حاشیه‌نویسی خودکار تصویر با استفاده از ارتباط معنایی بین نواحی مبتنی بر تئوری تصمیم چندشرطی

هنگامه دلجویی[۱]، دانشجوی کارشناسی ارشد، امیرمسعود افتخاری‌مقدم[۲]، استادیار
۱و ۲ - دانشگاه آزاد اسلامی- واحد قزوین- دانشکده مهندسی برق، رایانه و فن‌آوری اطلاعات- قزوین- ایران-
h.deljooi@qiau.ac.ir, eftekhari@qiau.ac.ir

چکیده: امروزه به دلیل وجود شکاف معنایی، حاشیه‌نویسی خودکار تصویر یک رویکرد مهم و چالش برانگیز می‌باشد. در این مقاله مدلی پیشنهاد می‌شود که بافتار ناحیه‌ای و موضوعات بصری را به منظور حاشیه‌نویسی خودکار تصاویر با هم ادغام می‌کند. بافتار ناحیه‌ای، ارتباط بین نواحی موجود در تصویر را در نظر می‌گیرد، در صورتی که موضوعات بصری، یک توزیع سراسری از موضوعات را در کل تصویر فراهم می‌کند. روش‌های حاشیه‌نویسی قبلی ارتباط بین نواحی تصویر را نادیده می‌گرفتند در صورتی که این نواحی دقیقاً بیانگر مفهوم یک تصویر هستند، بنابراین در نظر گرفتن ارتباط بین آن‌ها برای حاشیه‌نویسی تصاویر مناسب است. مدل پیشنهادی، بافتار ناحیه‌ای و موضوعات بصری را از تصویر استخراج کرده و آن‌ها را با استفاده از تئوری تصمیم چندشرطی مبتنی بر روش (Technique for Order Preference by Similarity to the Ideal Solution) TOPSIS ادغام می‌کند. بافتار ناحیه‌ای و موضوعات بصری توسط رویکرد (Probability Latent Semantic Analysis) PLSA روی داده‌های آموزشی، آموزش داده شده‌اند. نتایج آزمایشات بر روی تصاویر Corel 5k گواه برتری کارایی ادغام این ۲ نوع اطلاعات برای حاشیه‌نویسی تصاویر می‌باشد.

واژه‌های کلیدی: حاشیه‌نویسی خودکار تصویر، مدل‌های آماری، بافتار ناحیه‌ای، موضوعات بصری، تئوری تصمیم چندشرطی، PLSA، TOPSIS.

Automatic Image Annotation Using the Semantic Relationship between the Regions Based on the Multi Criteria Decision Making

Hengame Deljooi[1], Amir Masoud Eftekhari Moghaddam[2],
1,2 - Department of Electrical, Computer and IT engineering- Qazvin Branch- Islamic Azad University-

Abstract: In the present time, one of the significant thought-provoking approaches to digital images, which has been adopted because of the existence of the semantic gap, is the automatic image annotation. In this paper, we present a model, which combines regional contexts and visual topics to automatic image annotation. The difference of the regional contexts and visual topics is due to their function and application over an image. Regional contexts represent the associations of different regions with each other while visual topics are responsible for the global distribution of different topics. Regarding the associations of different regions with each other in an image is very useful for image annotation, since these regions form the exact explanation of the image semantics. An important fact, which was not considered in the former image annotation methods. The proposed model extracts regional contexts and visual topics from the image, and incorporates them by MCDM (Multi Criteria Decision Making) approach based on TOPSIS (Technique for Order Preference by Similarity to the Ideal Solution) method. Regional contexts and visual topics are learned by PLSA (Probability Latent Semantic Analysis) approach from the training data. The efficacy and importance of integrating these two kinds of information for the image annotation are apparently demonstrated in the experiments on 5k Corel images.

Keywords: Automatic Image Annotation, Statistical Models, Regional Contexts, Visual Topics, MCDM (Multi Criteria Decision Making), PLSA, TOPSIS.

تاریخ ارسال مقاله: ۹۱/۵/۱۷
تاریخ اصلاح مقاله: ۹۲/۱/۱۸
تاریخ پذیرش مقاله: ۹۲/۲/۲
نام نویسنده مسئول: هنگامه دلجوئی
نشانی نویسنده مسئول: دانشگاه آزاد اسلامی قزوین دانشکده مهندسی برق و فناوری اطلاعات

۱- مقدمه

با شیوع ابزار دیجیتال مانند وب‌کم، دوربین‌های دیجیتال و مواردی دیگر دسترسی کاربران به داده‌های تصویری بسیار افزایش پیدا کرده است. مسأله مهم، جستجو و بازیابی این تعداد وسیع داده‌ها است [۱-۶]. به طور کلی تحقیقات بازیابی تصویر به ۳ دسته تقسیم می‌شود [۱-۴]:

رویکرد اول بازیابی تصویر مبتنی بر متن[۱] است که در آن تصاویر به صورت دستی و به کمک انسان حاشیه‌نویسی می‌شوند و سپس این تصاویر همچون اسناد متنی مورد بازیابی قرار می‌گیرند [۴]. اما به دلیل حجم گسترده تصاویر این کار غیر قابل انجام است، علاوه بر آن، نظر انسان‌ها در حاشیه‌نویسی وارد می‌شود.

رویکرد دوم بازیابی تصویر مبتنی بر محتوا[۲] است که در آن تصاویر به صورت خودکار با استفاده از ویژگی‌های سطح پایینشان شاخص‌گذاری و بازیابی می‌شوند. تحقیقات نشان می‌دهد که بیان پرس‌وجو در این روش معمولاً ارتباط نزدیکی با کاربر ندارد، چرا که قصد کاربر با یک تصویر و یا ویژگی‌های سطح پایین آن به راحتی بیان نمی‌شود، بنابراین اختلافی بین مفاهیم سطح بالای موجود در ذهن انسان و ویژگی‌های سطح پایینی که سامانه[۳] می‌پذیرد به وجود می‌آید که به آن شکاف معنایی[۴] می‌گویند[۷].

در رویکرد سوم سامانه بازیابی معنایی تصویر[۵] مطرح شد که قادر به تشخیص مفاهیم موجود در تصویر می‌باشد. توسعه سامانه بازیابی تصویر باعث ظهور حاشیه‌نویسی خودکار[۶] تصویر شد که در آن‌ها تصاویر به صورت خودکار حاشیه‌نویسی می‌شوند و سپس با استفاده از کلمات متنی انتساب یافته و سامانه بازیابی مبتنی بر متن، مورد بازیابی قرار می‌گیرند. ایده اصلی روش‌های حاشیه‌نویسی خودکار تصاویر که به کاهش شکاف معنایی نیز کمک می‌کنند، یادگیری یک مدل معنایی از مجموعه تصاویر نمونه به صورت خودکار و سپس اعمال این مدل به تصاویر جدید برای انتساب برچسب می‌باشد.

۱-۱- کارهای مرتبط

در سال‌های اخیر الگوریتم‌های زیادی برای حاشیه‌نویسی خودکار تصویر معرفی شده‌اند که می‌توانند به ۴ دسته: مدل‌های مبتنی بر فضای بردار[۷] [۸ و ۹]، روش‌های دسته‌بندی[۸] [۱۰-۱۳]، روش‌های مبتنی بر گراف[۹] [۱۴-۱۹]، و مدل‌های آماری[۱۰] [۲۰-۲۸] تقسیم شوند.

مدل‌های مبتنی بر فضای بردار از روش‌های معروف در بازیابی اطلاعات و به خصوص بازیابی متن هستند. این مدل‌ها تصاویر را به صورت اسناد متنی در نظر می‌گیرند و مؤلفه‌های بصری را که همچون کلمات می‌باشند با استفاده از روش‌های استخراج ویژگی بدست می‌آورند.

روش‌های دسته‌بندی هر کلمه یا مفهوم معنایی را به عنوان یک دسته مستقل در نظر می‌گیرند و هر کلمه را به یک دسته‌بند اختصاص می‌دهند. سپس فرآیند انتساب کلمات به تصاویر را با دسته‌بندی تصاویر به همین دسته‌های از پیش تعیین شده انجام می‌دهند. فرآیند انتساب کلمات به تصاویر به دو صورت تک برچسبی[۱۱] و چند برچسبی[۱۲] بیان شده است. تاکنون مدل‌های ارائه شده در حوزه دسته‌بندی از دسته‌بندها برای حاشیه‌نویسی تصویر به صورت تک برچسبی استفاده می‌کردند، اما در [۱۳] حاشیه‌نویسی تصویر به صورت چند برچسبی با استفاده از دسته‌بندها مطرح شده است.

اخیراً روش‌های مبتنی بر گراف موفقیت‌های زیادی را در حوزه آنالیز تصویر و ویدئو و همچنین حاشیه‌نویسی کسب کرده‌اند. چگونگی ساخت گراف شباهت در مبحث یادگیری گراف بسیار مهم می‌باشد. یک گراف خوب باید فهم عمیقی از ساختار داده را منعکس کرده و در حد امکان به کشف دانش نهفته کمک کند.

هدف مدل‌های آماری یادگیری یک مدل مرتبط برای اتصال کلمات به تصاویر می‌باشد که همبستگی[۱۳] یا احتمال توأم[۱۴] بین کلمات و تصاویر را بازنمایی کند. رویکرد پیشنهادی در این مقاله بر مبنای مدل‌های آماری می‌باشد، به همین دلیل در اینجا معروف‌ترین مدل‌های آماری که برای حاشیه‌نویسی تصویر معرفی شده‌اند، تشریح می‌گردند.

