متنی عربی و انگلیسی هم استفاده شده است. نسبت هر یک از تصاویر به کل تصویر موجود به شرح زیر گزارش می شود:

۱- چاپی یا دست نویس:

الف: تصاویر دست نویس ۱۳,۷۳٪

ب: تصاویر چاپی ۸۲,۳۳٪

ج: ترکیب چاپی و دست نویس ۳,۹۴٪

۲- زبان متن:

عمدتا فارسی ۵۵,۳٪

عمدتا عربی ۵,۳۸٪

عمدتا انگلیسی ۱۷,۶۴٪

فارسی و عربی ۴,۱۱٪

فارسی و انگلیسی ۱۵,۴۲٪

عربی و انگلیسی ۲,۱۵٪

۳- نوع محتویات:

عمدتا متن ۵۴,۰۱٪

عمدتا گرافیک ۱۱,۸۲٪

متن و گرافیک ۳۴,۱۷٪

۴- نوع ساختار یا چیدمان:

متن جدا، گرافیک جدا ۷۰,۵۶٪

متن و گرافیک همپوشان ۲۹,۴۴٪

در زیر نمونه هایی از این تصاویر را ملاحظه می فرمایید:

تیمه چای گاهی تزه مردم دارد؟

این نمودارها اطلاعات ارزشمندی را در مورد وضعیت چای در ایران ارائه می‌دهند. از جمله:

- تولید چای در ایران: ۳۳۰ هزار تن
- صادرات چای: ۳۰۰ هزار تن
- ارزش صادرات چای: ۱۵۰ میلیون دلار
- کشورهای اصلی واردکننده: چین، روسیه، هند، عراق، ترکیه، پاکستان، هندوستان، قطر، امارات متحده امارات.
- پایتخت چای ایران: استان فارس (۳۵٪ تولید)
- سایر استان‌های تولید کننده: کهگیلویه و بویراحید، بوشهر، هرمزگان، سیستان و بلوچستان، خراسان جنوبی، اصفهان، آذربایجان غربی، کردستان، گلستان، مازندران، زنجان، گیلان، قزوین، البرز، تهران، چابکمه، کرمان، فارس، سیستان و بلوچستان، کهگیلویه و بویراحید، بوشهر، هرمزگان، خراسان جنوبی، اصفهان، آذربایجان غربی، کردستان، گلستان، مازندران، زنجان، گیلان، قزوین، البرز، تهران، چابکمه، کرمان.

لیست تکنیک الیز

ردیف	تکنیک	تاریخ
1	تکنیک ۱	۱۳۸۵
2	تکنیک ۲	۱۳۸۵
3	تکنیک ۳	۱۳۸۵
4	تکنیک ۴	۱۳۸۵
5	تکنیک ۵	۱۳۸۵
6	تکنیک ۶	۱۳۸۵
7	تکنیک ۷	۱۳۸۵
8	تکنیک ۸	۱۳۸۵
9	تکنیک ۹	۱۳۸۵
10	تکنیک ۱۰	۱۳۸۵
11	تکنیک ۱۱	۱۳۸۵
12	تکنیک ۱۲	۱۳۸۵
13	تکنیک ۱۳	۱۳۸۵
14	تکنیک ۱۴	۱۳۸۵
15	تکنیک ۱۵	۱۳۸۵
16	تکنیک ۱۶	۱۳۸۵
17	تکنیک ۱۷	۱۳۸۵
18	تکنیک ۱۸	۱۳۸۵
19	تکنیک ۱۹	۱۳۸۵
20	تکنیک ۲۰	۱۳۸۵
21	تکنیک ۲۱	۱۳۸۵
22	تکنیک ۲۲	۱۳۸۵
23	تکنیک ۲۳	۱۳۸۵
24	تکنیک ۲۴	۱۳۸۵
25	تکنیک ۲۵	۱۳۸۵
26	تکنیک ۲۶	۱۳۸۵
27	تکنیک ۲۷	۱۳۸۵
28	تکنیک ۲۸	۱۳۸۵
29	تکنیک ۲۹	۱۳۸۵
30	تکنیک ۳۰	۱۳۸۵

جمع به عروفت: ۱۳۸۵

تاریخ: ۱۳۸۵

معادله ۱

$$m^2 + (a+b+c)x + d = 0$$

محل‌های این معادله عبارتند از: $x_1 = \frac{-a-b-c + \sqrt{a^2+b^2+c^2+d}}{2}$ و $x_2 = \frac{-a-b-c - \sqrt{a^2+b^2+c^2+d}}{2}$

معادله ۲

$$m^2 + (a+b+c)x + d = 0$$

محل‌های این معادله عبارتند از: $x_1 = \frac{-a-b-c + \sqrt{a^2+b^2+c^2+d}}{2}$ و $x_2 = \frac{-a-b-c - \sqrt{a^2+b^2+c^2+d}}{2}$

جدول اول

ردیف	نام خانوادگی	تاریخ تولد	تاریخ فوت
1	محمد علی	۱۳۸۵	۱۳۸۵
2	علیرضا	۱۳۸۵	۱۳۸۵
3	فریبا	۱۳۸۵	۱۳۸۵
4	سید علی	۱۳۸۵	۱۳۸۵
5	رضا	۱۳۸۵	۱۳۸۵
6	مهدی	۱۳۸۵	۱۳۸۵
7	سید محمد	۱۳۸۵	۱۳۸۵
8	فرهاد	۱۳۸۵	۱۳۸۵
9	علیرضا	۱۳۸۵	۱۳۸۵
10	فریبا	۱۳۸۵	۱۳۸۵
11	سید علی	۱۳۸۵	۱۳۸۵
12	رضا	۱۳۸۵	۱۳۸۵
13	مهدی	۱۳۸۵	۱۳۸۵
14	سید محمد	۱۳۸۵	۱۳۸۵
15	فرهاد	۱۳۸۵	۱۳۸۵
16	علیرضا	۱۳۸۵	۱۳۸۵
17	فریبا	۱۳۸۵	۱۳۸۵
18	سید علی	۱۳۸۵	۱۳۸۵
19	رضا	۱۳۸۵	۱۳۸۵
20	مهدی	۱۳۸۵	۱۳۸۵
21	سید محمد	۱۳۸۵	۱۳۸۵
22	فرهاد	۱۳۸۵	۱۳۸۵
23	علیرضا	۱۳۸۵	۱۳۸۵
24	فریبا	۱۳۸۵	۱۳۸۵
25	سید علی	۱۳۸۵	۱۳۸۵
26	رضا	۱۳۸۵	۱۳۸۵
27	مهدی	۱۳۸۵	۱۳۸۵
28	سید محمد	۱۳۸۵	۱۳۸۵
29	فرهاد	۱۳۸۵	۱۳۸۵
30	علیرضا	۱۳۸۵	۱۳۸۵

انرژی خورشیدی

این سیستم به شما کمک می‌کند تا انرژی خورشیدی را به راحتی و به صرفه به کار ببرید. ویژگی‌های آن عبارتند از:

- صرفه‌جویی در هزینه‌های انرژی
- عدم آلودگی و صدا
- مدت طولانی عمر

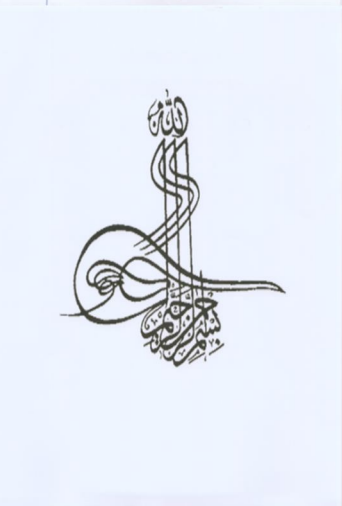
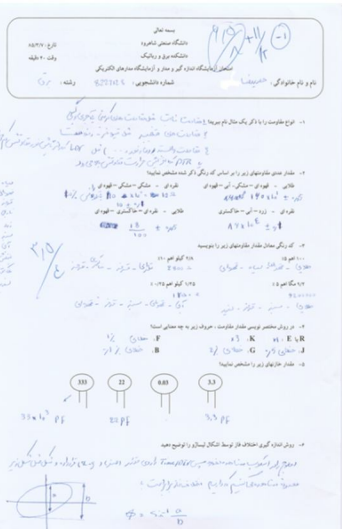
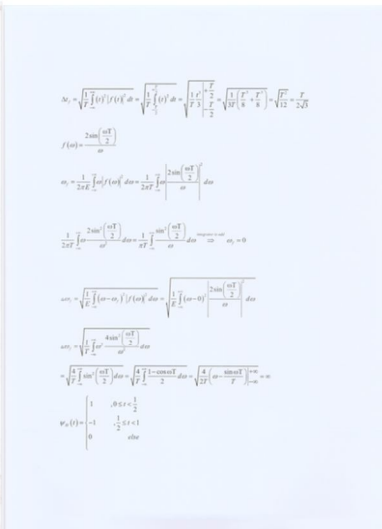
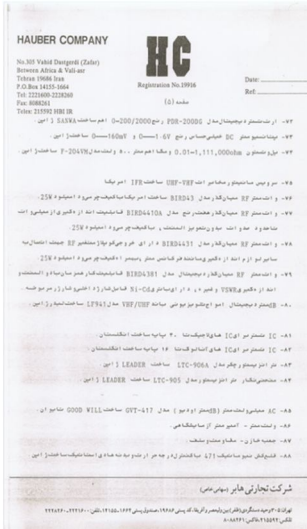
The Flight 8086 Training System

این سیستم آموزشی برای تیم‌های عملیاتی و پشتیبانی در صنایع نفت و گاز طراحی شده است. ویژگی‌های کلیدی آن عبارتند از:

- تیم‌های عملیاتی و پشتیبانی
- آموزش تخصصی در زمینه تعمیرات و نگهداری
- بهره‌برداری از تجهیزات پیشرفته

نظر مردم برای خیلی اهمیت دارد

این سیستم با نظرسنجی‌های مداوم، نظرات مشتریان را جمع‌آوری و تحلیل می‌کند تا بتواند خدمات بهتری ارائه دهد.



شکل (۵-۱) نمونه‌هایی از تصاویر موجود در پایگاه داده

۲-۵ نتایج مربوط به دودویی سازی تصاویر متنی

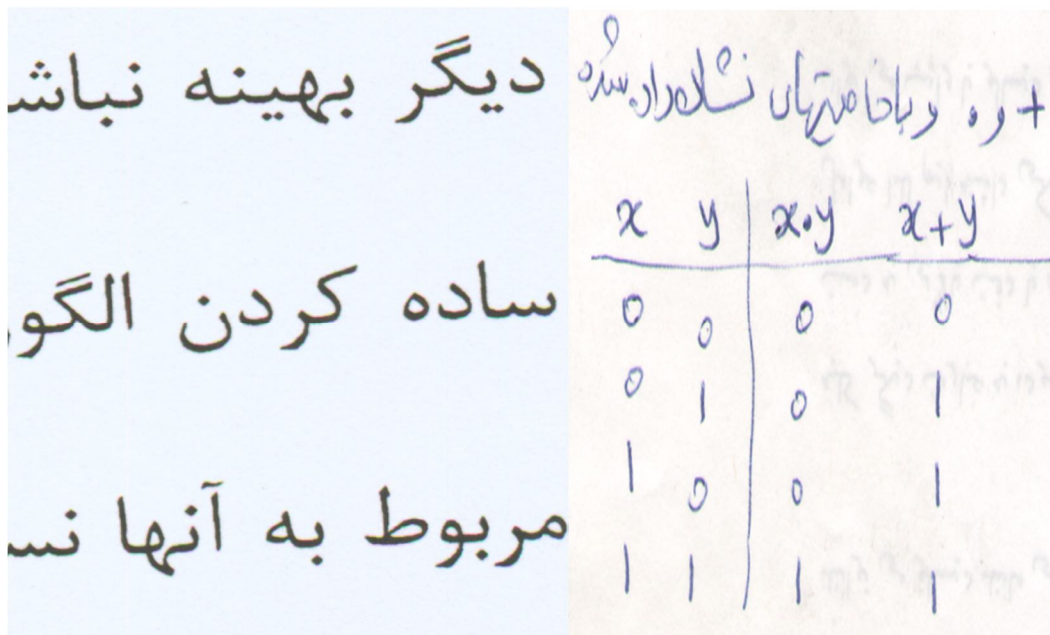
در شبیه‌سازی روش پیشنهادی در فصل قبل، از پایگاه تصاویر متنی سطح خاکستری موجود در پایگاه داده فوق‌الذکر استفاده شده است. اندکی بیش از ۳۰ درصد از این تصاویر برای آموزش شبکه‌های عصبی به کار رفته، در روش پیشنهادی، مورد استفاده قرار گرفتند.

