

بهبود بازیابی تصاویر رنگی با استفاده از رنگ، بافت و شکل در روش کیسه کلمات بصری مبتنی بر امضاء

زهرا هادیان برسیانی^۱، فرساد زمانی بروجنی^{۲*}

*نویسنده مسئول، دریافت: ۹۸/۰۷/۱۲، بازنگری: ۹۹/۰۳/۱۲، پذیرش: ۹۹/۰۷/۱۳

^۱ دانشجوی کارشناسی ارشد، مهندسی کامپیوتر، دانشکده فنی مهندسی، دانشگاه آزاد اسلامی واحد خوراسگان (اصفهان)، اصفهان، ایران،
^۲ استادیار، دانشکده فنی مهندسی، دانشگاه آزاد اسلامی واحد خوراسگان (اصفهان)، اصفهان، ایران

چکیده

تولید و ذخیره‌سازی تصاویر به‌صورت دیجیتالی باعث به وجود آمدن سیستم‌های بازیابی تصویر، با قابلیت بازیابی تصاویر با استفاده از محتوای بصری آن‌ها گردید. در بین سیستم‌های بازیابی تصویر مبتنی بر متن و یا محتوا، روش‌های مبتنی بر محتوا عملکرد جامع‌تر و بهتری دارند. اما یکی از نکاتی که در موفقیت این سیستم‌ها تأثیر بسزایی دارد ویژگی‌های مستخرج از تصاویر است. به این معنی که اگر ویژگی‌های مناسب و کاملی از تصاویر استخراج نشود، دقت این روش‌ها در بازیابی کاهش می‌یابد. یکی از روش‌هایی که اخیراً در حوزه بازیابی مبتنی بر محتوا بسیار مورد توجه محققین قرار گرفته است، تکنیک کیف کلمات بصری مبتنی بر امضاء است که بدون استفاده از روش‌های خوشه‌بندی ویژگی‌ها را کوانتیزه می‌کند. باوجوداینکه برتری این روش نسبت به روش کیف کلمات در مقالات اخیر گزارش شده است، اما برای رسیدن به نتایج مناسب، تصاویر را به بلوک‌های بسیار کوچکی تقسیم می‌کند تا به این صورت بتواند با تکیه بر ویژگی‌های محلی، به کلمات مشابهی در تصاویر مختلف دست یابد. به بیان دیگر، در تکنیک کیف کلمات بصری مبتنی بر امضاء ویژگی‌های سراسری نادیده گرفته می‌شود و بنابراین این روش برای هر نوع تصویری مناسب نیست و در بسیاری از پایگاه داده‌ها عملکرد ضعیفی دارد. از این‌رو در این مقاله در جهت بهبود دقت بازیابی تصاویر از ترکیب سه ویژگی سراسری گابور، هیستوگرام جهت لبه و هیستوگرام گرادیان جهت‌دار با دو ویژگی محلی رنگ و بافت سراسری غالب به‌صورت کیف کلمات بصری مبتنی بر امضاء استفاده شده است، تا به این صورت بتوان دانش کافی از محتوای تصاویر استخراج نمود و در نتیجه دقت بازیابی را بهبود داد. نتایج ارزیابی بر روی دو پایگاه داده Oxford و Wang حاکی از آن است که مدل پیشنهادی در مقایسه با روش پیشین در معیار دقت بهبود داشته است.

کلمات کلیدی: بازیابی مبتنی بر محتوا، هیستوگرام جهت لبه، هیستوگرام گرادیان جهت‌دار، کیف کلمات مبتنی بر امضاء.

۱- مقدمه

سیستم‌ها این است که اگر متن همراه با تصویر، به‌درستی محتوای تصویر را ارائه ندهد، باعث عدم موفقیت سیستم در بازیابی تصاویر مشابه می‌شود. همچنین، این سیستم‌ها کاربرد جامعی ندارند و در صورت عدم وجود متن همراه با تصویر کاملاً ناکارآمد هستند. در مقابل، روش‌های بازیابی مبتنی بر محتوا ویژگی‌های بصری تصاویر از جمله رنگ و بافت و شکل را استخراج کرده و تصاویری که دارای ویژگی‌های مشابه هستند را بازیابی می‌نمایند [۲]. در اغلب این سیستم‌ها ویژگی‌های بصری سطح پایین به‌صورت سراسری و یا محلی استخراج می‌شوند [۳]. استفاده هم‌زمان از هر دو نوع ویژگی فوق می‌تواند در موفقیت سیستم تأثیر بالایی داشته باشد، اما همچنان بسیاری از سیستم‌های بازیابی تصویر فقط به