مدل Co-Occurrence [۲۲] که توسط Mori معرفی شد از اولین کارهای موجود در این حوزه می‌باشد که به هم‌رخدادی کلمات با نواحی ایجاد شده در تصویر با استفاده از یک شبکه با قاعده و منظم توجه می‌کند. سپس طرحی توسط Duygulu به نام Machine Translation Model (TM) [۲۳] پیشنهاد شد که در آن فرآیند اتصال کلمات به تصاویر همچون یک مدل ترجمه بود که یک مجموعه از نواحی را (که با خوشه‌بندی[۱۵] نواحی تصویر بدست می‌آید) به مجموعه‌ای از کلمات حاشیه ترجمه می‌کرد. Jeon در مدل Cross-Media Relevance Model (CMRM) [۲۴] از توزیع توأم نواحی و کلمات استفاده می‌کند و مجموعه‌ای از کلمات را به مجموعه‌ای از نواحی مرتبط می‌کند. بعد از آن Lavrenko، Continuous-Space Relevance Model

(CRM) [۲۵] را پیشنهاد کرد که برای بهبود مدل CMRM معرفی شد و در آن هر تصویر به تعدادی نواحی تقسیم می‌شود و هر ناحیه با یک بردار ویژگی پیوسته توصیف می‌شود. در یک مجموعه تصاویر آموزش حاشیه‌نویسی شده، احتمال توأم ویژگی‌های تصویر و کلمات تخمین زده می‌شود، سپس احتمال تعلق نواحی تصویر به یک کلمه پیش‌بینی می‌شود. در مقایسه با CMRM در CRM ویژگی‌های پایدار و پیوسته به طور مستقیم مدل می‌شوند، بنابراین بر خوشه‌بندی تکیه نمی‌کند. Feng مدل دیگری را بنام Multiple Bernoulli Relevance Model (MBRM) [۲۶] برای بهبود CMRM و CRM پیشنهاد کرد که بر مبنای توزیع ضرب برنولی برای تولید کلمات به جای توزیع چندجمله‌ای موجود در CRM به حاشیه‌نویسی می‌پردازد.

دو مشکل چالشی در این مدل‌ها وجود دارد. اول، بیشتر الگوریتم‌های موجود یکی از این دو رویکرد را دارند، یعنی از ویژگی‌های سراسری[۱۶] یا ویژگی‌های ناحیه‌ای[۱۷] به صورت انحصاری استفاده می‌کنند. در رویکرد استفاده از ویژگی‌های سراسری، بردار ویژگی سراسری مانند هیستوگرام رنگ از تصویر استخراج می‌شود. ویژگی سراسری دارای مزیتی در دسته‌بندی صحنه‌های ساده مانند "غروب خورشید"، "کوه"، "ساختمان" و غیره می‌باشد. در رویکرد استفاده از ویژگی ناحیه‌ای، هر تصویر به نواحی متعددی قطعه‌بندی می‌شود و به عنوان مجموعه‌ای از بردارهای ویژگی بصری نمایش داده می‌شود. انگیزه اساسی نمایش تصویر مبتنی بر ناحیه این است که ظاهر بسیاری از اشیاء مانند "گربه"، "ببر" و "هواپیما" معمولاً در قسمت کوچکی از تصویر ظاهر می‌شود. اگر قطعه‌بندی مورد قبولی بدست آید، مثلاً هر شئ بتواند به نواحی واضح و همگنی قطعه‌بندی شود، آنگاه نمایش مبتنی بر ناحیه بسیار با معنی خواهد بود. هر کدام از این دو نمایش ویژگی، انواع مختلفی از اطلاعات را فراهم می‌کنند و هر کدام از آن‌ها مزایای خود را در دسته‌بندی به طبقه‌های مشخص دارند. بنابراین ادغام این دو نوع ویژگی با تنوع زیادی از طبقه‌ها برای حاشیه‌نویسی مفیدتر است.

دوم، رویکردهای مرسوم هر لغت را به صورت جدا و بدون توجه به ارتباط بافتار متنی[۱۸] آن در نظر می‌گیرند. در واقع حاشیه‌نویسی یک تصویر با یک لغت مستقل از حاشیه‌نویسی با لغت دیگر روی همان تصویر است. در نتیجه ارتباط بافتار متنی بین لغت‌های حاشیه‌نویسی شده نادیده گرفته می‌شود. بافتار متنی بین لغات، ارتباط هم‌رخدادی[۱۹] را بیان می‌کند. این نوع اطلاعات متنی به صورت واقعی در حاشیه‌نویسی دستی جاسازی شده است، به طوری که بشر معمولاً یک تصویر را با مجموعه‌ای از لغات منسجم و با معنی و دارای ارتباط منطقی به صورت کلی حاشیه‌نویسی می‌کند. برای مثال در تصاویر "خارج از خانه"، اگر کلمه "ابر" حاشیه‌نویسی شده باشد، در حاشیه‌نویسی قبلی احتمال وجود کلمه "آسمان" بیشتر از حاشیه‌نویسی کلمه "خیابان" است (بدون آزمایش محتوای بصری و جزئی تصویر). بنابراین استخراج بافتار متنی از داده‌های آموزشی برای حاشیه‌نویسی تصاویر مفید است.

جدیدترین مدلی که این دو مشکل را برطرف کرده است Extended CMRM [۲۷] می‌باشد که ویژگی‌های سراسری و ناحیه‌ای را با بافتار متنی برای حاشیه‌نویسی ادغام می‌کند.

۱-۲- طرح پیشنهادی

مشکل مهمی که در تمام روش‌های بالا وجود دارد این است که تمام الگوریتم‌های ذکر شده از توزیع مستقیم نواحی برای حاشیه‌نویسی تصویر استفاده می‌کنند (نواحی از طریق قطعه‌بندی تصویر مبتنی بر ناحیه یا ماهیت اشیاء بدست می‌آیند) در صورتی که توجه به ارتباط بین این نواحی که هر کدام بیانگر یک کلمه یا مفهوم هستند در بهبود کلمات حاشیه نهایی به ما کمک می‌کنند. بنابراین بافتار ناحیه‌ای[۲۰] که نشان دهنده ارتباط بین نواحی می‌باشد را بدست آورده و به جای استفاده از توزیع مستقیم نواحی از توزیع موضوعی بین نواحی در حاشیه‌نویسی استفاده می‌شود.

برای حل کردن مشکلی که در قسمت بالا اشاره شد یک رویکرد جدید برای حاشیه‌نویسی خودکار تصاویر معرفی شده است. CMRM معمولی فقط ویژگی‌های ناحیه‌ای را برای توصیف تصاویر به کار می‌برد، اما CMRM توسعه‌یافته، ویژگی‌های سراسری و ناحیه‌ای را با بافتار متنی ادغام می‌کرد و با تخمین احتمال توأم، برای حاشیه‌نویسی تصویر استفاده می‌کرد. در رویکرد که پیشنهادی می‌گردد علاوه بر اینکه ویژگی‌های سراسری به عنوان بردار توزیع موضوعات بصری[۲۱] توصیف می‌شود، ارتباط بین نواحی بدست آمده از فرآیند قطعه‌بندی را هم در نظر گرفته و بافتار ناحیه‌ای به صورت توزیع موضوعات بین نواحی بدست آورده است. بافتار ناحیه‌ای و موضوعات بصری توسط PLSA (Probability Latent Semantic Analysis) [۲۹] روی داده‌های آموزشی آموزش داده شده‌اند. سپس برای ادغام این دو نوع اطلاعات از الگوریتم TOPSIS (Technique for Order Preference by Similarity to the Ideal Solution) [۳۱ و

۳۰] موجود در تئوری تصمیم چند شرطی استفاده می‌شود. به طور کلی رویکرد پیشنهادی مقاله شامل این سه مرحله می‌باشد:

۱. مدل کردن بافتار ناحیه‌ای یا ارتباط بین نواحی موجود در تصویر به صورت توزیع موضوعات بین نواحی به منظور ایجاد کلمات حاشیه مرتبط با مفهوم نواحی تصویر،
۲. استفاده از ویژگی‌های سراسری به صورت بردار توزیع موضوعات بصری به منظور در نظر گرفتن مفهوم کلی تصویر،
۳. ادغام دو نوع اطلاعات بدست آمده از تصویر با استفاده از الگوریتم TOPSIS موجود در تئوری تصمیم چند شرطی.

چارچوب رویکرد پیشنهادی در شکل (۱) نشان داده شده است. تصاویر با استفاده از نواحی موجود در آن‌ها و مجموعه‌ای از تکه‌ها نمایش داده شده‌اند. مدل پیشنهادی با استفاده از دو نوع اطلاعات بافتار ناحیه‌ای $B(I)$ و موضوعات بصری $H(I)$ بر روی داده‌های آموزشی، آموزش داده می‌شود.

ادامه این مقاله به این صورت سازماندهی شده است: بخش ۲ مدل ECMRM را معرفی می‌کند. در بخش ۳ رویکرد پیشنهادی تشریح می‌شود. بخش ۴ نتایج تجربی را مطرح می‌کند و نتیجه‌گیری هم در بخش ۵ بیان می‌شود.

۲- Extended Cross Media Relevance Model (ECMRM)

CMRM مرسوم [۲۴] فقط به یک نوع اطلاعات در تصویر مثلاً مجموعه‌ای از نواحی توجه می‌کند، اما در مورد تصاویری که نمی‌توان آن‌ها را به صورت مجموعه‌ای از نواحی درآورد و نمایش داد باید به نوع دیگری از اطلاعات هم توجه شود. بنابراین مدل ECMRM [۲۵] پیشنهاد می‌شود که از موضوعات بصری هم به عنوان نوع دیگری از اطلاعات استفاده شود. این اطلاعات جدید تصویر با نمایش نواحی بصری ادغام می‌شود. علاوه بر این مدل CMRM کلمات را به صورت تک‌تک و بدون توجه به توزیع توأم کلمات مختلف حاشیه‌نویسی می‌کند. برای حل کردن این مشکل بافتار متنی برای مدل کردن توزیع توأم کلمات مختلف پیشنهاد می‌شود. برای حاشیه‌نویسی یک تصویر با چندین لغت ابتدا تصویر با بافتار متنی حاشیه‌نویسی می‌شود و سپس توزیع کلمات از روی بافتار متنی ایجاد می‌شوند. در یک جمله ساده بافتار متنی یک ارتباط هم‌رخدادی بین کلمات مختلف است. یادگیری بافتار متنی بر مبنای رویکرد PLSA [۲۹] می‌باشد که توسط Hofmann ابداع شده است، PLSA برای یادگیری خودکار موضوعات از اسناد متنی پیشنهاد شده است.