مقادیر پارامترهای روش پیشنهادی در درجه تفکیک مکانی ۳۰۰ نقطه بر اینچ در جدول (۱) آورده شده است. مقدار پارامتر N با درجه تفکیک مکانی رابطه تقریباً خطی دارد.

جدول (۱-۵): مقادیر پارامترهای مورد استفاده در روش دودویی‌سازی پیشنهادی

نام پارامتر	مقدار
N	۳۲
M	۳

دو نمونه تصویر متنی و نتیجه‌ی دودویی‌سازی آن با روش پیشنهادی در شکل‌های به ترتیب ۲-۵ و ۳-۵ نشان داده شده است. شکل ۴-۵ نیز نتیجه‌ی دودویی‌سازی این تصویر را به کمک روش اتسو نشان می‌دهد. با مقایسه‌ی شکل‌های (۳) و (۴) برتری روش پیشنهادی به ویژه در دودویی‌سازی اجزای متنی ظریف مانند ترسیمات خطی مشاهده می‌شود.



شکل (۲-۵): یک نمونه تصویر متنی اصلی در درجه‌ی تفکیک مکانی ۳۰۰ نقطه بر اینچ

+ و ریاضیات را نشان داده دیگر بهینه نباشد

ساده کردن الگوریتم

مربوط به آنها نیست

x	y	$x \cdot y$	$x + y$
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	1

شکل (۳-۵): نتیجه‌ی دودویی‌سازی تصویر متنی شکل (۲-۵) به کمک روش پیشنهادی

+ و ریاضیات را نشان داده دیگر بهینه نباشد

ساده کردن الگوریتم

مربوط به آنها نیست

x	y	$x \cdot y$	$x + y$
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	1

شکل (۴-۵): نتیجه‌ی دودویی‌سازی تصویر متنی شکل (۲-۵) به کمک روش اُتسو

روش پیشنهادی در نرم‌افزار MATLAB 2010b روی سیستمی با ۱ گیگابایت حافظه‌ی RAM

شبیه‌سازی گردید. تحت این شرایط، مدت زمان اجرای روش پیشنهادی اول به طور متوسط، کمتر از

نصف مدت زمان مربوط به روش اُتسو و در حالت کلی اندکی بیش از این زمان به دست آمد. میتوان نشان داد که پیچیدگی محاسباتی روش پیشنهادی از مرتبه N^2 است، که N اندازه بلوکهای پردازش هر پیکسل است. همچنین، طبق معیار متوسط امتیاز نظر سنجی (MOS) و با کسب نظر حدود ۵۰ مشاهده‌گر، مقدار متوسط این معیار برای روش پیشنهادی برابر با ۴,۲ و برای روش اُتسو برابر با ۳,۸ به دست آمد که نشان‌دهنده‌ی بالاتر بودن کیفیت تصویر دودویی شده با روش پیشنهادی نسبت به روش استاندارد اُتسو است. مهمترین علت برتری روش پیشنهادی نسبت به روش اُتسو، این است که روش پیشنهادی علاوه بر برخوردار بودن از قدرت تعیین افقی آستانه (حتی بهتر از روش اُتسو)، در مواقعی که روش آستانه افقی ضعیف عمل می‌کند با کمک طبقه بند مبتنی بر شبکه عصبی، مقدار آستانه به طور مناسب انتخاب می‌شود.

۳-۵ نتایج روش پیشنهادی اول مربوط به فشرده سازی تصاویر متنی

با استفاده از تبدیل KLT

برای شبیه سازی روش پیشنهادی از پایگاه تصاویر متنی فوق الذکر با تاکید بر تصویر عمدتاً متنی با کیفیت معمولی و درجه تفکیک مکانی ۳۰۰ نقطه بر اینچ استفاده شده است، که از جاروب برخی اسناد فارسی و انگلیسی تشکیل شده است. برخی از تصاویر موجود در این پایگاه در شکل ۵-۵ نشان داده شده است. قسمت عمده این تصاویر شامل متن است. برخی از تصاویر مانند شکل (۵-۵-ج) و (۵-۵-د) ممکن است شامل قسمتهای گرافیکی و ترسیمات خطی نیز باشند.

In July 1845 Texas formally accepted an American proposal to be annexed to the United States. Already strained relations between the United States and Mexico rapidly worsened. President James K. Polk ordered General Zachary Taylor and his troops to Corpus Christi. In March 1846, under instructions, Taylor took up positions on the Rio Grande. On April 26 an American squadron of dragoons surrounded by Mexicans and unable to fight its way out surrendered. Several Americans lost their lives in the defeat.

On May 8 the Mexicans intercepted Taylor at Palo Alto but were driven back. The next day Mexicans again gave way in battle at Resaca de la Palma. In June Taylor began a march toward Monterrey, taking that city on September 25. Two months later the Americans took Saltillo and with little effort filed into Tampico.

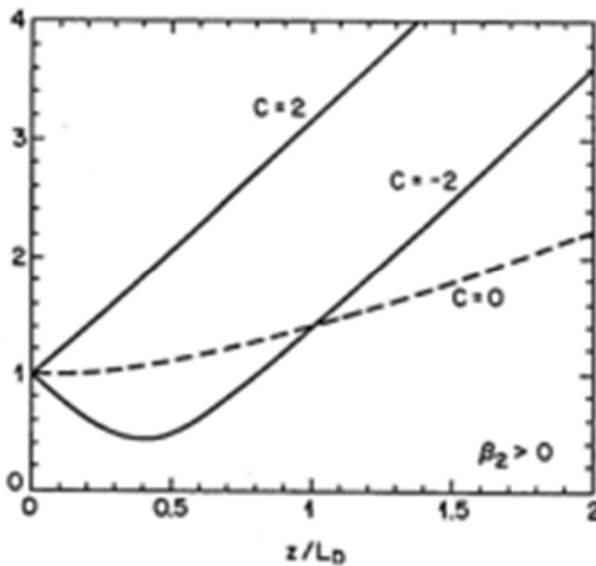
Santa Anna now took the field against the American forces in northern Mexico, finally engaging the Americans at Buena Vista in February. Upon learning of the Mexican's design, General John E. Wool, marched from San Antonio to join Taylor's forces, abandoning his prior plans to take Chihuahua. After extremely bitter fighting Santa Anna pulled out his army, leaving Taylor in control of northern Mexico.

While Taylor pursued the enemy, Colonel Stephen W. Kearny took the "Army of the West" into New Mexico, capturing Santa Fe on August 16, 1846. Kearny then divided his forces, taking part to California and sending the remainder under Alexander W. Doniphan against Chihuahua. After General Winfield Scott captured Mexico City the two countries finally reached a settlement.

Source: Maps adapted from *The West Point Atlas of American Wars*, Volume 1, Frederick A. Paenger, 1969.

Copyright 1976, Board of Regents, The University of Texas System. All rights reserved.

ب



د

شکل (۵-۵): برخی از تصاویر متنی موجود در پایگاه مورد استفاده در این پایان نامه

۱۶ / فصلنامه کتاب

به‌موازات خطوط تلفنی، نصب و راه‌اندازی کرده است. نظام تلفنی بر پایه تبادل خطوط راه دور انگلستان مستقر است و شبکه اطلاعاتی بر پایه PACX 2000 نظامی‌های این شبکه‌ها استوار است. مهم‌ترین عملکرد شبکه اطلاعاتی این است که امکان ارتباط آسانی را برای گروه زیادی از پایانه‌ها همزمان فراهم سازد. ریزرایانه‌ها و رایانه‌های میزبانی که میان ساختمان‌های دانشگاه پراکنده هستند، جایگزین کاربرد شبکه این دانشگاه استفاده می‌کنند. این سیستم است که برنامه عوارضی‌گردن کتابخانه‌ها از آن سود می‌برد (پاکو، ۱۳۸۷).

۸. در تحقیق دیگری اثر شبکه کتابخانه‌های دانشگاهی بر نظام‌های اشتراک منابع و امانت بین کتابخانه‌های دانشگاهی و چالش‌های تحویل مدارک مورد لزوم که بسیار مورد تأکید است انجام شده است (دی‌جی‌پارو، ۱۳۸۷).

۹. دانشگاه ایالتی نیویورک براساس تأییدی که بر استفاده از شبکه‌های رایانه‌ای درون پردیس‌ها برای دستیابی به اطلاعات بیشتر به صورت الکترونیکی دارد به توسعه شبکه‌های کتابخانه‌های دانشگاهی پرداخته است. هدف این‌گونه شبکه‌ها ایجاد ارتباط پیوسته با دیگر نظام‌های اطلاعاتی در پردیس است. کتابخانه‌های دانشگاه ایالتی آریزونا نیز به همین طریق شبکه اطلاعاتی داخل پردیس خود را تشکیل داده‌اند (اریجی، ۱۳۸۸).

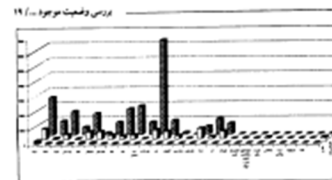
۱۰. برای کامل کردن نظام کتابخانه‌های دانشگاه کپیتان در افریقای جنوبی و ایجاد شبکه اطلاع‌رسانی در این دانشگاه بررسی انجام گرفته و گزارش آن برای ارائه به همایش رایانه‌ای گردن کتابخانه‌های دانشگاهی در افریقای جنوبی تهیه شده است (پاترسون، ۱۳۸۸).

۱۱. در زمینه تأثیر فناوری و اقتصاد اطلاعات بر کتابخانه‌های دانشگاهی تحقیقی انجام گرفته و توسعه و راه‌اندازی برای انجام همکاری در ساخت، ایجاد و گسترش شبکه اطلاعات کتابخانه‌های دانشگاهی بعنوان یک الزام در تحقق شبکه اطلاعات ملی ارائه می‌دهد (مورس، ۱۳۸۹).