یکی از مسائلی که در حوزه تصاویر دیجیتالی مطرح می‌باشد، بازیابی تصاویر مشابه، با تصویر پرس‌وجوی کاربر است. این مسئله به دلیل رشد روزانه تصاویر در محیط‌های مختلف به چالشی بسیار بزرگ تبدیل شده است. در چنین شرایطی وجود سیستمی که بتواند تصاویر دارای محتوای مشابه را بازیابی کند، ضروری به نظر می‌رسد [۱]. مطالعات انجام شده در این زمینه باعث به وجود آمدن دو دسته اصلی سیستم‌های مبتنی بر محتوا^۱ و سیستم‌های مبتنی بر متن^۲ گردید. سیستم‌های مبتنی بر متن، با استفاده از متون همراه با تصویر یا حاشیه تصویر، بازیابی تصاویر مشابه را انجام می‌دهند. اما یکی از اصلی‌ترین مشکلات این

در مرحله استخراج ویژگی در یک سیستم بازیابی تصویر، ویژگی‌های مختلف را می‌توان به دو صورت سراسری [۸، ۹] و محلی [۱۰، ۱۱] استخراج نمود. ویژگی‌های سراسری به ویژگی‌هایی اطلاق می‌شود که از کل یک تصویر استخراج می‌شوند. در مقابل ویژگی‌های محلی به ویژگی‌هایی اطلاق می‌شود که از بخش‌های مختلف یک تصویر استخراج می‌شوند. در این روش تصویر با استفاده از روش‌های تقسیم‌بندی به قسمت‌هایی تقسیم می‌شود و از هر قسمت تصویر، ویژگی مورد نظر استخراج می‌شود. [۱۲، ۱۳]. عموماً ترکیب دو نوع ویژگی فوق می‌تواند باعث موفقیت بیشتر یک سیستم بازیابی شود [۱۴].

۲-۳- کیف کلمات بصری

در سیستم‌های بازیابی تصاویر غالباً از ترکیب چندین ویژگی مختلف برای رسیدن به نتایج بهتر استفاده می‌شود. اما یکی از مشکلاتی که در استخراج ویژگی‌های متعدد رخ می‌دهد، بزرگ شدن طول بردارهای ویژگی تصاویر است. یکی از روش‌هایی که برای رفع این مشکل استفاده می‌گردد، روش کیف کلمات بصری^۹ است [۱۵، ۱۶]. این مدل از دو مرحله اصلی استخراج ویژگی و تولید فرهنگ لغت تشکیل شده است و در دو دسته کلی مبتنی بر خوشه‌بندی و مبتنی بر امضا می‌باشد.

روش BoVW: در روش کیف کلمات مبتنی بر خوشه‌بندی، کلیه ویژگی‌های مورد استفاده از تصاویر استخراج می‌شوند؛ ویژگی‌ها توسط الگوریتم‌های خوشه‌بندی مانند k-means در k گروه همگن خوشه‌بندی می‌شوند. خروجی این مرحله، فرهنگ لغت کلمات بصری می‌باشد و مراکز خوشه‌های تشکیل شده توسط الگوریتم k-means، کلمات بصری هستند [۱۷، ۱۸]. در مرحله دوم، برای هر ویژگی استخراج شده از تصویر، فاصله بین آن با تمامی کلمات بصری (مراکز خوشه‌ها) موجود در فرهنگ لغت محاسبه شده، نزدیک‌ترین کلمه بصری متناظر آن مشخص شده و به آن نگاشت می‌شود. در نهایت برای هر تصویر، کیف کلمات بصری آن ایجاد خواهد شد. کیف کلمات بصری یک هیستوگرام حاوی تعداد رخدادهای هر کلمه بصری است. خروجی این تکنیک یک ماتریس $n \times m$ بعدی است که n تعداد تصاویر و m تعداد کلمات بصری حاصل شده از مرحله خوشه‌بندی است [۱۸]. روش‌های کیف کلمات بصری مبتنی بر خوشه‌بندی، به دلیل محدودیت‌های روش‌های خوشه‌بندی دارای نقاط ضعف مختلفی هستند. غالباً روش‌های خوشه‌بندی به صورت تصادفی مراکز اولیه خوشه‌ها را انتخاب می‌کند، یا نیازمند به تعیین تعداد خوشه می‌باشند که باعث می‌شود تکنیک کیف کلمات بصری نیز با محدودیت‌های فوق مواجه باشد. برای رفع این مشکل، کیف کلمات بصری مبتنی بر امضا^{۱۰} معرفی شدند که بدون استفاده از خوشه‌بندی ویژگی‌های استخراج شده از تصاویر را به کلمات بصری نگاشت می‌کند [۴، ۱۹].