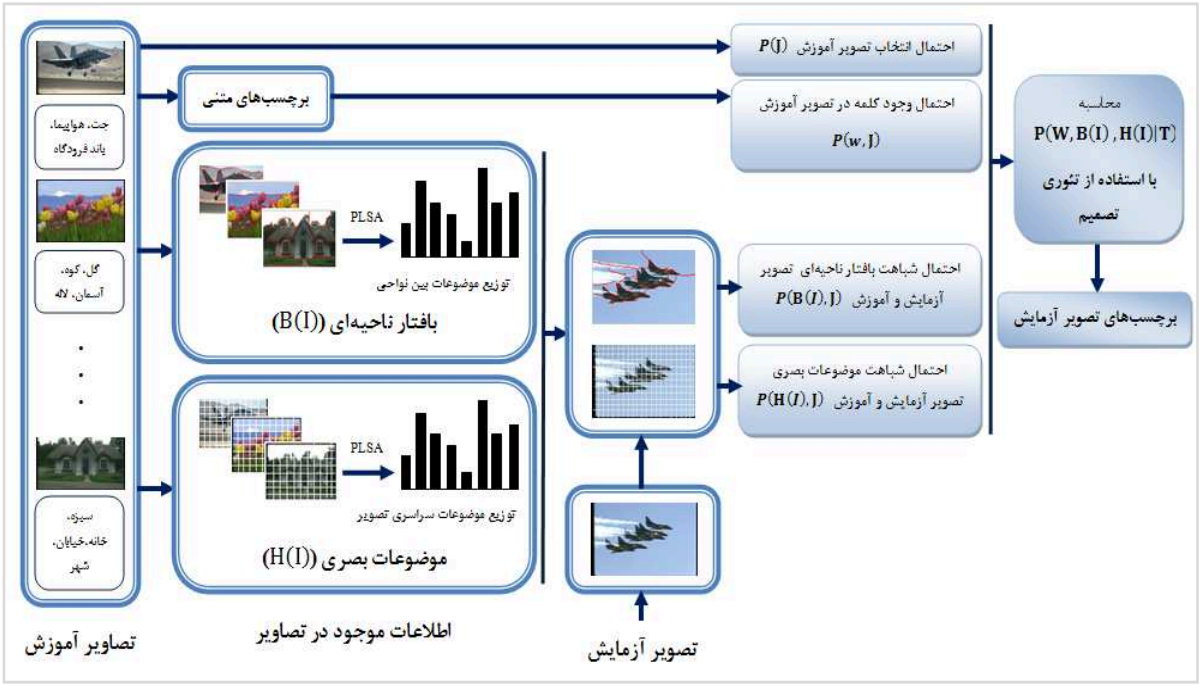
مشابه یادگیری موضوعات از اسناد متنی، می‌توان موضوعات بصری را هم از مجموعه تصاویر یاد گرفت. نکته اصلی، نمایش تصویر به صورت مجموعه‌ای از لغات مشابه نمایش برداری سند متنی[۳۲] است. تصاویر به صورت شبکه انبوه[۲۲] و منظمی افرازبندی[۲۳] می‌شوند و این تکه‌ها[۲۴] به صورت مجموعه‌ای نامرتب از تصاویر در نظر گرفته می‌شوند. سپس توصیفگر SIFT [۳۳] به طول ۱۲۸ از آن تکه‌ها استخراج شده و در نهایت با خوشه‌بندی [۳۴] کمی‌سازی می‌شوند که برای بازشناسی اشیاء بسیار مؤثر است [۳۵] . مرکز خوشه‌ها کلمات بصری[۲۵] نامیده می‌شود. سپس با منسوب کردن برچسب‌های کلمات بصری به هر تکه تصویر، می‌توان تصویر را به مجموعه‌ای از کلمات بصری انتقال داد. اکنون مجموعه کلمات بصری داده شده است، سپس می‌توان PLSA را برای یادگیری مجموعه موضوعات بصری اعمال کرد، هر کدام از آن‌ها با توزیع چندجمله‌ای کلمات بصری مشخص می‌شود.

با داشتن بافتار متنی C، نواحی بصری $R=(b_1,b_2,...,b_m)$ که به کمک قطعه‌بندی [۳۴] بدست آمده و توزیع موضوع بصری $H(I)$، روش حاشیه‌نویسی ECMRM به این صورت می‌باشد:

$$P(C,R,H(I))=\sum_{J\in\tau}P(J)P(C,b_1,...,b_m,H(I)|J) \qquad (۱)$$

در مقایسه مدل ECMRM با CMRM، دو نکته متفاوت برای توضیح دادن وجود دارد: اول این که CMRM اصلی، تصاویر را با استفاده از ویژگی‌های ناحیه‌ای $R=\{b_1,b_2,...,b_m\}$ حاشیه‌نویسی می‌کند. اگرچه مدل توسعه‌یافته در معادله (۱)، هم از ویژگی‌های ناحیه‌ای R و هم از ویژگی‌های سراسری $H(I)$ استفاده می‌کند که توزیع سراسری موضوعات بصری را در تصویر I نشان می‌دهد. در واقع ویژگی‌های سراسری و ناحیه‌ای را با هم ادغام می‌کند. دوم این که، مدل CMRM، احتمال یک کلمه W را به صورت مستقیم پیش‌بینی می‌کند اما مدل ECMRM احتمال بافتار متنی C را پیش‌بینی می‌کند. این نشان می‌دهد که مدل جدید استقلال بین کلمات داده شده در تصویر را در نظر نمی‌گیرد.

استقلال دوطرفه بین بافتار متنی C و نواحی تصویر و توزیع موضوعات بصری در نظر گرفته می‌شود، بنابراین $P(C,b_1,...,b_m,H(I)|J)$ به این صورت ساده می‌شود:



شکل (۱): رویکرد پیشنهادی با استفاده از بافتار ناحیه‌ای و موضوعات بصری برای حاشیه‌نویسی خودکار تصویر.

$$P\left(C,b_1,...,b_m,H\left(I\right)\middle|J\right)= P\left(C\middle|J\right)P\left(H\left(I\right)\middle|J\right)\prod_{i=1}^{m}P\left(b_i\middle|J\right) \quad (۲)$$

$P\left(b\middle|J\right)$ همانند مدل CMRM، تخمین زده می‌شود. $P\left(C\middle|I\right)$ بعد از یادگیری بافتار متنی بر روی حاشیه‌های دستی در دسترس است. $P\left(H\left(I\right)\middle|J\right)$ هم به عنوان انشعاب Kullback-Leibler [۳۷] بین توزیع موضوع بصری تصویر I و J است، به این صورت:

$$P\left(H\left(I\right)\middle|J\right)=D_{kl}\left(H\left(I\right)\middle|H\left(J\right)\right)=\sum_{i=1}^{Q}P\left(q_i\middle|I\right)\log\frac{P\left(q_i\middle|I\right)}{P\left(q_i\middle|J\right)} \quad (۳)$$

که در آن D_{kl}، انشعاب Kullback-Leibler بین دو توزیع است. از تئوری Bayesian می‌دانیم که:

$$P\left(C\middle|I\right)=\frac{P\left(C,I\right)}{P\left(I\right)}=\frac{P\left(C,b_1,...,b_m,H\left(I\right)\right)}{P\left(I\right)} \quad (۴)$$

بنابراین نرمال‌سازی روی $P\left(C,b_1,...,b_m,H\left(I\right)\right)$، توزیع شرطی $P\left(C\middle|I\right)$ را ارائه می‌دهد. توزیع شرطی کلمات $P\left(W_j\middle|I\right)$ از I به صورت زیر با ترکیب کردن توزیع کلمات از همه بافتارهای متنی بدست می‌آید:

$$P\left(W_j\middle|I\right)=\sum_{i}^{S}P\left(W_j\middle|C_i\right)P\left(C_i\middle|I\right) \quad (۵)$$

۳- رویکرد پیشنهادی

نکته قابل ذکر این است که در مدل ECMRM به ارتباط بین نواحی توجهی نمی‌شود، همان‌گونه که مشخص می‌باشد هر کدام از این نواحی بیانگر یک مفهوم هستند و ارتباط معنایی بین این نواحی بسیار دقیق‌تر از ارتباط معنایی بین کلمات (بافتار متنی) است که توسط انسان به تصاویر متصل شده‌اند. چرا که این نواحی دقیقاً محتوای یک تصویر را نشان می‌دهند و در نهایت کلمات تولید شده به عنوان حاشیه، مرتبط با محتوای تصویر می‌باشند.

۳-۱- یادگیری بافتار ناحیه‌ای از تصاویر

مشابه یادگیری موضوعات از سند متنی در مدل ECMRM، می‌توان بافتار ناحیه‌ای را هم از مجموعه تصاویر یاد گرفت. در ابتدا طبق الگوریتم CMRM معمولی نواحی موجود در تصویر بر اساس قطعه‌بندی مبتنی بر ناحیه و بر اساس ماهیت اشیاء بدست می‌آیند لذا باید تصویر به صورت مجموعه‌ای از لغات نمایش داد[۳۶] در واقع یک تصویر بیان کننده یک سند می‌باشد و می‌توان اشیایی که در تصویر وجود دارند را به عنوان کلمات موجود در این سند فرض نمود (هر ناحیه به عنوان یک کلمه در نظر گرفته می‌شود). سپس از طریق اعمال الگوریتم PLSA می‌توان رابطه بین این نواحی بدست آورد. موضوعات در واقع بیان‌کننده ارتباط بین نواحی هستند و

می‌توان در حاشیه‌نویسی به جای استفاده از توزیع مستقیم نواحی از توزیع موضوعات مربوط به آن نواحی استفاده نمود. این موضوعات که ارتباط بین نواحی را بیان می‌کنند بافتار ناحیه‌ای نامیده می‌شوند.