۱۲. شبکه اطلاعات کنش‌ناقص کتابخانه‌های دانشگاهی اسکاتلند (سالیبن، یکی از تجلیات همکاری میان کتابخانه‌ها در اسکاتلند است. هدف این شبکه مستقیماً استفاده کنندگان کتابخانه‌ها هستند که با روشی «سفارشی‌تری» بدون کمک کتابداران به جست‌وجوی اطلاعات کنش‌ناقص منابع اطلاعاتی قابل دسترس در کتابخانه‌های دیگر می‌پردازند. اولین تولید سالیبن بسته ارتباطی مختصر و مفیدی است که به‌وسیله آن کاربران می‌توانند با شبکه دانشگاهی

- | | |
|-----------|-------------------|
| 1. Parkes | 2. De Gennaro |
| 3. Rigg | 4. Fainman |
| 5. Morris | 6. do-it-yourself |

الف

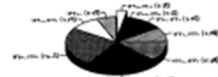


شماره ۱: توزیع انتشارات در مناطق مختلف، به‌شمار

اهلیت، طرح و فن آوری

در جدول ۲ مقایسه آماری میان دانشجویان، اساتید و کارکنان در سالهای ۱۳۸۶، ۱۳۸۵ و ۱۳۸۴ انجام گرفته که با اختلاف ۴۵٪ نفر در کل دانشگاه دامنه استفاده‌کننده این کتابخانه‌ها یا افزایش رو به‌روست که در مقایسه با سال ۱۳۸۲ افزایش سالانه ۲۵۰۰ نفر دانشجویان دانشجو را نشان می‌دهد. باقی است این افزایش دانشجو باید افزایش کارکنان و اساتید را به میان داشته باشد. لذا تنها تعداد کارکنان ۱۲۲ نفر افزایش داشته و لی تعداد اساتید ۷۸ نفر کاهش داشته است. این کتابخانه‌ها مهدی‌ارائه خدمات اطلاعاتی به ۲۲۶ نفر دانشجویی مقول نیز در مناطق مختلف تعریفی هستند.

تأسیس کتابخانه‌ها برای ارائه خدمات اطلاعاتی در طران عصر شبکه در دهه‌های مختلف انجام گرفته که بیشترین تعداد کتابخانه‌ها در دهه‌های ۵۰ که حدوداً ۲۵ درصد آنها و در دهه ۶۰ حدوداً ۶۱ درصد آنها تأسیس شده است. (شماره ۲: توزیعی تأسیس کتابخانه‌ها را در دهه‌های مختلف نشان می‌دهد).



شماره ۲: توزیعی تأسیس کتابخانه‌ها در دهه‌های مختلف تاریخ دانشگاه

ج

در این قسمت نتایج حاصل از اعمال روش پیشنهادی، استاندارد JPEG و استاندارد JPEG2000 روی تصاویر موجود در پایگاه داده مورد بررسی، در کار این پایان‌نامه، با هم مقایسه می‌شود. برای پیاده سازی استاندارد JPEG از نرم افزار Advanced JPEG Compressor و برای پیاده سازی استاندارد JPEG2000 از نرم افزار JPEG2000 Compressor استفاده شده است. استاندارد JPEG مبتنی بر

بلوک بندی تصویر و اعمال تبدیل DCT روی هر بلوک است. در استاندارد JPEG2000 استفاده از بلوک بندی اختیاری می‌باشد. در این روش، از تبدیل موجک استفاده می‌شود.

شکل ۵-۶-الف یک تصویر متنی اصلی و شکل ۵-۶-ب نتیجه فشرده‌سازی با روش پیشنهادی (با $NL=8$) و با میزان فشرده‌سازی حدود ۶۰ برابر را نشان می‌دهد. ملاحظه می‌شود که روش پیشنهادی تمایل به کاهش کانتراست (یا تباين) تصویر متنی دارد بنابراین، برای افزایش کانتراست می‌توان از تعدیل هیستوگرام استفاده کرد [۹۴]. نتیجه اعمال این بهسازی بر روی شکل ۵-۶-ب در شکل ۵-۶-ج نشان داده شده است. البته برای وضوح بیشتر، تنها قسمتی از شکل نشان داده شده است. شکل ۵-۶-د و ۵-۶-ه به ترتیب نتیجه اعمال استاندارد JPEG و JPEG2000 را روی همان قسمت از شکل ۵-۶-الف و در حالت میزان فشرده‌سازی تقریبی را ۵۰ برابر نشان می‌دهند.

با مقایسه شکل‌های ۵-۶-ج، ۵-۶-د و ۵-۶-ه ملاحظه می‌شود که علاوه بر بیشتر بودن میزان فشرده‌سازی در روش پیشنهادی (نسبت به دو روش دیگر) کیفیت تصویر متنی بازسازی شده نیز تا حد قابل توجهی بیشتر است. در تصویر حاصل از JPEG اثر بلوکی و در تصویر حاصل از JPEG2000 اثر نشتی به وضوح قابل مشاهده است در حالیکه در روش پیشنهادی، این اثرات نامطلوب به چشم دیده نمی‌شوند. البته در مقادیر بزرگ میزان فشرده‌سازی، اثر نامطلوب تولید نوارهای عمودی به چشم می‌خورد. نمونه‌ای از این اثر نواری در شکل ۵-۷ نشان داده شده است که مربوط به میزان فشرده‌سازی حدود ۷۰ برابر است. ملاحظه می‌شود که در این حال نیز کیفیت تصویر متنی افت چندانی زیادی پیدا نکرده است. در این حالت با استفاده از روش میانگین متحرک (۳-۵-۶-۲) نسبت به اصلاح تصویر اقدام می‌شود که تصویر بدست آمده از لحاظ دیداری هیچ آسیبی نمی‌بیند و مشکل اثر نواری هم از بین می‌رود.

In July 1845 Texas formally accepted an American proposal to be annexed to the United States. Already strained relations between the United States and Mexico rapidly worsened. President James K. Polk ordered General Zachary Taylor and his troops to Corpus Christi. In March 1846, under instructions, Taylor took up positions on the Rio Grande. On April 26 an American squadron of dragoons surrounded by Mexicans and unable to fight its way out surrendered. Several Americans lost their lives in the defeat.

On May 8 the Mexicans intercepted Taylor at Palo Alto but were driven back. The next day Mexicans again gave way in battle at Resaca de la Palma. In June Taylor began a march toward Monterrey, taking that city on September 25. Two months later the Americans took Saltillo and with little effort filed into Tampico.

Santa Anna now took the field against the American forces in northern Mexico, finally engaging the Americans at Buena Vista in February. Upon learning of the Mexican's design, General John E. Wool, marched from San Antonio to join Taylor's forces, abandoning his prior plans to take Chihuahua. After extremely bitter fighting Santa Anna pulled out his army, leaving Taylor in control of northern Mexico.

While Taylor pursued the enemy, Colonel Stephen W. Kearny took the "Army of the West" into New Mexico, capturing Santa Fe on August 16, 1846. Kearny then divided his forces, taking part to California and sending the remainder under Alexander W. Doniphan against Chihuahua. After General Winfield Scott captured Mexico City the two countries finally reached a settlement.

Source: Maps adapted from *The West Point Atlas of American Wars*, Volume 1, Frederick A. Praeger, 1959.

Copyright 1976, Board of Regents, The University of Texas System. All rights reserved.

الف

In July 1845 Texas formally accepted an American proposal to be annexed to the United States. Already strained relations between the United States and Mexico rapidly worsened. President James K. Polk ordered General Zachary Taylor and his troops to Corpus Christi. In March 1846, under instructions, Taylor took up positions on the Rio Grande. On April 26 an American squadron of dragoons surrounded by Mexicans and unable to fight its way out surrendered. Several Americans lost their lives in the defeat.

On May 8 the Mexicans intercepted Taylor at Palo Alto but were driven back. The next day Mexicans again gave way in battle at Resaca de la Palma. In June Taylor began a march toward Monterrey, taking that city on September 25. Two months later the Americans took Saltillo and with little effort filed into Tampico.

Santa Anna now took the field against the American forces in northern Mexico, finally engaging the Americans at Buena Vista in February. Upon learning of the Mexican's design, General John E. Wool, marched from San Antonio to join Taylor's forces, abandoning his prior plans to take Chihuahua. After extremely bitter fighting Santa Anna pulled out his army, leaving Taylor in control of northern Mexico.

While Taylor pursued the enemy, Colonel Stephen W. Kearny took the "Army of the West" into New Mexico, capturing Santa Fe on August 16, 1846. Kearny then divided his forces, taking part to California and sending the remainder under Alexander W. Doniphan against Chihuahua. After General Winfield Scott captured Mexico City the two countries finally reached a settlement.

Source: Maps adapted from *The West Point Atlas of American Wars*, Volume 1, Frederick A. Praeger, 1959.

Copyright 1976, Board of Regents, The University of Texas System. All rights reserved.

ب

In July 1845 Texas formally
to be annexed to the United States
between the United States and Mexico.
James K. Polk ordered General Zachary Taylor to
Corpus Christi. In March 1846, Taylor's army
positions on the Rio Grande. On March 25,
of dragoons surrounded by Mexican forces
out surrendered. Several American soldiers were
killed.

On May 8 the Mexicans in
were driven back. The next day
at Resaca de la Palma. In June
Monterrey, taking that city on
the Americans took Saltillo and
Tampico.

ع

In July 1845 Texas formally
to be annexed to the United States
between the United States and Mexico.
James K. Polk ordered General Zachary Taylor to
Corpus Christi. In March 1846, Taylor's army
positions on the Rio Grande. On March 25,
of dragoons surrounded by Mexican forces
out surrendered. Several American soldiers were
killed.

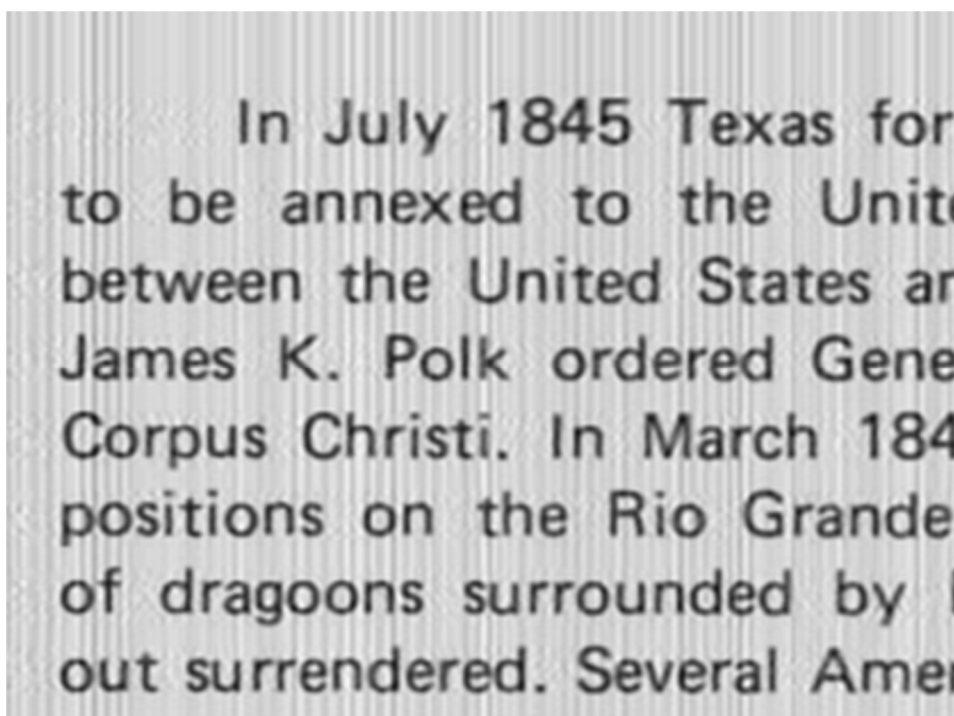
On May 8 the Mexicans in
were driven back. The next day
at Resaca de la Palma. In June
Monterrey, taking that city on
the Americans took Saltillo and
Tampico.

د

In July 1845 Texas forma
to be annexed to the United
between the United States and
James K. Polk ordered General
Corpus Christi. In March 1846,
positions on the Rio Grande. C
of dragoons surrounded by Me
out surrendered. Several Americi
On May 8 the Mexicans ir
were driven back. The next day
at Resaca de la Palma. In Ju
Monterrey, taking that city on
the Americans took Saltillo
Tampico.

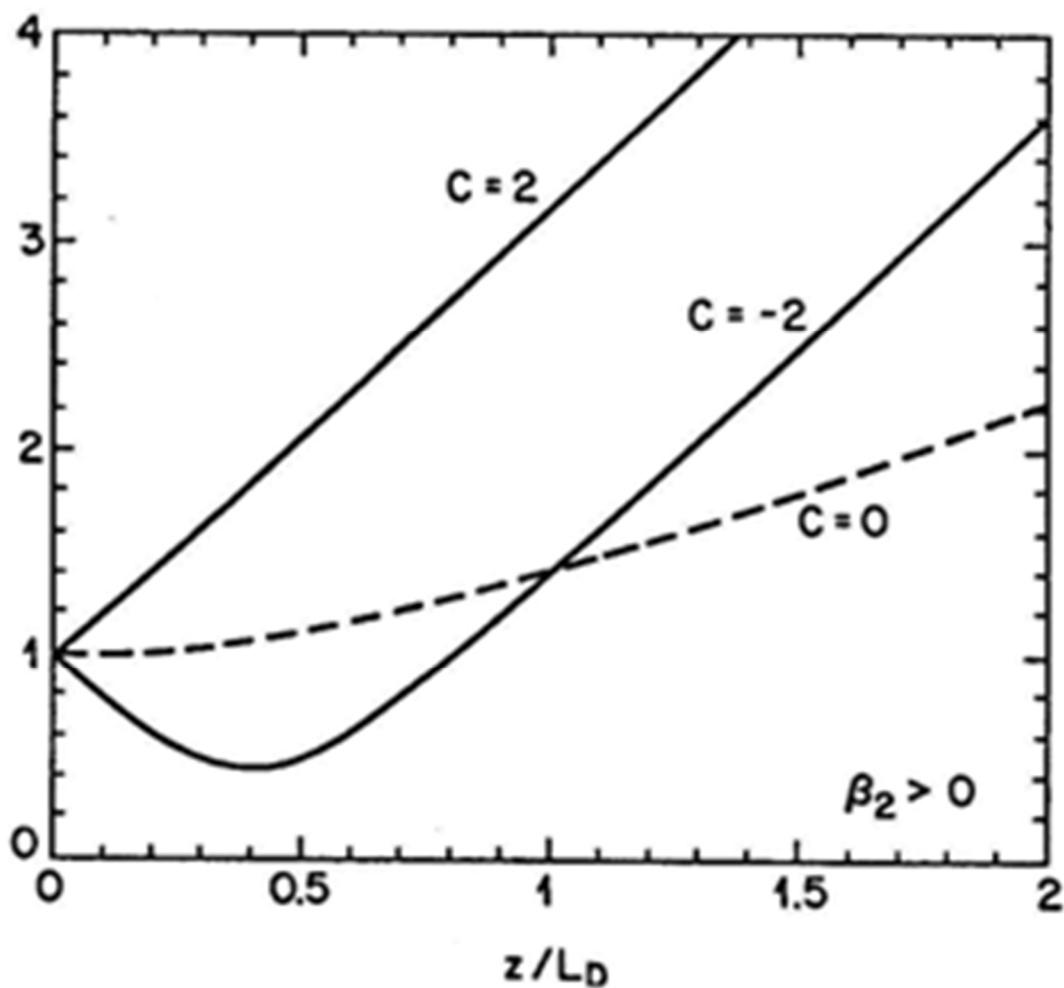
۵

شکل (۵-۶): (الف تا ه) تصویر اصلی، تصویر فشرده شده با روش پیشنهادی، افزایش کانتراست تصویر، تصویر فشرده شده با JPEG، تصویر فشرده شده با JPEG2000

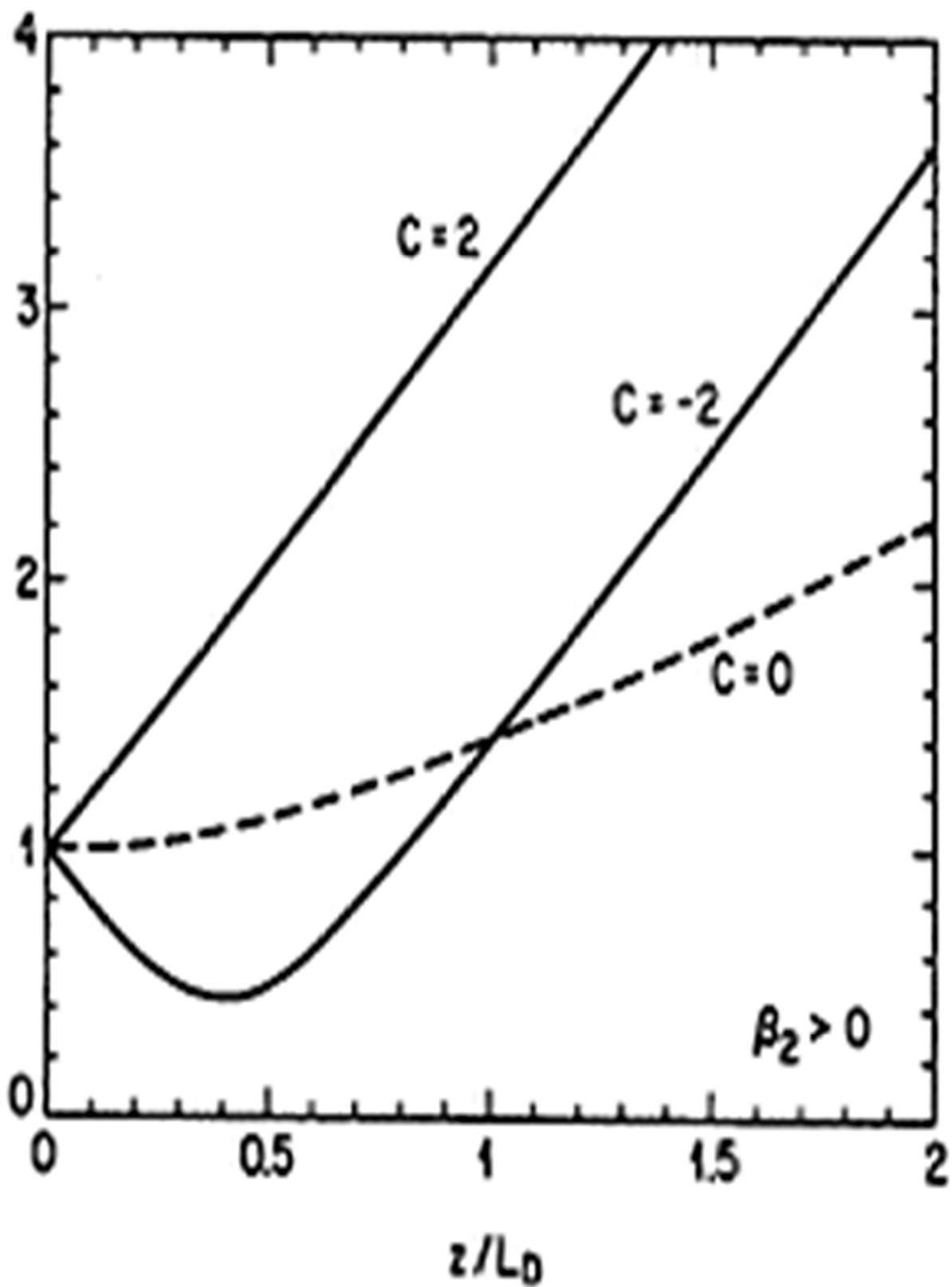


شکل (۵-۷): نمونه‌ای از اثر نواری تولید شده در میزانهای فشرده‌سازی بزرگ

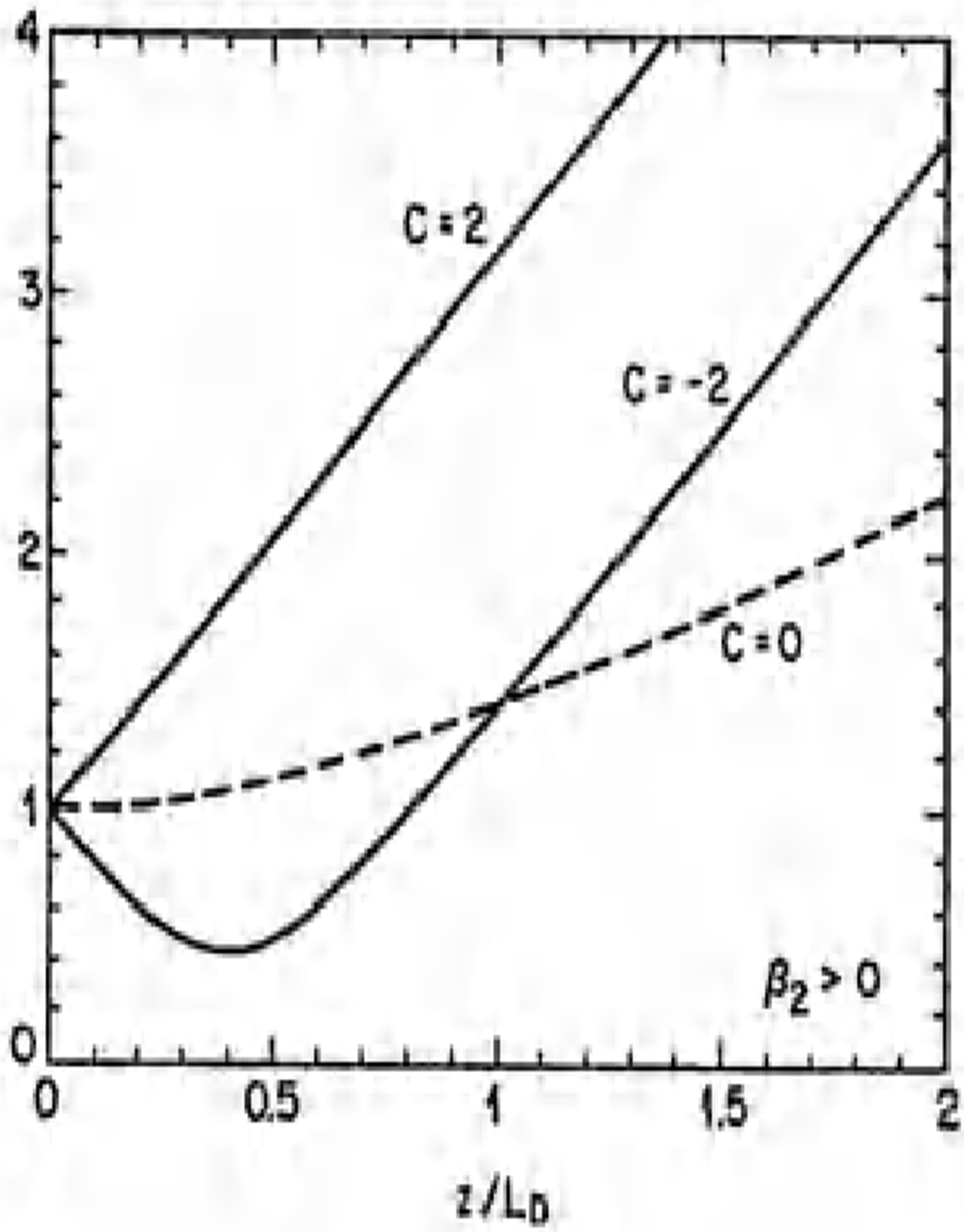
روش پیشنهادی در مورد تصاویر شامل قسمتهای عمدتاً گرافیکی نیز به خوبی عمل می‌کند. شکل‌های ۸-الف تا ۸-د یک نمونه تصویر متنی اصلی، تصویر فشرده شده با روش پیشنهادی، تصویر فشرده شده با روش JPEG، تصویر فشرده شده با روش JPEG2000 را نشان می‌دهند. در هر سه روش فشرده‌سازی، میزان فشرده‌سازی حدود ۵۰ برابر می‌باشد. در این مورد نیز اثرات بلوکی شدن و نشتی به وضوح قابل مشاهده است.