الگوریتم PLSA [۲۹] به این صورت عمل می‌کند، فرض بر این است که یک مجموعه‌ای از اسناد متنی موجود می‌باشد، $D=\{d_1,d_2,...,d_n\}$ ، که هر کدام به این صورت می‌باشند:

$$d_i=\left[n(d_i,w_1),n(d_i,w_2),...,n(d_i,w_m)\right] \qquad (۶)$$

که $n(d_i,w_j)$ تعداد رویدادهای کلمه w_j در سند d_i است و m سایز واژگان است. PLSA فرض می‌کند که هر لغت در سند با موضوع پنهان و جزئی Z_k تولید می‌شود که $Z_k \in I$ است و I، واژگان و لغات موضوع پنهان است وقتی Z_k متغیر پنهان است، احتمال شرطی کلمه w_j در سند d_i، حاشیه‌سازی در اطراف موضوع است. به این صورت:

$$P(w_j|d_i)=\sum_{k}^{K}P(w_j|Z_k,d_i)P(Z_k|d_i) \qquad (۷)$$

که K تعداد موضوعات پنهان است، $P(w_j|Z_k,d_i)$ احتمال شرطی کلمه w_j در موضوع داده شده Z_k و سند d_i است. $P(Z_k|d_i)$ احتمال شرطی موضوع Z_k در d_i داده شده است. علاوه بر آن در PLSA فرض براین است که احتمال شرطی تولید یک کلمه با موضوع مشخص مستقل از سند است. به عبارت دیگر:

$$P(w_j|Z_k,d_i)=P(w_j|Z_k) \qquad (۸)$$

بنابراین معادله (۷) به این صورت ساده می‌شود:

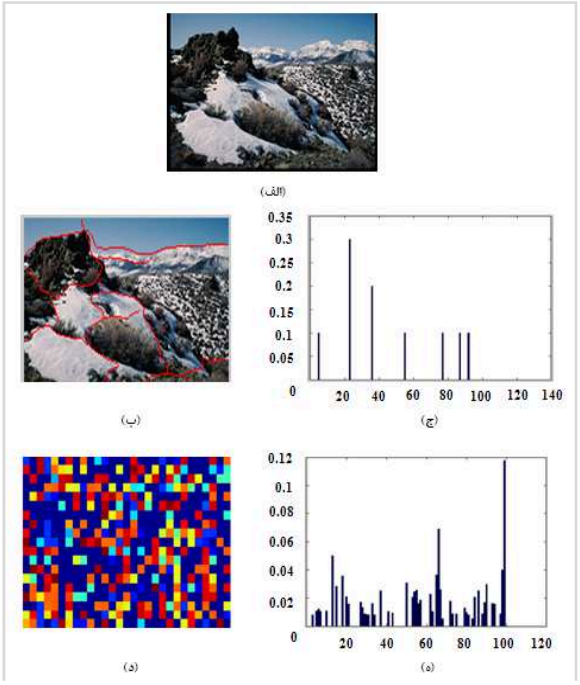
$$P(w_j|d_i)=\sum_{k}^{K}P(w_j|Z_k)P(Z_k|d_i) \qquad (۹)$$

پارامترهای $P(w_j|Z_k)$ و $P(Z_k|d_i)$ با الگوریتم EM یاد گرفته می‌شوند [۳۶].

۳-۲- یادگیری موضوعات بصری از تصاویر

مشابه یادگیری موضوعات از سند متنی، می‌توان موضوعات بصری را هم از مجموعه تصاویر یاد گرفت. نکته اصلی، نمایش تصویر به صورت مجموعه‌ای از لغات مشابه نمایش برداری سند متنی است[۳۲]، در این مقاله تصاویر به صورت شبکه انبوه و منظمی افرازبندی شده و این تکه‌ها به صورت مجموعه‌ای نامرتب از تصاویر در نظر گرفته شده است. سپس توصیفگر Gabor [۳۹-۴۱] را به تکه‌های تصویر اعمال کرده، استخراج ویژگی انجام شده و تکه‌های تصاویر با خوشه‌بندی [۳۴] کمی‌سازی می‌شود که برای بازشناسی اشیاء بسیار مؤثر است [۳۵] توصیفگر Gabor هیستوگرام جهت [۴۲ و ۴۳] را استخراج کرده و بهتر از گرادیان ساده یا SIFT عمل می‌کند. به مرکز خوشه‌ها کلمات بصری می‌گویند. سپس می‌توان با منسوب کردن برچسب‌های کلمات بصری به هر تکه تصویر، تصویر را به مجموعه‌ای از کلمات بصری انتقال داد. در نهایت می‌توان PLSA را برای یادگیری مجموعه موضوعات بصری اعمال کرد هر کدام از آن‌ها با توزیع چندجمله‌ای کلمات بصری مشخص می‌شود.

بین نواحی تصویر که با قطعه‌بندی بدست می‌آیند و تکه‌های بدست آمده از افرازبندی تفاوت وجود دارد. تکه‌های تصاویر که در یک موضوع گروه‌بندی شده‌اند خاصیت انباشتگی مکانی و سازگاری بصری ندارند. در صورتی که در قطعه‌بندی، پیکسل‌ها با ویژگی‌های بصری و اطلاعات مکانی خود گروه‌بندی می‌شوند. شکل (۲) نواحی تصویر، تکه‌های تصویر و توزیع بافتار ناحیه‌ای و موضوعات بصری را در تصویر نشان می‌دهد. در واقع موضوعات بصری و بافتار ناحیه‌ای (موضوعات بین نواحی تصویر که با قطعه‌بندی بدست می‌آیند) بر روی جنبه‌های مختلف تصویر تمرکز می‌کنند مکمل هم هستند و انتظار می‌رود که ادغام آن‌ها کارایی بهتری بدست آورد.



شکل (۲): نمایش دو نوع اطلاعات: (الف) تصویر اصلی، (ب) نواحی تصویر بدست آمده با قطعه‌بندی تصویر، (ج) توزیع بافتار ناحیه‌ای مربوط به تصویر اصلی، (د) تکه‌های تصویر بدست آمده از افرازبندی که تکه‌های یک‌رنگ به یک موضوع تعلق دارند، (ه) توزیع موضوعات بصری مربوط به تصویر اصلی.

۳-۳- ادغام بافتار ناحیه‌ای و موضوعات بصری تصاویر

رویکرد پیشنهادی تصویر تست I را با ادغام $B(I)$ و با در نظر گرفتن کلمه داده شده (w)، حاشیه‌نویسی می‌کند:

$$P(W,B(I),H(I))=\sum_{J\in\tau}P(J)P(W,B(I),H(I)|J) \quad (۱۰)$$

با مقایسه معادلات (۱) و (۱۰) مشاهده می‌شود: CMRM توسعه‌یافته در معادله (۱)، تصاویر را با استفاده از ویژگی‌های ناحیه‌ای و به صورت مجموعه‌ای از نواحی $R=\{b_1,b_2,...,b_m\}$ حاشیه‌نویسی می‌کند، به عبارت دیگر از توزیع مستقیم نواحی استفاده می‌کند. در صورتی که در معادله (۱۰)، به ارتباط بین این نواحی هم توجه شده و از توزیع موضوعات بین نواحی تحت عنوان بافتار ناحیه‌ای $B(I)$ استفاده می‌شود. با توجه به اینکه ارتباط معنایی بین نواحی یک تصویر بسیار دقیق‌تر از ارتباط بین کلمات است، بافتار متنی موجود در معادله (۱) حذف شده و از بافتار ناحیه‌ای و موضوعات بصری در کنار هم استفاده می‌شود. چرا که این دو نوع اطلاعات، هر کدام کلمات متفاوتی را استخراج می‌کنند بنابراین ادغام آن‌ها برای ایجاد کلمات مرتبط و سازگار با محتوای تصویر مناسب است. در حالت کلی وقتی که نواحی موجود در یک تصویر در دسترس باشند، توزیع کلمات بصری نمی‌تواند مستقل از نواحی باشد، بنابراین توزیع نواحی و توزیع کلمات بصری به هم وابسته هستند. از طرفی با در نظر گرفتن معادله (۱۰) می‌توان نتیجه گرفت که:

$$p(w|I)=\sum_{J\in\tau}p(w|J)\alpha_j \quad (۱۱)$$

ضریب α_jها میزان شباهت تصویر آزمایش I را با تصاویر آموزش محاسبه می‌کند که در رویکرد پیشنهادی با توجه به معادله (۱۰)، بافتار ناحیه‌ای و موضوعات بصری می‌باشند. در واقع میزان سهم توزیع کلمات تصاویر آموزش J را در تشکیل توزیع کلمات تصویر آزمایش I معین می‌کند. این مسأله به آسانی قابل درک است که هرچه یک تصویر آموزش به تصویر آزمایش I بیشتر شباهت داشته باشد توزیع کلمات آن دو نیز شبیه‌تر و به عبارت دیگر میزان سهم توزیع کلمات تصویر آموزش و یا ضریب α_j می‌بایست بیشتر باشد. با توجه به وابستگی $B(I)$ و $H(I)$ و به منظور ساده‌سازی معادله (۱۰) و محاسبه ضریب α_j یا میزان شباهت بین تصویر آزمایش و تصاویر آموزش، از رویکرد تئوری تصمیم چندشرطی استفاده می‌شود[۳۰]. تئوری تصمیم چند شرطی[۲۶] شامل ایجاد و ساخت بهترین تصمیم از میان مجموعه‌ای از پیشنهادها[۲۷] یا حالت‌ها[۲۸] می‌باشد به طوری که این پیشنهادها با استفاده از چندین شرط[۲۹] یا ویژگی[۳۰] مورد ارزیابی قرار بگیرند. این روش در حوزه‌های مختلفی از مهندسی ساختار از جمله اختصاص منابع، رتبه‌بندی الگوهای ترتیبی و اختصاص امکانات خاص به کار برده شده است. همچنین در سرمایه‌گزاری‌های مالی برای سامانه تولید هم مورد استفاده قرار گرفته است.