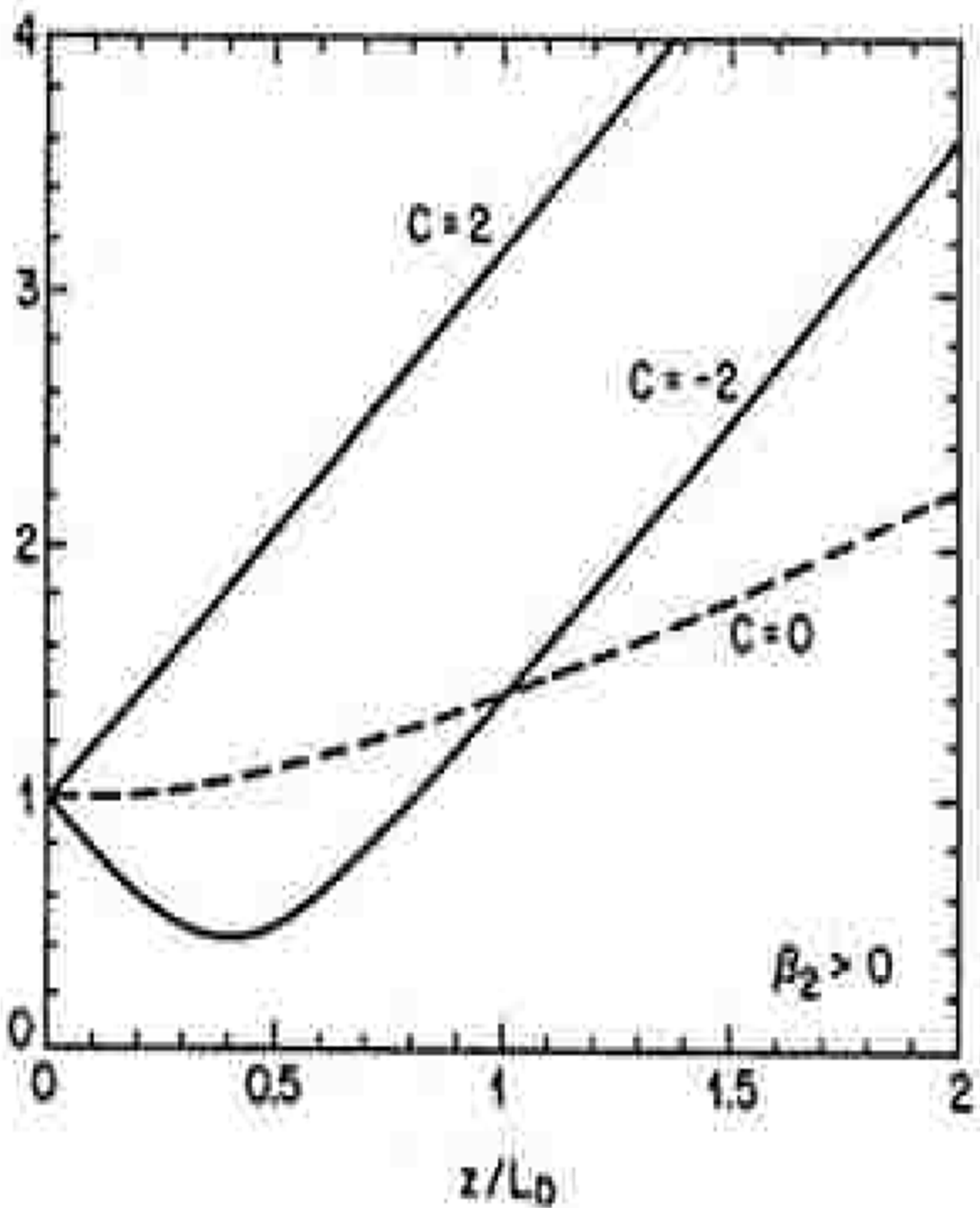
(الف)



(ب)



(c)



(د)

شکل (۸-۵): بررسی اثر روشهای فشرده‌سازی پیشنهادی، JPEG، و JPEG2000 در مورد تصاویر عمده‌تاً شامل قسمتهای ترسیمات خطی

پس از اعمال روش پیشنهادی به تصاویر موجود در پایگاه تصاویر، در مقدار SNR تقریباً یکسان، متوسط نسبت میزان فشرده‌سازی روش پیشنهادی به استاندارد JPEG حدود ۲,۲ برابر و در مورد استاندارد JPEG2000 حدود ۱,۶ برابر به دست آمد.

۴-۵ نتایج روش دوم جهت فشرده سازی تصاویر متنی توسط

پیش پردازش تصاویر جهت استفاده بهینه از استاندارد JPEG

برای شبیه سازی روش پیشنهادی از پایگاه تصاویر متنی چاپی موجود با کیفیت معمولی و درجه تفکیک مکانی ۳۰۰ نقطه بر اینچ استفاده شده است که از جاروب برخی اسناد فارسی و انگلیسی حاصل شده است. برخی از تصاویر موجود در این پایگاه در شکل ۵-۹ نشان داده شده است. قسمت عمده این تصاویر شامل متن است. برخی از تصاویر مانند شکل (۵-۹-ج) و (۵-۹-د) ممکن است شامل قسمتهای گرافیکی و ترسیمات خطی نیز باشند.



۱۶ / فصلنامه کتاب

به‌موازات خطوط تلفنی، نصب و راه‌اندازی کرده است. نظام تلفنی بر پایه تبادل خطوط راه دور انگلستان مستقر است و شبکه اطلاعاتی بر پایه FAX 2000 نظام‌های این شبکه‌ها استوار است. مهم‌ترین عملکرد شبکه اطلاعاتی این است که امکان ارتباط آسانی را برای گروه زیادی از پایانه‌ها همزمان فراهم سازد. ریزرایانه‌ها و رایانه‌های میزبانی که میان ساختمان‌های دانشگاه پراکنده هستند. جایزترین کاربرد شبکه این دانشگاه استفاده‌های مختلفی است که برنامه خودکار کردن کتابخانه‌ها از آن سود میبرد (پارکز^۱، ۱۹۸۷).

۸. در تحقیق دیگری اثر شبکه کتابخانه‌های دانشگاهی بر نظام‌های اشتراک منابع و امانت بین کتابخانه‌های دانشگاهی و چالش تحویل مدارک مورد لزوم که بسیار مورد تأکید است انجام شده است (دی چینارو^۲، ۱۹۸۷).

۹. دانشگاه ایالتی نیویورک براساس تأکیدی که بر استفاده از شبکه‌های رایانه‌ای درون پردیس‌ها برای دستیابی به اطلاعات بیشتر به صورت الکترونیکی دارد به توسعه شبکه‌های کتابخانه‌های دانشگاهی پرداخته است. هدف این‌گونه شبکه‌ها ایجاد ارتباط پیوسته با دیگر نظام‌های اطلاعاتی در پردیس است. کتابخانه‌های دانشگاه ایالتی آریزونا نیز به همین طریق شبکه اطلاعاتی داخل پردیس خود را تشکیل داده‌اند (رایجو^۳، ۱۹۸۸).

۱۰. برای کامل کردن نظام کتابخانه‌های دانشگاه کپ‌تان در افریقای جنوبی و ایجاد شبکه اطلاع‌رسانی در این دانشگاه بررسی انجام گرفته و گزارش آن برای ارائه به همایش رایانه‌ای کردن کتابخانه‌های دانشگاهی در افریقای جنوبی تهیه شده است (پاترسون^۴، ۱۹۸۸).

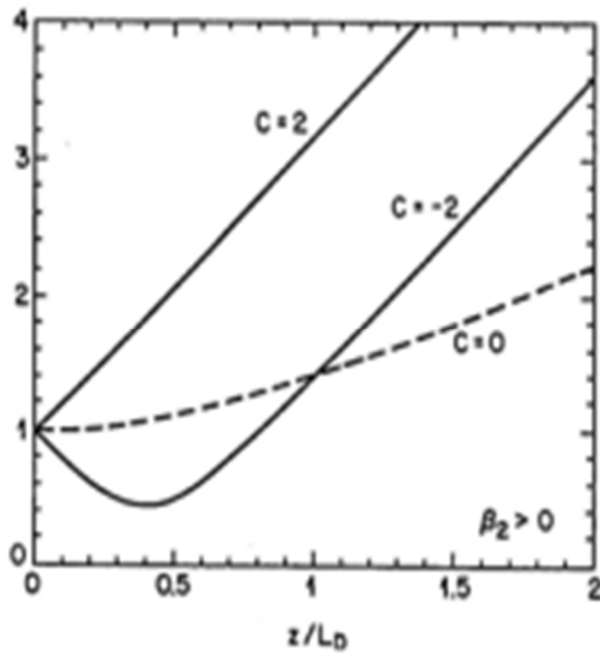
۱۱. در زمینه تأثیر فن آوری و اقتصاد اطلاعات بر کتابخانه‌های دانشگاهی تحقیقی انجام گرفته و توصیه و راهنمای برای انجام همکاری در ساخت، ایجاد و گسترش شبکه اطلاعات کتابخانه‌های دانشگاهی بعنوان یک لازم در تحقق شبکه اطلاعات ملی ارائه می‌دهد (مورس^۵، ۱۹۸۹).

۱۲. شبکه اطلاعات کتابخانه‌های دانشگاهی اسکاتلند (سالیبن) یکی از تجلیات همکاری میان کتابخانه‌ها در اسکاتلند است. هدف این شبکه مستقیماً استفاده کنندگان کتابخانه‌ها هستند که با روش استفاده فردی^۶ بدون کمک کتابداران به جست‌وجوی اطلاعات کتابخانه‌ای متابع اطلاعاتی قابل دسترس در کتابخانه‌های دیگر می‌پردازند. اولین تولید سالیبن بسته ارتباطی مختصر و مفید است که به وسیله آن کاربران می‌توانند با شبکه دانشگاهی

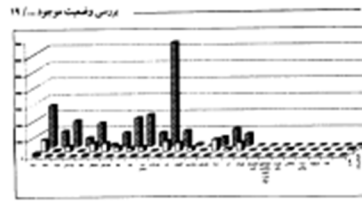
- | | |
|-----------|-------------------|
| 1. Parkes | 2. De Gennaro |
| 3. Riggs | 4. Paterson |
| 5. Morris | 6. do-it yourself |

ب

الف

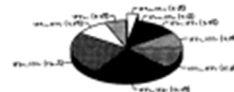


د



شماره ۱: فرآیند دانشجو در مناطق مختلف تحصیلی

اهداف، نحوه و فنی دارد.
 در جدول ۳ مقایسه آماری میان دانشجویان استانان و کارکنان در سالهای ۱۳۷۶، ۱۳۷۷، ۱۳۷۸ و ۱۳۷۹ انجام گرفته که با اختلاف ۳۵٪ فر در کل دانشگاه جامعاً استفاده کننده این کتابخانه با افزایش رو به روست که در مقایسه با سال ۱۳۷۶ افزایش سالانه ۲۵۰۰ نفر دانشجو را نشان می‌دهد. بهر است این افزایش دانشجو باید افزایش کارکنان و استادان را به دنبال داشته باشد. لذا تنها تعداد کارکنان ۱۹۹۲ نفر افزایش داشته ولی تعداد استادان ۷۸ نفر کاهش داشته است. این کتابخانه هم‌اکنون راه خدمات اطلاعاتی به ۲۲۶ نفر دانشجوی معقول نیز در مناطق مختلف تحصیلی هستند.
 تأسیس کتابخانه برای ارائه خدمات اطلاعاتی در طول عمر دانشگاه در دهه‌های مختلف انجام گرفته که بیشترین تعداد کتابخانه در دهه‌های ۵۰ که حدوداً ۲۵ درصد آنها و در دهه ۴۰ حدوداً ۲۱ درصد آنها تأسیس شده است. (شماره ۶ فرآیند تأسیس کتابخانه را در دهه‌های مختلف نشان می‌دهد.)