در بین الگوریتم‌های موجود در تئوری تصمیم چند شرطی، از الگوریتم TOPSIS [۳۰ و ۳۱] استفاده شده است. از جمله قابلیت‌های این الگوریتم، امکان تعیین هر نوع شرط، وضوح نتایج آن و کاهش پیچیدگی‌های مربوط به پارامترها می‌باشد. همچنین این الگوریتم برای محاسبه میزان شباهت به کار می‌رود. در این الگوریتم دو مفهوم وجود دارد، پیشنهاد ایده‌آل[۳۱] پیشنهادی است که بهترین سطح و بیشترین میزان را برای همه شروط کسب کرده باشد و پیشنهاد ایده‌آل منفی[۳۲] پیشنهادی است که بدترین سطح و ارزش را برای همه شروط دارا باشد. الگوریتم TOPSIS پیشنهادی را انتخاب می‌کند که نزدیک‌ترین مقدار به پیشنهاد ایده‌آل و دورترین مقدار از پیشنهاد ایده‌آل منفی باشد.

ورودی‌های الگوریتم TOPSIS به این صورت می‌باشد:

- فرض بر این است که m پیشنهاد یا حالت و n شرط یا ویژگی در اختیار است و مقدار هر حالت با توجه به هر شرط در اختیار باشد. در رویکرد ارائه شده پیشنهادها همان تصاویر آموزش و شروط مورد نظر، بافتار ناحیه‌ای و موضوعات بصری هستند. در واقع برای تعیین شرط، باید توابعی را تعریف کرد که میزان شباهت را محاسبه می‌کنند و در این حالت به صورت تجربی توابع زیر در نظر گرفته می‌شود:

$$\left[p(B(I)|J)^2, p(B(I)|J), p(B(I)|J)\,p(H(I)|J)\right] \quad (۱۲)$$

- با تشکیل ماتریس m×n مقدار x_{ij} به هر سلول آن با توجه به i پیشنهاد یا حالت و j شرط اختصاص داده می‌شود.
- J را مجموعه‌ای از شرط‌ها یا ویژگی‌های مؤثر در نظر گرفته می‌شود (مقدار بیشتر، بهتر است).
- J' را مجموعه‌ای از شرط‌ها یا ویژگی‌های منفی در نظر گرفته می‌شود (مقدار کمتر، بهتر است).

مرحله ۱: ماتریس تصمیم نرمالیزه شده ایجاد شده است.

- این مرحله ابعاد ویژگی‌های متفاوت را یکی می‌کند تا امکان مقایسه با استفاده از شرط‌ها فراهم شود.
- داده‌ها صورت زیر نرمالیزه می‌شود:

$$r_{ij} = \frac{x_{ij}}{\sqrt{x_{ij}^2}} \quad \text{for i=1,...,m:j=1,...,n} \qquad (۱۳)$$

مرحله ۲: ماتریس تصمیم نرمالیزه شده وزن‌دار بدست می‌آید.

- فرض بر این است که مجموعه‌ای از وزن‌ها w_j برای هر شرط j=1,...,n در اختیار می‌باشد. این وزن‌ها به صورت تجربی و با توجه به اهمیت هر شرط انتخاب می‌گردد. در رویکرد پیشنهادی با توجه به سه معیار انتخاب شده، نیاز به تعیین سه وزن به صورت $\left[a*\frac{6}{7} \quad \frac{a}{7} \quad 1\text{-}a\right]$ برای آن‌ها هست.
- هر ستون از ماتریس تصمیم نرمالیزه شده در وزن آن ضرب می‌شود.
- هر مؤلفه ماتریس جدید به این صورت محاسبه می‌شود:

$$v_{ij} = w_{ij} \cdot r_{ij} \qquad (۱۴)$$

مرحله ۳: برای هر شرط، بیشترین مقدار و کم‌ترین مقدار جواب ایده‌آل[۳۳] و جواب ایده‌آل منفی[۳۴] در نظر گرفته می‌شود.

- جواب ایده‌آل:

$$A^* = \left\{v_1^*,...,v_n^*\right\}$$
$$\text{where } v_j^* = \left\{\max\left(v_{ij}\right) \text{ if j}\in \text{J:}\min\left(v_{ij}\right) \text{ if j}\in \text{J}'\right\} \qquad (۱۵)$$

- جواب ایده‌آل منفی:

$$A' = \left\{v_1',...,v_n'\right\}$$
$$\text{where } v' = \left\{\min\left(v_{ij}\right) \text{ if j}\in \text{J:}\max\left(v_{ij}\right) \text{ if j}\in \text{J}'\right\} \qquad (۱۶)$$

مرحله ۴: معیار جدایی[۳۵] برای هر حالت (تصاویر آموزش) محاسبه می‌گردد:

- معیار جدایی از انتخاب ایده‌آل برابر است با:

$$S_i^* = \left[\sum_j \left(v_j^* - v_{ij}\right)^2\right]^{\frac{1}{2}} \quad \text{i=1,...,m} \qquad (۱۷)$$

- معیار جدایی از انتخاب ایده‌آل منفی برابر است با:

$$S_i' = \left[\sum_j \left(v_j' - v_{ij}\right)^2\right]^{\frac{1}{2}} \quad \text{i=1,...,m} \qquad (۱۸)$$

مرحله ۵: C_i^* که همان α_j است محاسبه می‌شود:

$$C_i^* = \frac{S_i'}{\left(S_i^* + S_i'\right)} \quad 0 < C_i^* < 1 \qquad (۱۹)$$

C_i^*ای انتخاب می‌شود که به ۱ نزدیکتر باشد.

با توجه به نوع شرط‌هایی که تعریف شد، نیاز به محاسبه $P\left(B(I)\middle|J\right)$ و $P\left(H(I)\middle|J\right)$ ضروری است. این دو مؤلفه به صورت انشعاب Kullback-Leibler [۳۷] بین توزیع دو تصویر I و J محاسبه می‌شوند، $P\left(H(I)\middle|J\right)$ طبق معادله (۳) در دسترس است:

$$P\left(B(I)\middle|J\right) = D_{kl}\left(B(I)\middle|B(J)\right) = \sum_{i=1}^{Q} P\left(q_i\middle|I\right)\log\frac{P\left(q_i\middle|I\right)}{P\left(q_i\middle|J\right)} \qquad (۲۰)$$

۴- نتایج تجربی

در این قسمت جزئیات مربوط به مجموعه داده و نتایج آزمایش‌ها تشریح می‌شود. مجموعه داده و معیارهای ارزیابی به ترتیب در بخش ۴-۱ و ۴-۲ و نتایج آزمایشات و مقایسه با سایر روش‌های حاشیه‌نویسی هم در بخش ۴-۳ تشریح می‌گردد.

۴-۱- مجموعه داده

برای ارزیابی کارایی رویکرد پیشنهادی روی مجموعه ۵۰۰۰ تایی Corel آزمایش انجام شد. این مجموعه داده شامل ۵۰ موضوع اصلی می‌باشد که هر موضوع از ۱۰۰ تصویر تشکیل شده است. این مجموعه داده در مرحله اول توسط Duygulu [۲۳] جمع‌آوری شد و سپس توسط محققان دیگر گسترش یافت به همین دلیل به عنوان یک مجموعه داده استاندارد در تحقیقات حاشیه‌نویسی تصویر می‌باشد. هر تصویر با ۱ الی ۵ کلمه به صورت دستی حاشیه‌نویسی شده است و تعداد کلمات متفاوت در دیکشنری این مجموعه داده ۳۷۴ کلمه است.

در آزمایش‌ها، کل مجموعه داده را به دو دسته مجموعه آموزش و مجموعه آزمایش تقسیم کرده که شامل ۴۵۰۰ تصویر آموزشی و ۵۰۰ تصویر آزمایش می‌باشد. برای قطعه‌بندی از الگوریتم JSEG [۳۶] استفاده کرده و تعداد نواحی موجود در هر تصویر ۱ الی ۱۱ ناحیه می‌باشد، که به طور میانگین ۵ ناحیه در هر تصویر وجود دارد. نواحی موجود در هر تصویر با یک بردار ویژگی به طول ۳۶ بازنمایی می‌شوند. همچنین برای ایجاد شبکه انبوه تکه‌های تصاویر با پیکسل‌های ۱۳×۱۳ بدون هم‌پوشانی، در نظر گرفته شده‌اند. تعداد میانگین تکه‌های تصویر در هر تصویر در حدود ۵۵۰ است. نواحی تصویر به ۵۰۰ نوع ناحیه تصویر خوشه‌بندی می‌شوند، به طور مشابه توصیفگر Gabor از تکه‌های تصویر هم به ۵۰۰ مرکز خوشه‌بندی می‌شود آزمایشات با تعداد ۱۰۰ از موضوعات بصری V و تعداد ۸۰ از بافتار متنی T و تعداد ۱۲۰ از بافتار ناحیه‌ای R گرفته است.