شماره ۲: فرآیند تأسیس کتابخانه در دهه‌های مختلف در دانشگاه

ج

شکل (۵-۹): برخی از تصاویر متنی موجود در پایگاه مورد استفاده در این روش

در این قسمت نتایج حاصل از اعمال روش پیشنهادی و استاندارد JPEG روی تصاویر موجود در پایگاه داده مورد بررسی در کار این پایان‌نامه با هم مقایسه می‌شود. مقادیر پارامترهای N و Th1 به ترتیب برابر ۸ و ۱، ۰ انتخاب شده‌اند. انتخاب این پارامترها به صورت تجربی انجام شده است. برای پیاده سازی استاندارد JPEG از نرم افزار Advanced JPEG Compressor استفاده شده است.

شکل (۵-۱۰-الف) یک تصویر متنی اصلی، شکل (۵-۱۰-ب) نتیجه فشرده‌سازی با اعمال به ترتیب روشهای پیشنهادی و سپس JPEG، و بالاخره، شکل (۱۰-ج) نتیجه فشرده‌سازی با اعمال مستقیم روش JPEG به تصویر (۱۰-الف) را نشان می‌دهند. دو شکل اخیر در میزان فشرده‌سازی تقریباً یکسان به دست آورده شده‌اند.

با مقایسه دو شکل اخیر ملاحظه می‌شود که اولاً کیفیت کلی تصویر حاصل از روش پیشنهادی به طور قابل ملاحظه‌ای بهتر از روش JPEG است. ثانیاً اثر بلوکی در تصویر حاصل از اعمال مستقیم JPEG دیده می‌شود اما در روش پیشنهادی این اثر تقریباً دیده نمی‌شود. در روش پیشنهادی اثر بلوکی شدن

تصویر، در میزانهای فشرده‌سازی بالا دیده می‌شود. علت دیده نشدن اثر بلوکی در روش پیشنهادی مربوط به بازچینی ضرایب تبدیل و انتقال مولفه‌های فرکانس میانی به سمت فرکانسهای پایین‌تر است که منجر به هموارتر شدن تصویر می‌گردد.

In July 1845 Texas formally accepted an American proposal to be annexed to the United States. Already strained relations between the United States and Mexico rapidly worsened. President James K. Polk ordered General Zachary Taylor and his troops to Corpus Christi. In March 1846, under instructions, Taylor took up positions on the Rio Grande. On April 26 an American squadron of dragoons surrounded by Mexicans and unable to fight its way out surrendered. Several Americans lost their lives in the defeat.

On May 8 the Mexicans intercepted Taylor at Palo Alto but were driven back. The next day Mexicans again gave way in battle at Resaca de la Palma. In June Taylor began a march toward Monterrey, taking that city on September 25. Two months later the Americans took Saltillo and with little effort filed into Tampico.

Santa Anna now took the field against the American forces in northern Mexico, finally engaging the Americans at Buena Vista in February. Upon learning of the Mexican's design, General John E. Wool, marched from San Antonio to join Taylor's forces, abandoning his prior plans to take Chihuahua. After extremely bitter fighting Santa Anna pulled out his army, leaving Taylor in control of northern Mexico.

While Taylor pursued the enemy, Colonel Stephen W. Kearny took the "Army of the West" into New Mexico, capturing Santa Fe on August 16, 1846. Kearny then divided his forces, taking part to California and sending the remainder under Alexander W. Doniphan against Chihuahua. After General Winfield Scott captured Mexico City the two countries finally reached a settlement.

Source: Maps adapted from *The West Point Atlas of American Wars*, Volume 1, Frederick A. Praeger, 1959.

Copyright 1976, Board of Regents, The University of Texas System. All rights reserved.

الف

In July 1845 Texas formally
to be annexed to the United States
between the United States and Mexico
James K. Polk ordered General Zachary Taylor to
Corpus Christi. In March 1846,
positions on the Rio Grande. On
of dragoons surrounded by Mexican
out surrendered. Several American
On May 8 the Mexicans in
were driven back. The next day
at Resaca de la Palma. In June
Monterrey, taking that city on
the Americans took Saltillo
Tampico.

ب

In July 1845 Texas formally
to be annexed to the United States
between the United States and Mexico
James K. Polk ordered General Zachary Taylor to
Corpus Christi. In March 1846,
positions on the Rio Grande. On
of dragoons surrounded by Mexican
out surrendered. Several American
On May 8 the Mexicans in
were driven back. The next day
at Resaca de la Palma. In June
Monterrey, taking that city on
the Americans took Saltillo
Tampico.

ج

شکل (۵-۱۰): (الف تا ج) تصویر اصلی، تصویر فشرده شده با اعمال روش پیش‌پردازش پیشنهادی و سپس روش
JPEG و تصویر فشرده شده با اعمال مستقیم JPEG.

پس از اعمال روش پیشنهادی به تصاویر موجود در پایگاه تصاویر، در مقدار SNR تقریباً یکسان، متوسط نسبت میزان فشرده‌سازی روش پیشنهادی به میزان فشرده‌سازی حاصل از اعمال مستقیم استاندارد JPEG حدود ۱,۵ برابر به دست آمد.

۵-۵ جمع بندی و نتیجه گیری

در این فصل پایگاه داده ایجاد شده معرفی شد. پایگاه داده ارائه شده تصاویر متنوع با دقت‌های (dpi) متفاوت را شامل می‌شد. نتایج مربوط به روش پیشنهادی برای دودویی سازی و دو روش برای فشرده‌سازی تصاویر متنی نیز ارائه گردید. این روش‌ها جهت ایجاد لایه‌های متفاوت و فشرده‌سازی هر لایه با روشی بهینه، برای به کارگیری در فشرده‌سازی تصاویر متنی، با MRC لازم می‌باشد.

روش دودویی سازی پیشنهادی دارای سرعت مناسب و وقفی بودن انتخاب آستانه و استفاده از تحلیل چند مقیاسی برای انتخاب آستانه میباشد و همچنین از دو روش متفاوت جهت دودویی سازی استفاده شد، در این روش از شبکه عصبی پرسپترون به عنوان طبقه بند و از فیلتر بانک به جهت تحلیل چند مقیاسی استفاده شد. به علت ذات موازی کاری در این روش پیشنهادی قابلیت استفاده از آن در FPGAها و DSPها جهت کارایی بالاتر امکان پذیر می باشد. و در نهایت این روش نسبت به روش اتسو به صورت قابل ملاحظه ای دارای کارایی بالاتری می‌باشد.

در روش پیشنهادی اول جهت فشرده سازی تصاویر متنی از تبدیل KLT استفاده شد که نسبت به سایر تبدیلات ریاضی مرسوم مانند DCT و تبدیل موجک دارای کارایی بالاتری می باشد. همچنین کیفیت تصویر متنی باز سازی شده تا حد قابل توجهی نسبت به استانداردهای مرسوم (JPEG و JPEG2000) بهبود یافته است و همچنین از کارایی مناسبی بر روی نمونه های گرافیکی برخوردار است. البته در مقادیر بزرگ فشرده سازی اثر نا مطلوب تولید نوارهای عمودی به چشم می‌خورد.

در روش پیشنهادی دوم با اعمال یک پیش پردازش بر روی تصاویر متنی این تصاویر را جهت استفاده بهینه از استاندارد JPEG آماده می‌کند. این روش عمدتاً مبتنی بر دو تکنیک وزن دهی مناسب ضرایب تبدیل DCT و تکنیک بازچینی دنباله ضرایب فرکانسی عمل می‌کند. اولین تکنیک باعث حفظ تقریبی کیفیت لبه های تصویر متنی و دومین تکنیک باعث حفظ تقریبی میزان فشرده سازی استاندارد JPEG می شود. در این روش همچنین کیفیت کلی تصویر حاصل از روش پیشنهادی به طور قابل ملاحظه ای بهتر از روش استاندارد JPEG بوده و اثر بلوکی در تصویر، حاصل از اعمال مستقیم روش JPEG وجود ندارد. این بهبود ناشی از بازچینی ضرایب تبدیل DCT و انتقال مولفه های فرکانس میانی به سمت فرکانس های پایین تر است که منجر به هموارتر شدن تصویر می گردد.

جمع‌بندی و نتیجه‌گیری

۱-۶ خلاصه و جمع‌بندی مطالب گفته شده

امروزه، بیشتر تولیدات تاریخی، هنری و علمی روی کاغذ منتشر می‌شوند. بسیاری از کتابخانه‌های بزرگ دنیا در حال تبدیل منابع خود به شکل دیجیتال می‌باشند زیرا ذخیره اطلاعات مربوط به اسناد مختلف به شکل دیجیتالی، تا حد زیادی مقرون به صرفه‌تر از نگهداری آنها در قفسه‌های کتابخانه است. بنا به دلایل اقتصادی، برای ذخیره چنین حجم عظیمی از اطلاعات به فشردگی آنها نیاز داریم. در گذشته چنین نیازی وجود نداشت، زیرا حجم اطلاعات چندان زیاد نبود، اما امروزه با توسعه سریع علم و پیشرفت تکنولوژی، روزانه به مقدار بسیار زیادی به حجم اطلاعات افزوده می‌شود. بنابراین، فشردگی اطلاعات تصاویر اسناد (مرکب) موضوع جدیدی است که در دهه اخیر توجه زیادی به آن شده است.

روش فشردگی پیشنهادی (MRC) مورد استفاده در سیستم مذکور، نیاز به روشی برای فشردگی تصاویر دودویی و روشی برای جداسازی متن از پس‌زمینه آن را طلب می‌کرد. در این راستا یک روش برای دودویی سازی تصاویر متنی و دو روش نیز برای فشردگی تصاویر متنی ارائه شد.

روش پیشنهادی برای دودویی سازی تصاویر متنی در این پایان نامه جهت تصاویر متنی سطح خاکستری ارائه شد که مبتنی بر ایده‌ی ترکیب طبقه‌بندها (با استفاده از شبکه عصبی) و استفاده از تحلیل چندمقیاسی (با استفاده از فیلتربانک) بود. چهار مزیت روش پیشنهادی عبارتند از محاسبه‌ی

آستانه به صورت وفقی و محلی، استفاده از تحلیل چندمقیاسی، استفاده از ایده‌ی ترکیب طبقه بندها، برخورداری از سرعت اجرای بالا و مناسب برای پیاده‌سازی سخت‌افزاری.