۴-۲- معیارهای ارزیابی

برای هر روش، ۵ کلمه با بیشترین احتمال به عنوان کلمات حاشیه نهایی در نظر گرفته شده است. همچنین کارایی حاشیه‌نویسی با میانگین دقت و فراخوانی و معیار F_1 برای کل تصاویر آزمایش، ارزیابی شد. معیار F_1 با مقدارهای دقت و فراخوانی به این صورت بدست می‌آیند:

$$\Pr ecision(W) = \frac{r}{n} \quad (۲۱)$$

$$\mathrm{Re}call(W) = \frac{r}{N} \quad (۲۲)$$

$$F_1 = \frac{2 \times \Pr ecision \times \mathrm{Re}call}{(\Pr ecision + \mathrm{Re}call)} \quad (۲۳)$$

r بیانگر تعداد تصاویری است که با کلمه w به درستی حاشیه‌نویسی شده‌اند، n تعداد تصاویری است که با کلمه w به صورت خودکار حاشیه‌نویسی شده‌اند و N تعداد تصاویری است که با w به صورت دستی حاشیه‌نویسی شده‌اند.

همچنین تعداد کلمات با فراخوانی بزرگ‌تر از صفر نیز در نظر گرفته می‌شود، در واقع این معیار تعداد کلماتی را نشان می‌دهد که سیستم حاشیه‌نویسی خودکار تصویر به صورت مؤثر و کارا یاد گرفته است.

۴-۳- ارزیابی کارایی و مقایسه نتایج

در این بخش نتایج رویکرد پیشنهادی در مجموعه داده ذکر شده مورد بررسی قرار می‌گیرد. آزمایشات به ۲ قسمت تقسیم می‌شود، ابتدا تأثیر بافتار ناحیه‌ای در حاشیه‌نویسی تصویر مورد ارزیابی قرار می‌گیرد و سپس ادغام آن با موضوعات بصری به کمک تئوری تصمیم چندشرطی سنجیده می‌شود. در نهایت مقایسه‌ی کلی بین رویکرد پیشنهادی و سایر روش‌های آماری انجام می‌شود.

۴-۳-۱- آنالیز تأثیر بافتار ناحیه‌ای

در اولین آزمایش، تأثیر بافتار ناحیه‌ای مورد ارزیابی قرار می‌گیرد. نتایج مقایسات کلی در جدول (۱) نشان داده شده است، با مقایسه نتایج در جدول، بهبود مدل ECMRM [۲۷] نسبت به CMRM اصلی [۲۴] که ویژگی‌های سراسری و ناحیه‌ای و بافتار متنی را با هم ادغام می‌کند به وضوح مشخص است و مقدار F_1 را از ۰/۱۱۱۵به ۰/۳۲۷۰ می‌رساند. اما در رویکرد پیشنهادی با استفاده از بافتار ناحیه‌ای مقدار F_1 را به ۰/۴۲۱۴ بهبود داده است. همچنین بهبود مقادیر دقت، فراخوانی و تعداد کلمات با فراخوانی بزرگ‌تر از صفر دیده می‌شود. به صورت خلاصه استفاده از بافتار ناحیه‌ای (نادیده گرفتن بافتار متنی) به طور قابل ملاحظه‌ای کارایی را افزایش می‌دهد، زیرا نواحی یا اشیاء موجود در تصویر دقیقاً بیانگر محتوای آن تصویر هستند. علاوه بر آن هر کدام از این نواحی همچون یک کلمه عمل می‌کنند، در نتیجه در نظر گرفتن ارتباط بین این نواحی بسیار دقیق‌تر از بافتار متنی (ارتباط بین کلمات) برای ایجاد کلمات حاشیه می‌باشد. بافتار متنی دارای نویز و خطا است، زیرا برچسب‌های تصاویر آموزش به صورت دستی و توسط انسان ایجاد شده‌اند، بنابراین دارای ناسازگاری با مفهوم تصویر هستند.

۴-۳-۲- آنالیز ادغام بافتار ناحیه‌ای و موضوعات بصری

در این آزمایش، کارایی رویکرد پیشنهادی که شامل ادغام بافتار ناحیه‌ای و موضوعات بصری مبتنی بر تئوری تصمیم چندشرطی می‌باشد مورد ارزیابی قرار می‌گیرد. در جدول (۱) نتایج آزمایش‌های این قسمت ارائه شده است. تأثیر بافتار ناحیه‌ای بر بهبود کارایی که در آزمایش قبلی نشان داده شد، به وضوح مشخص است که ادغام آن با موضوع بصری در مدل پیشنهادی به بهبود بیشتر کارایی کمک می‌کند، به دلیل اینکه این دو نوع اطلاعات هر کدام بر جنبه‌های مختلفی از تصویر تمرکز دارند، کلمات متفاوتی را از تصویر استخراج می‌کنند و مکمل هم هستند. این بهبود در جدول (۱) دیده می‌شود و مقدار F_1 را از ۰/۴۲۱۴ به ۰/۴۷۷۶ افزایش داده است. شکل (۳) تعدادی کلمه نمونه و مقدار دقت و فراخوانی آن‌ها را در رویکرد پیشنهادی با ادغام بافتار ناحیه‌ای و موضوعات بصری نشان می‌دهد، شکل (۴) هم مقدار F_1 چندین کلمه نمونه را در رویکرد CMRM، ECMRM و مدل پیشنهادی نشان می‌دهد. تفاوت بین مقدار F_1 رویکرد پیشنهادی و ECMRM در شکل (۴) نویزهای موجود در حاشیه‌نویسی دستی (بافتار متنی) روش ECMRM را نشان می‌دهد. کاملاً قابل درک است که در نظر گرفتن ارتباط بین نواحی و ادغام آن با موضوعات بصری برای ایجاد کلمات حاشیه مرتبط با محتوای تصویر بسیار مفید است و مقدار F_1 را افزایش می‌دهد.

۴-۳-۳- مقایسه با سایر روش‌های آماری

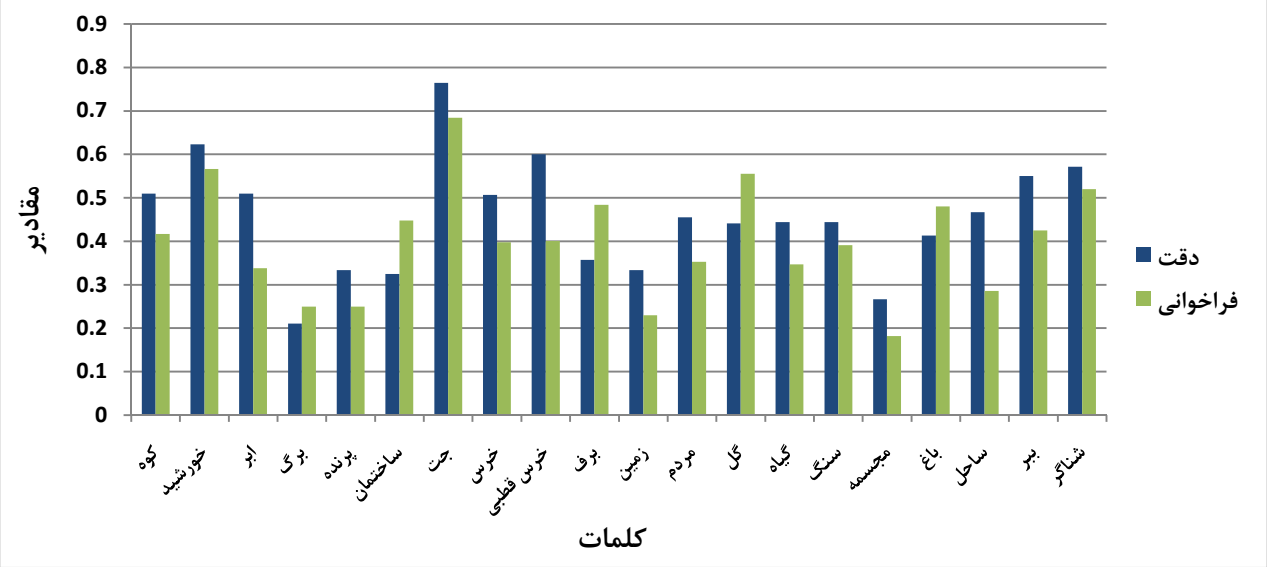
در دو آزمایش قبلی تأثیر بافتار ناحیه‌ای و ترکیب آن با موضوعات بصری مورد ارزیابی قرار گرفت. در این قسمت، رویکرد پیشنهادی یعنی ترکیب بافتار ناحیه‌ای و موضوعات بصری با سایر مدل‌های آماری در حوزه حاشیه‌نویسی تصویر مورد مقایسه قرار می‌گیرد. شکل (۵) میانگین دقت و فراخوانی

را در چندین مدل آماری و شکل (۶) تعداد کلمات با فراخوانی بزرگ‌تر از صفر را در همان مدل‌های آماری نشان می‌دهد که برتری رویکرد پیشنهادی به سایر روش‌ها به وضوح مشخص است. تعدادی از تصاویر آزمایش و حاشیه‌های ایجاد شده توسط مدل CMRM، ECMRM و رویکرد پیشنهادی در شکل (۷) دیده می‌شود. قابل مشاهده است که مدل پیشنهادی حاشیه‌های بهتری را نسبت به ECMRM ایجاد کرده که معنای موجود در تصویر را بیان می‌کنند.

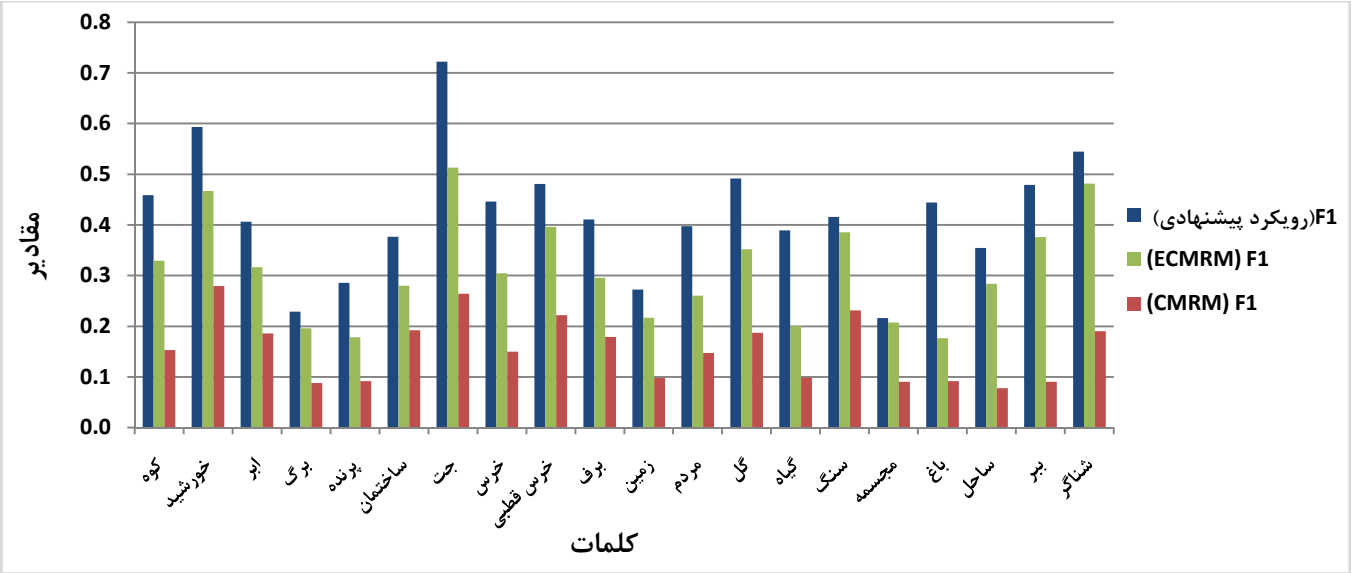
در رویکرد پیشنهادی با توجه به اعمال خوشه‌بندی، الگوریتم PLSA و تئوری تصمیم زمان مورد نیاز برای استخراج ویژگی و مرحله آموزش بالاتر از سایر روش‌های آماری می‌باشد اما فرآیند استخراج ویژگی، کسب دانش از تصاویر و حاشیه‌نویسی آن‌ها به صورت آفلاین انجام می‌شود. همچنین با توجه به بررسی معیارهای ارزیابی کارایی در الگوریتم‌های حاشیه‌نویسی تصویر یعنی دقت و فراخوانی و بهبود قابل توجه این معیارها در مدل پیشنهادی، می‌توان از این بار محاسباتی چشم‌پوشی نمود.

جدول (۱): مقدار میانگین دقت، فراخوانی و F_1 به ازای مقادیر V=۱۰۰، T=۸۰، R=۱۲۰.

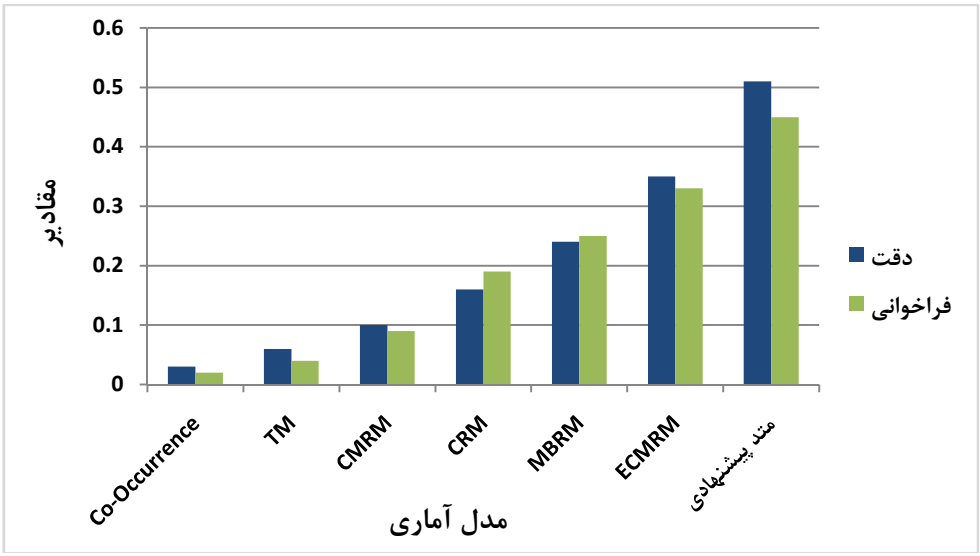
نام مدل	CMRM [24]	ECMRM [27]	مدل پیشنهادی با استفاده از بافتار ناحیه‌ای	مدل پیشنهادی با ادغام بافتار ناحیه‌ای و موضوعات بصری
دقت	۰/۱۲۸۸	۰/۳۰۵۰	۰/۴۴۰۷	۰/۵۰۶۳
فراخوانی	۰/۰۹۸۳	۰/۳۵۲۴	۰/۴۰۳۸	۰/۴۵۱۹
F_1	۰/۱۱۱۵	۰/۳۲۷۰	۰/۴۲۱۴	۰/۴۷۷۶
تعداد کلمات با فراخوانی بزرگ‌تر از صفر	۶۶	۱۲۹	۱۳۳	۱۴۱



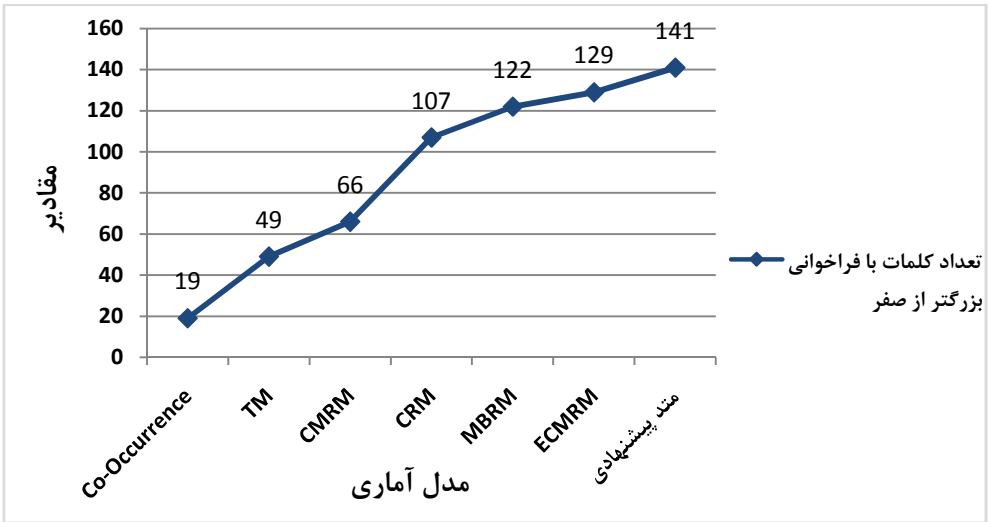
شکل (۳): تعدادی کلمه نمونه و مقدار دقت و فراخوانی آن‌ها در رویکرد پیشنهادی.



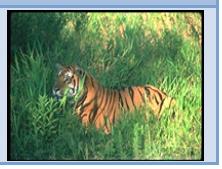
شکل (۴): تعدادی کلمه نمونه و مقدار F_1 آن‌ها در CMRM، ECMRM و مدل پیشنهادی.



شکل (۵): میانگین دقت و فراخوانی در چندین مدل آماری.



شکل (۶): تعداد کلمات با فراخوانی بزرگ‌تر از صفر در چندین مدل آماری.

تصویر	برچسب صحیح	CMRM	ECMRM	رویکرد پیشنهادی
	کوه، آسمان، خورشید، آب	ساختمان، غروب خورشید، شهر، آب، درخت	ساختمان، غروب خورشید، آب، مردم، درخت	درخت، آسمان، ابر، غروب خورشید، آب
	ابر، سبزه، کوه، شهر	هواپیما، آسمان، جت، آب، سبزه	آب، هواپیما، آسمان، سبزه، درخت	آسمان، سبزه، کوه، مردم، درخت
	در بزرگ، سبزه، معبد، درخت	مردم، ساختمان، پرچم، میدان رژه، آسمان	آب، آسمان، مردم، ساختمان، قایق	ساختمان، سبزه، درخت، آسمان، مردم
	جت، کوه، هواپیما	آسمان، هواپیما، جت، پرنده، مرغابی بزرگ	آسمان، هواپیما، جت، پرنده، مردم	آسمان، هواپیما، جت، درخت، ابر
	قهرمان، استخر، شناگر، آب	مردم، استخر، شناگر، آب، آسمان	شناگر، استخر، آب، مردم، آسمان	شناگر، استخر، آب، مردم، مسابقه
	گربه، جنگل، سبزه، ببر	آب، زمین، سبزه، باغ، گیاه	سبزه، گیاه، گربه، باغ، زمین	گربه، سبزه، گیاه، جنگل، باغ

شکل (۷): تعدادی تصاویر نمونه و حاشیه‌های آن‌ها در CMRM، ECMRM و رویکرد پیشنهادی.

۵- نتیجه‌گیری

در این مقاله، یک رویکرد حاشیه‌نویسی خودکار برای بهبود ECMRM معرفی شد. به جای استفاده از توزیع مستقیم نواحی در ECMRM، از توزیع موضوعات بین نواحی استفاده می‌شود. همچنین ارتباط بین نواحی یا بافتار ناحیه‌ای بسیار دقیق‌تر از ارتباط بین کلمات یا بافتار متنی موجود در ECMRM است. رویکرد پیشنهادی، بافتار ناحیه‌ای را با موضوعات بصری ادغام می‌کند، به دلیل وابستگی میان این دو نوع اطلاعات در تصویر، به منظور ادغام آن‌ها از الگوریتم TOPSIS موجود در تئوری تصمیم چندشرطی استفاده می‌شود. برای بدست آوردن توزیع سراسری موضوعات، موضوعات بصری توسط رویکرد PLSA بر روی داده‌های آموزش یاد گرفته می‌شوند. PLSA برای مدل کردن بافتار ناحیه‌ای یا موضوعات بین نواحی هم استفاده می‌شود. روش پیشنهادی بر روی مجموعه داده Corel ۵۰۰۰ تایی آزمایش شد و نتایج نشان داد که استفاده از بافتار ناحیه‌ای کارایی را بهبود داده و ادغام آن با موضوعات بصری از طریق تئوری تصمیم چند شرطی، کارایی بهتری را به همراه خواهد داشت.