اولین روش پیشنهادی جهت فشرده‌سازی تصاویر متنی مبتنی بر استفاده از تبدیل KLT به منظور فشرده‌سازی تصاویر متنی ارائه گردید. کارایی فشرده‌سازی روش پیشنهادی با دو استاندارد متداول امروزی، JPEG و JPEG2000 مقایسه گردید. نتایج حاصل از شبیه‌سازی نشان می‌دهند که در SNR تقریباً یکسان، متوسط میزان فشرده‌سازی روش پیشنهادی نسبت به استانداردهای JPEG و JPEG2000 بیشتر می‌باشد.

یک ضعف مهم استاندارد JPEG مشاهده شدن اثر بلوکی و در مورد استاندارد JPEG2000 مشاهده شدن اثر نشتی در مقادیر متوسط میزان فشرده‌سازی است. در حالیکه، روش پیشنهادی تنها در مقادیر بزرگ میزان فشرده‌سازی از خود اثر نواری نشان می‌دهد.

دومین روش پیشنهادی جهت فشرده‌سازی تصاویر متنی در این پایان‌نامه، یک روش پیش پردازش تصاویر متنی قبل از اعمال فرآیند فشرده‌سازی JPEG می‌باشد که کارایی فشرده‌سازی آن را نسبت به حالت اعمال مستقیم روش JPEG به تصویر متنی افزایش داد.

روش پیش‌پردازش پیشنهادی عمدتاً مبتنی بر دو تکنیک وزن‌دهی مناسب ضرایب تبدیل DCT و تکنیک بازچینی دنباله‌ی ضرایب فرکانسی عمل می‌کند. نتایج روش پیشنهادی با روش اعمال مستقیم فشرده‌سازی JPEG مقایسه گردید و با توجه به نتایج به دست آمده مشخص شد که در مقدار SNR تقریباً یکسان، متوسط نسبت میزان فشرده‌سازی روش پیشنهادی به میزان فشرده‌سازی حاصل از اعمال مستقیم استاندارد JPEG دارای برتری می‌باشد.

کارایی روش‌های فشرده‌سازی پیشنهادی، از جنبه‌های مختلف شامل نرخ فشرده‌سازی روش‌های پیشنهادی برای فشرده‌سازی تصاویر متنی، مدت زمان اجرای نسبی روش‌های پیشنهادی برای فشرده‌سازی تصاویر متنی، کارایی جداسازی روش پیشنهادی برای جداسازی متن از پس‌زمینه و

نهایتاً نرخ فشرده‌سازی روش پیشنهادی برای فشرده‌سازی تصاویر اسناد مورد ارزیابی و مقایسه قرار گرفت.

همچنین در این پایان نامه به معرفی پایگاه داده ایجاد شده جهت تصاویر متنی نیز پرداخته شد که که تصاویر موجود در آن دارای درجه تفکیک‌های متفاوت ۱۰۰، ۲۰۰، ۳۰۰، ۶۰۰ نقطه بر اینچ (dpi) و شامل تنوع رسم الخط فارسی و عربی و انگلیسی می باشد. هر یک از این تصاویر نیز به ترکیب های تمام متن یا ترکیبی از متن و گرافیک تقسیم می شود که برحسب افزایش یا کاهش میزان وجود گرافیک در این تصاویر در رده بندی جدیدی گنجانده شد.

۲-۶ نتیجه گیری

با توجه به انجام آزمایشهای مختلف انجام شده در فصل پنجم، ملاحظه گردید که طبق هر یک از معیارهای بیان شده در بخش قبلی، روش پیشنهادی تا حد قابل توجهی از عملکرد بهتری برخوردار بود. این عملکرد برای تصاویر عمدتاً متنی، بهبود بیشتری نسبت به سایر انواع تصاویر ارائه نمود.

۳-۶ پیشنهادهایی برای ادامه کار

جهت بهبود کارایی سیستم پیشنهادی در کار این رساله می توان موارد مختلفی را مورد بررسی قرار داد از جمله:

۱- افزایش تصاویر موجود در پایگاه داده ایجاد شده از حیث تنوع نوع رسم الخط به خصوص تصاویر مربوط به رسم الخط عربی و همچنین افزودن قسمت هایی به عنوان تصاویر آسیب دیده که می تواند شامل مباحثی نظیر کجی تصویر و وجود نویز و خمش در محل شیرازه ومات بودن تصویر و...باشد.

۲- در نظر گرفتن تمهیداتی جهت کاهش مدت زمان اجرای الگوریتمهای مختلف مورد استفاده در روشهای پیشنهادی.

۳- استفاده از فیلترهای مختلف یا پیش‌پردازشهای مختلف جهت افزایش کیفیت دیداری تصویر فشرده‌شده و افزایش همزمان نرخ فشرده‌سازی.

۴- استفاده از بازشناسی حروف (OCR) جهت افزایش نرخ فشرده‌سازی. ترکیب استفاده از بازشناسی حروف با مدل محتوای ترکیبی گرچه تاکنون به صورت جدی برای افزایش نرخ فشرده‌سازی مورد استفاده قرار نگرفته است اما به عنوان موضوع امیدبخشی می‌تواند مورد توجه قرار گیرد. در صورت امکان بهتر است این بازشناسی مستقل از نوع رسم‌الخط باشد.

۵- ایجاد راهکاری جهت بهبود عملکرد تبدیل KLT در تصاویر متنی، که به صورت عمدتاً گرافیکی می‌باشد.

فهرست مراجع

- [١] O. E. Kia and D. S. Doermann, "Residual Coding in Document Image Compression," IEEE Trans. on Image Proc. vol. 9, no. 6, pp.961-969, 2000.
- [٢] T. Kanungo, R. M. Haralick, and I. T. Phillips, "Global and Local Document Degradation Models", Proceedings of the International Conference on Document Analysis and Recognition, pp. 730-734, 1993.
- [٣] D. Salomon, "A Concise Introduction to Data Compression," Springer-Verlag, London, 2008.
- [٤] K. Sayood Editor, "Lossless Compression Handbook," Academic Press, 2003.
- [٥] D. Salomon, "Data Compression, The Complete Reference," Fourth Edition, Springer-Verlag, London, 2007.
- [٦] I. H. Witten, A. Moffat and T. C. Bell, "Managing Gigabytes: Compressing and Indexing Documents and Images," Second Edition, Academic Press, 1999.
- [٧] I. M. Pu, "Fundamental Data Compression," Butterworth-Heinemann, 2006.
- [٨] 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.
- [٩] 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.
- [١٠] G. Feng and C. A. Bouman, "High-Quality MRC Document Coding," IEEE Trans. Image Proc., vol. 15, no. 10, pp. 3152-3169, 2006.
- [١١] 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.
- [١٢] 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.

- [١٣] M. Sarfraz, A. Zidouri and S. N. Nawaz, On Offline Arabic Character Recognition, in Computer-Aided Intelligent Recognition: Techniques and Applications, M. Sarfraz Editor, John Wiley & Sons Ltd., 2005.
- [١٤] A. Gersho and R. Gray, "Vector Quantization and Signal Compression", Norwell, MA: Kluwer, 1992.
- [١٥] A. Jain, "Fundamentals of Digital Image Processing", Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [١٦] M. F. Barnsley and L. P. Hurd, "Fractal Image Compression", Wellesley, MA: Peters, 1993.
- [١٧] C. W. Kok and T. Q. Nguyen, "Document Image Compression by Sub-band System," IEEE International Symposium on Circuits and Systems, ISCAS '96, vol. 2, pp. 688-691, May 1996.
- [١٨] R. L. de Queiroz, R. Buckley, and M. Xu, "Mixed Raster Content (MRC) Model for Compound Image Compression," SPIE Conf. on Visual Communications and Image Processing, vol. 3653, pp. 1106–1117, Feb. 1999.
- [١٩] E. Y. Lam, "Compound Document Compression with Model-Based Biased Reconstruction," Journal of Electronic Imaging, vol. 13, no. 1, pp. 191-194, 2004.
- [٢٠] H. Cheng and C. A. Bouman, "Multilayer Document Compression Algorithm," IEEE ICIP, Oct. 1999.
- [٢١] D. Huttenlocher, P. Felzenszwalb, and W. Rucklidge, "Digipaper: A Versatile Color Document Image Representation," IEEE ICIP, vol. 1, pp. 219–223, Oct. 1999.
- [٢٢] K. Barthel, S. McPartlin, and M. Thierschmann, "New Technology for Raster Document Image Compression," SPIE Conf. on Document Recognition and Retrieval VII, vol. 3967, pp. 286–290, Jan. 2000.
- [٢٣] P. G. Howard, F. Kossentini, S. Forchhammer and W. J. Rucklidge, "The Emerging JBIG2 Standard," IEEE Trans. on Circuits and Sys for Video Tech., vol. 8, no. 7, pp. 838-848, 1998.
- [٢٤] Y. Ye and P. Cosman, "Fast and Memory Efficient Text Image Compression With JBIG2," IEEE Trans. on Image Proc., vol. 12, no. 8, pp. 944-956, 2003.
- [٢٥] P. G. Howard, "Text Image Compression Using Soft Pattern Matching," The Computer Journal, vol. 40, no. 2/3, pp. 146-156, 1997.
- [٢٦] Y. Ye and P. Cosman, "Dictionary Design for Text Image Compression with JBIG2," IEEE Trans. on Image Proces., vol. 10, no. 6, pp. 818-828, 2001.
- [٢٧] R.N. Ascher, G. Nagy, "A Means for Achieving a High Degree of Compaction on Scan-Digitized Printed Text," IEEE Trans. Comput., vol. 23, pp. 1174-1179, 1974.

- [٢٨] W. K. Pratt, P. J. Capitant, W. H. Chen, E. R. Hamilton, and R. H. Wallis, "Combined Symbol Matching Facsimile Data Compression System," Proc. IEEE, vol. 68, no. 7, pp. 786-796, 1980.
- [٢٩] N. F. Brickman and W. S. Rosenbaum, "Word Autocorrelation Redundancy Match (WARM) Technology," IBM J. Res. Devel., vol. 26, no. 6, pp. 681-686, 1982.
- [٣٠] O. Johnsen, J. Segen, and G. L. Cash, "Coding of Two-Level Pictures by Pattern Matching and Substitution," Bell Syst. Tech. J, vol. 62, no. 8, pp. 2513-2545, 1983.
- [٣١] Y. Yang, H. Yan and D. Yu, "Content-Lossless Document Image Compression Based on Structural Analysis and Pattern Matching," Pattern Recognition, vol. 33, pp. 1277-1293, 2000.
- [٣٢] M. B. Carvalho, E. A. B. Silva and W. A. Finamore, "Multidimensional Signal Compression Using Multiscale Recurrent Patterns," Signal Processing, vol. 82, pp. 1559-1580, 2002.
- [٣٣] S. Inglis and I. H. Witten, "Compression-Based Template Matching," Data Compression Conference, DCC '94 Proceedings, pp. 106 – 115, 1994.
- [٣٤] Q. Zhang and J. M. Danskin, "A Pattern-Based Lossy Compression Scheme for Document Images," Electronic Publishing, vol. 8, no. 2&3, pp. 221-233, 1995.
- [٣٥] I. H. Witten, T. C. Bell, H. Emberson, S. Inglis and A. Moffat, "Textual Image Compression: Two-Stage Lossy/Lossless Encoding of Textual Images", Proceedings of the IEEE, vol. 82, no. 6, 1994.
- [٣٦] I. H. Witten, A. Moffat and T. C. Bell, "Managing Gigabytes: Compressing and Indexing Documents and Images," Second Edition, Academic Press, 1999.
- [٣٧] F. Ono, P. G. Howard and D. Fernandes, "The Emerging JBIG-2 Standard," ISO/IEC JTC 1/SC 29/WG1 (ITU-T SG8), N1397, July 1997.
- [٣٨] S. Chen, H. Yan and Z. Xu, "Compression of Chinese Document Images Based on Morphological Analysis and Pattern Matching," Optical Engineering, vol. 45, no. 10, 107001, 2006.
- [٣٩] O. N. Gerek, A. E. Cetin and A. H. Tewfik, "Wavelet Domain Textual Coding of Ottoman Script Images," Proc. SPIE, vol. 2727, pp. 568-578, 1996.
- [٤٠] S. D. Wei and S. H. Lai, "Efficient Normalized Cross Correlation Based on Adaptive Multilevel Successive Elimination," ACCV 2007, pp. 638-646, 2007.