مراجع

[1] A. Tousch, S. Herbin, J. Audibert, “Semantic hierarchies for image annotation: A survey,” Elsevier Ltd., Pattern Recognition, Vol. 40, pp. 333-345, 2012.

[2] D. Zhang, Md. M. Islam, G. Lu, “A review on automatic image annotation techniques,” Elsevier Ltd., Pattern Recognition, Vol. 45, 2011.

[3] R. Datta, D. Joshi, J. Li, J.Z. Wang, “Image Retrieval: Ideas, Influences, and Trends of the New Age,” ACM Comput. Surv, Vol. 40, No. 2, Article 5, 2008.

[4] A. Hanbury, “A survey of methods for image annotation,” Elsevier Ltd., Science Direct, Vol. 19, pp. 617-627, 2008.

[5] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, “Content-based image retrieval at the end of the early years,” IEEE Trans. on Pattern Anal. Mach.Intell, Vol. 22, No. 12, pp. 1349-1380, 2000.

[6] J. Tang, "Automatic Image Annotation and Object Detection," A thesis for the degree of Doctor of philosophy, University Of Southampton, 2008.
[7] Y. Liu, D. Zhang, G. Lu, W. Ma, "A survey of content-based image retrieval with high-level semantics," Pattern Recognition, Vol. 40, pp. 262-282, 2007.
[8] J. Y. Pan, H. J. Yang, P. Duygulu, Ch. Faloutsos, "Automatic image captioning," In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), pp. 1987-1990, 2004.
[9] J. S. Hare, P. H. Lewis, "Saliency-based models of image content and their application to auto-annotation by semantic propagation," In Proceedings of Multimedia and the Semantic Web / European Semantic Web Conference 2005.
[10] C. Cusano, G. Ciocca, R. Schettini, "Image Annotation Using Svm," In Proceedings of Internet Imaging IV, SPIE 5304, Vol. 5304, pp. 330-338, 2003.
[11] E. Chang, K. Goh, G. Sychay, G. Wu, "CBSA: content based soft annotation for multimodal image retrieval using Bayes point machines," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 26-38, 2003.
[12] J. Li, J. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," IEEE Trans. on PAMI, 25(19), pp. 1075-1088, 2003.
[13] A. Fakhari, A. M. Eftekhari Moghadam, "Combination of classification and regression in decision tree for multi-labeling image annotation and retrieval," Elsevier Ltd., Applied Soft Computing, Vol. 13, pp. 1292-1302, 2013.
[14] J. Tang, H. Li, G. Qi, T. Chua, "Image annotation by graph-based inference with integrated multiple/single instance representations," IEEE Trans. on Multimedia, Vol. 12, No. 2, 2010.
[15] J. Liu, M. Li, Q. Liu, H. Lu, S. Ma, "Image annotation via graph learning," Elsevier Ltd., Pattern Recognition, Vol. 42, pp. 218-228, 2009.
[16] J. Liu, B. Wang, H. Lu, S. Ma, "A graph-based image annotation framework," Pattern Recognition Letters, Vol. 29, pp.407-415, 2008.
[17] J. Liu, M. Li, W. Ma, Q. Liu, H. Lu, "An Adaptive Graph Model For Automatic Image Annotation," In Proceedings of the 8th ACM international workshop on Multimedia information retrieval, 2006.
[18] H. Tong, J. He, M. Li, W. Ma, H.J. Zhang, C. Zhang, "Manifold-Ranking Based Keyword Propagation For Image Retrieval," EURASIP J. Appl. Signal Process. Spec. Issue Inf. Min. Multimedia Database 21, pp. 1-10, 2006.
[19] D. Zhou, O. Bousquet, T. Lal, J. Weston, B. Scholkopf, "Ranking On Data Manifolds," In Proceedings of 18th Annual Conference on Neural Information Processing System, pp. 169-176, 2003.
[20] S. Abd manaf, M. J. Nordin, "Review on Statistical Approaches for Automatic Image Annotation," In International Conference on Electrical Engineering and Informatics, Selangor, Malaysia, August 2009.
[21] G. Carneiro, A. B. Chan, P. J. Moreno, N. Vasconcelos, "Supervised learning of semantic classes for image annotation and retrieval," IEEE Trans. on PAMI, Vol. 29, No. 3, 2007.
[22] Y. Mori, H. Takahashi, R. Oka, "Image-to-word transformation based on dividing and vector quantizing images with words," In Proceedings of First International Workshop Multimedia Intelligent Storage and Retrieval Management, 1999.
[23] P. Duygulu, K. Barnard, J. Freitas, D. Forsyth, "Object recognition as machine translation: learning a lexicon for a fixed image vocabulary," In Proceedings of the 7th European Conference on Computer Vision, Vol. 2353, pp. 97-112, 2002.
[24] J. Jeon, V. Lavrenko, R. Manmatha, "Automatic image annotation and retrieval using Cross-Media Relevance Model," In Proceedings of the 26th annual international ACM SIGIR, pp. 119-126, 2003.
[25] V. Lavrenko, R. Manmatha, J. Jeon, "A model for learning the semantics of pictures," In Proceedings of Advance in Neutral Information Processing, 2003.
[26] S. Feng, R. Manmatha, V. Laverenko, "Multiple Bernoulli Relevance Models for image and video annotation," In IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), pp. 1002-1009, 2004.
[27] Y. Wang, T. Mei, Sh. Gong, X. Sh. Hua, "Combining global, regional and contextual features for automatic image annotation," Elsevier Ltd., Pattern Recognition, Vol. 42, pp. 259-266, 2009.
[28] Zh. Li, H. Ma, Zh. Shi, Zh. Shi, "A Probabilistic Model for Automatic Image Annotation and Retrieval," In IEEE Ninth International Conference on Computer and Information Technology, 2009.
[29] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," Machine Learning, Vol. 42, pp. 177-196, 2001.
[30] Ch. L. Hwang, K. Paul Yoon, "Multiple Attribute Decision Making: Methods and Applications," Springer-Verlag, New York, 1981.
[31] Y.J. Lai, T.Y. Liu, Ch. L. Hwang, "TOPSIS for MODM," European Journal of Operational Research 76(3), pp. 486-500, 1994.
[32] G. Csurka, Ch. R. Dance, L. Fan, J. Willamowski, C. Bray, "Visual categorization with bags of keypoints," In Proceedings of ECCV Workshop on Statistical Learning in Computer Vision, pp. 1-16, 2004.
[33] D. G Lowe, "Distinctive image features from scale invariant keypoints," Int. J. Comput. Vision, vol. 60, no. 2, pp. 91-110, 2004.
[34] R. Xu, D. Ii, "Survey Of Clustering Algorithms," IEEE Transactions On Neural Networks, Vol. 16, No. 3, 2005.
[35] L. Fei Fei, P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories," In Proceedings of the IEEE Computer Society

Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, pp. 524-531, 2005.
[36] Y. Deng, B. S. Manjunath, "Unsupervised Segmentation of Color-Texture Regions in Images and Video," IEEE Trans. on PAMI, Vol. 23, No. 8, pp. 800-810, 2001.
[37] S. Kullback, R. A. Leibler, "On Information and Sufficiency," The Annals of Mathematical Statistics, Vol. 22, No. 1, pp. 79-86, 1951.
[38] A. P. Dempster, N. M. Laird, D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society, Vol. 39, No. 1, pp. 1-38, 2007.
[39] K. Hotta, "Scene Classification Based on Multi-resolution Orientation Histogram of Gabor Features," In ICVS'08 Proceedings of the 6th international conference on Computer vision systems, Vol. 5008, pp. 291-301, 2008.
[40] K. Hotta, "Scene classification based on local autocorrelation of similarities with subspaces," In 16th IEEE International Conference on Image Processing (ICIP), pp. 2053-2056, 2009.
[41] K. Hotta, "Object Categorization Based on Kernel Principal Component Analysis of Visual Words," In Proceedings of IEEE Workshop on Application of Computer Vision, pp. 1-8, 2008.
[42] K. Grauman, T. Darrell, "Discriminative Classification with Sets of Image Features," In Proceeding of International Conference on Computer Vision, pp. 1458-1465, 2005.
[43] S. Lazebnik, C. Schmid, J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2169-2178, 2006.

زیرنویس‌ها

[1] Text Based Image Retrieval (TBIR)
[2] Content Based Image Retrieval (CBIR)
[3] System
[4] Semantic Gap
[5] Semantic Based Image Retrieval (SBIR)
[6] Automatic Image Annotation (AIA)
[7] Vector Space Models
[8] Classification Methods
[9] Graph Based Methods
[10] Statistical Models
[11] Single Labeling
[12] Multi Labeling
[13] Correlation
[14] Joint Probability
[15] Clustering
[16] Global Features
[17] Regional Features
[18] Textual Contexts
[19] Co-Occurrence
[20] Regional Contexts
[21] Visual Topics
[22] Grid Based
[23] Partitioning
[24] Patches
[25] Visual Words
[26] Multi Criteria Decision Making (MCDM)
[27] Alternatives
[28] Option
[29] Criteria
[30] Attribute
[31] Ideal Alternative
[32] Negative Ideal Alternative
[33] Ideal Solution
[34] Negative Ideal Solution
[35] Separation Measure