- [٤١] L. D. Stefano, S. Mattoccia and F. Tombari, "An Algorithm for Efficient and Exhaustive Template Matching," ICIAR 2004, pp. 408-415, 2004.
- [٤٢] G. Cirrincione and M. Cirrincione, "Neural Networks for Matching in Computer Vision," KES 2007/WIRN 200, pp. 688-695, 2007.
- [٤٣] D. Marimon and T. Ebrahimi, "Efficient Rotation-Discriminative Template Matching," CIARP 2007, pp. 221-230, 2007.
- [٤٤] J. Eggert, C. Zhang and E. Corner, "Template Matching for Large Transformations," ICANN 2007, pp. 169-179, 2007.
- [٤٥] J. Song, B. Chen, Z. Chi, X. Qiu and W. Wang, "Face Recognition Based on Bi-level Template Matching," ICIC 2007, pp. 1131-1139, 2007.
- [٤٦] ISO/IEC International Standard 11544, Progressive Bi-Level Image Compression, JBIG, ITU-Recommendation T.82.
- [٤٧] T. C. Bell, J. G. Cleary and I. H. Witten, Text Compression, Englewood Cliffs, NJ: Prentice Hall, 1990.
- [٤٨] H. Freeman, "Techniques for the digital computer analysis of chain-encoded arbitrary plane curves," Proc. Nat. Electronics Conf., vol.17, pp.421-432, 1961.
- [٤٩] T. Pavlidis, "Algorithms for Graphics and Image," Processing. Computer Science Press, Maryland, 1982.
- [٥٠] B. Jahne, "Digital Image Processing," Springer-Verlag Berlin Heidelberg, 6th edition, 2005.
- [٥١] R. C. Gonzalez and R. E. Woods, "Digital Image Processing," 2nd edition, Prentice Hall, 2002.
- [٥٢] J. C. Ross, "The Image Processing Handbook", 5th edition, CRC Press, 2007.
- [٥٣] H. C. Liu and L.S. Srinath, "Corner Detection from Chain-Code," Pattern Recognition, vol. 23, pp. 51-68, 1990.
- [٥٤] G. Cirrincione and M. Cirrincione, "Neural Networks for Matching in Computer Vision," KES 2007/WIRN 200, pp. 688-695, 2007.
- [٥٥] D. Marimon and T. Ebrahimi, "Efficient Rotation-Discriminative Template Matching," CIARP 2007, pp. 221-230, 2007.

- [Δϕ] J. Eggert, C. Zhang and E. Corner, "Template Matching for Large Transformations," ICANN 2007, pp. 169-179, 2007.
- [Δψ] H. Freeman and R. Shapira, "Determining the Minimum-Area Enclosing Rectangle of an Arbitrary Closed Curve," *Comm. ACM*, vol. 18, pp. 409-413, 1975.
- [Δλ] Z. Galil and K. Park, "Truly Alphabet-Independent Two-Dimensional Pattern Matching," *Proc. IEEE Syrup. Foundations Computer Science*, pp. 247-256, 1992.
- [Δϱ] J. Lebledz, "Shape Similarity to Alphanumeric Sign", ICCVG 2004, Warsaw Poland.
- [ϕ•] D. Yu and H. Yan, "An Efficient Algorithm for Smoothing, Linearization and Detection of Structural Feature Points of Binary Image Contours," *Pattern Recognition*, vol. 30, pp. 57-69, 1997.
- [ϕ\] S. M. Maillet and Y. M. Sharaiha, "Binary Digital Image Processing: A Discrete Approach", Academic Press, 2000.
- [ϕϒ] L. D. Wu, "On the Chain Code of a Line," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, vol. 4, pp. 347-353, 1982.
- [ϕϓ] B. B. Chaudhuri (Ed.), "Digital Document Processing: Major Directions and Recent Advances," Springer-Verlag, 2007.
- [ϕϔ] D. Solomon, "Data Compression, The Complete Reference," Fourth Edition, Springer-Verlag, London, 2007.
- [ϕΔ] G. Navarro and M. Raffinot, "Flexible Pattern Matching in Strings," Cambridge University Press, 2002.
- [ϕϕ] 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.
- [ϕψ] M. Sezgin and B. Sankur, "Survey Over Image Thresholding Techniques and Quantitative Performance Evaluation," *J. Elec. Imaging*, vol. 13, no. 1, pp. 146-165, 2004.
- [ϕλ] O. D. Trier and T. Taxt, "Evaluation of Binarization Methods for Document Images," *IEEE Trans. Pattern Anal. Mach. Intell. PAMI*, vol. 17, no. 3, pp. 312-315, 1995.

- [६९] 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.
- [७०] 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.
- [७१] H. Cheng and C. A. Bouman, “Document Compression Using Rate-Distortion Optimized Segmentation,” Journal of Electronic Imaging, vol. 10, no. 2, pp. 460–474, 2001.
- [७२] H. Cheng and C. A. Bouman, “Multilayer Document Compression Algorithm,” IEEE ICIP, pp. 244-248, Oct. 1999.
- [७३] 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.
- [७४] 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.
- [७५] Z. Fan and M. Xu, “A Simple Segmentation Algorithm for Mixed Raster Contents Image Representation,” Proc. IS&T/SPIE Symposium on Electronic Imaging Science & Technology, Visual Communications and Image Processing, San Jose, CA, vol. 4663, pp. 63-71, 2002.
- [७६] 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.
- [७७] 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.
- [७८] 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 P. Lopresti, Jiangying Zhou, Editors, Proceedings of SPIE, vol. 4307, pp. 158-166, 2001.
- [७९] R. L. de Queiroz, Z. Fan, and T. D. Tran, “Optimizing Block-Thresholding Segmentation for Multilayer Compression of Compound Images,” IEEE Trans. Image Process., vol. 9, no. 9, pp. 1461–1471, 2000.

- [۸۰] D. Mukherjee, N. Memon, and A. Said, "JPEG-Matched MRC Compression of Compound Documents," IEEE Proceedings, vol. 3, pp. 434-437, 2001.
- [۸۱] D. Mukherjee, C. Chrysafis, and A. Said, "JPEG2000-Matched MRC Compression of Compound Documents," IEEE ICIP, vol. 3, pp. 73–76, 2002.
- [۸۲] C. Kuo, A. R. Rao, and G. Thompson, "Color Document Analysis," Proceeding of SPIE, Color Imaging: Device-Independent Color, Color Hardcopy, and Applications VII, vol. 4663, pp.72- 80, 2002.
- [۸۳] A. Zaghetto and R. L. de Queiroz, "Iterative pre- and post-processing for MRC layers of scanned documents", Proc. IEEE Intl. Conf. on Image Processing, ICIP, San Diego, CA, USA, Oct. 2008.
- [۸۴] B. Jahne, "Practical Handbook on Image Processing for Scientific and Technical Applications," 2nd Edition, CRC Press, 2004.
- [۸۵] O. Johnsen, J. Segen, and G. L. Cash, "Coding of Two-Level Pictures by Pattern Matching and Substitution," Bell Syst. Tech. J, vol. 62, no. 8, pp. 2513-2545, 1983.
- [۸۶] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, Introduction to Algorithms, McGraw-Hill, 2nd ed., 2001.
- [۸۷] J. Makhoul, K. Francis, S. Richard, and W. Ralph, "Performance Measures for Information Extraction," Proceedings of DARPA Broadcast News Workshop, Herndon, VA, February 1999.
- [۸۸] R. Baeza-Yates, B. Ribeiro-Neto, "Modern Information Retrieval," New York: ACM Press, Addison-Wesley, 1999.
- [۸۹] M. Cheriet, N. Kharma, C.-L. Liu, C. Y. Suen, Character Recognition Systems: A Guide for Students and Practitioners, John Wiley & Sons, 2007.
- [۹۰] Hu, M. K. "Visual Pattern Recognition by Moment Invariant," IRE Transactions on Information Theory, IT – 8, pp. 179–187, 1962.
- [۹۱] Rumelhart, Hinton & Williams, 1986a; McClelland & Rumelhart, 1988
- [۹۲] H. Grailu, M. Lotfizad, and H. Sadoghi-Yazdi, Farsi and Arabic Document Images Lossy Compression Based on the Mixed Raster Content Model, International Journal on Document Analysis and Recognition (IJ DAR), Volume 12 Issue 4, November 2009
- [۹۳] R. C. Gonzalez , Digital Image Processing. John Wiley and Sons, 2009.

دانشگاه تربیت مدرس، دانشکده برق، رساله دکتری، گرایلو هادی [۹۴]

دکتر مجتبی لطفی زاد، 1390، پردازش دیجیتال تصاویر، R.C.GONZALES and R.E.WOOD [۹۵]
ویراست سوم، وسعید میرقاسمی و امیر مسعود عمویی

[۹۶]DR.HANS-GEORG STARK, WAVELETS AND SIGNAL PROCESSING

هادی ویسی - کبرا مفاخری-سعید باقری، ۱۳۸۸، مبانی شبکه‌های عصبی، لوران فاست [۹۷]

Abstract:

Today, most art, historical, and scientific products are produced on paper. Most libraries round the world are converting their resources into digital format for economical reasons. Storing such huge amount of digital information requires using efficient compression methods.

In this thesis, we aim to propose a MRC-based compression method for text images. Text images are usually decomposed into three layers of foreground, background, and mask. The mask layer is a binary image which determines the textual regions. The foreground layer shows color of textual regions and finally, the background layer includes graphical regions and line drawings.

In thesis we created a text image database which includes about 1449 text images. These images have various types of scripts, qualities, and text-graphic combinations. We proposed and used a gray-scale-text-image binarization method to produce the mask layer which is based on two ideas of combining classifiers (using neural networks) and multi-scale analysis (using filter bank). This proposed method has four advantages including adaptive computing a local threshold, using multi-scale analysis, combining classifiers, and low run time. The proposed method has a run time up to 50% lower than that of the Otsu's method. Also, due to MOS measure values collected using 50 human observers, the mean MOS value of the proposed method is 4.2 and for Otsu is 3.8.

We proposed two methods for compression of grayscale text images produced after employing the proposed text binarization. First method is based on KLT transform and the second one uses a proposed preprocessing and JPEG compression. The first compression method has compression ratio of 2.2 times (in average) relative to the JPEG and 1.6 relative to JPEG2000 in equal SNRs. The second proposed compression method produces better quality for reconstructed image than standard JPEG in equal bits per pixel. Also, in equal quality, this method has compression ratio of 1.5 times relative to the JPEG.

Keywords: Text image compression – Mixed Raster Content (MRC) model – Text separation – Binarization.



Shahrood University of Technology

Faculty of Electrical Engineering

Designing a CODEC System Based on MRC Model for Text Image Compression

Thesis submitted to the Graduate Studies Office in partial fulfillment of the
requirements for the

Degree of Master of Science (M.Sc.)

Aidin Khodashenas

Supervisor:

Dr. Hadi Grailu

Associate Supervisor:

Dr. Hossein Khosravi

January 2013