روش S-BoVW: در تکنیک کیف کلمات بصری مبتنی بر امضا، تصاویر به بلوک‌های ثابتی تقسیم‌بندی می‌شود. از هر بلوک ویژگی‌هایی همچون بافت و رنگ به صورت محلی استخراج می‌شود. هر یک از ویژگی‌های استخراج شده از هر بلوک، به صورت رشته‌ای متنی در نظر گرفته می‌شود و برای هر تصویر سندی متشکل از کلیه رشته‌های مرتبط با بلوک‌ها تهیه می‌گردد. برای پردازش هر سند تشکیل شده برای هر تصویر، از تکنیک‌های متن‌کاوی استفاده می‌گردد و مدل فضای بردار هر تصویر تشکیل می‌شود و با استفاده از تکنیک‌های وزن دهی متداول در متن‌کاوی همچون تکنیک TF-IDF وزنی برای هر کلمه موجود در هر سند محاسبه می‌شود. بردار نهایی تشکیل شده برای هر تصویر، برداری از وزن‌های TF-IDF است [۴، ۱۹]. در این تکنیک از معیارهای فاصله متداول مثل معیار تشابه کسینوسی برای محاسبه تشابه یا عدم تشابه بردارهای وزن TF-IDF استفاده می‌شوند و مشابه‌ترین تصاویر به تصویر پرس‌وجو به کاربر ارائه می‌شود [۴، ۱۹].

استفاده از یک نوع ویژگی محلی یا سراسری بسنده می‌کنند. از جمله پژوهش‌هایی که فقط از ویژگی‌های محلی برای توصیف تصاویر استفاده کرده است، می‌توان به روش سانتوس و همکاران در سال ۲۰۱۷ اشاره نمود [۴]. الگوریتم SDLCT، دو ویژگی محلی رنگ و بافت را از بخش‌های مختلف هر تصویر استخراج می‌کند و با استفاده از تکنیک کیف کلمات بصری مبتنی بر امضا و تکنیک وزن دهی TF-IDF بردار نهایی وزن تصاویر را تشکیل می‌دهد، و در نهایت از بردارهای وزنی تصاویر برای محاسبه تشابه بین آن‌ها و یافتن برترین تصاویر مشابه استفاده می‌کند. تکنیک کیف کلمات بصری مبتنی بر امضاء اگرچه بدون استفاده از روش‌های خوشه‌بندی ویژگی‌ها را کوانتیزه می‌کند، اما برای رسیدن به نتایج مناسب، تصاویر را به بلوک‌های بسیار کوچکی تقسیم می‌کند تا به این صورت بتواند به کلمات مشابهی در تصاویر مختلف دست یابد. به عبارت ساده تکنیک کیف کلمات بصری مبتنی بر امضاء برای تصاویری که در آنها اشیاء بخش وسیعی از تصویر را می‌پوشانند مناسب نیست و در بسیاری از پایگاه داده‌ها عملکرد ضعیفی دارد. در این راستا در این مقاله یک سیستم بازیابی تصویر مبتنی بر محتوا متشکل از ویژگی‌های محلی و سراسری جهت بهبود دقت بازیابی تصویر ارائه شده است. در این روش از سه ویژگی سراسری هیستوگرام جهت لبه، هیستوگرام گرادیان جهت دار و گابور در کنار تکنیک کیف کلمات بصری مبتنی بر امضاء برای بهبود عملکرد بازیابی تصویر استفاده شده است. به این صورت محدودیت تکنیک کیف کلمات مبتنی بر امضاء در پایگاه داده‌های خاص نیز برطرف می‌گردد.

در ادامه این مقاله، در بخش ۲ به معرفی سیستم‌های بازیابی تصویر پرداخته می‌شود. در بخش ۳ روش پیشنهادی به‌طور کامل معرفی خواهد شد و در بخش ۴ نتایج حاصل شده، ارائه می‌شود و در نهایت پیشنهادهایی برای کارهای آتی ارائه می‌گردد.

۲- سوابق تئوری و کارهای مرتبط

۲-۱- سیستم‌های بازیابی تصویر

سیستم‌های بازیابی تصویر، پس از دریافت تصویر پرس‌وجوی کاربر و جستجو در پایگاه داده، تصاویر مشابه به تصویر پرس‌وجو را به کاربر ارائه می‌دهند. در این فرآیند سرعت بازیابی و دقت بازیابی از اهمیت بسزایی برخوردار است [۵]. سیستم‌های بازیابی به ترتیب در دو بخش برون‌خط و برخط ویژگی تصاویر را استخراج کرده و بردارهای ویژگی را جهت بازیابی سریع تصاویر شاخص گذاری می‌کنند و پس از دریافت پرس‌وجوی کاربر و تشکیل بردار ویژگی آن، با استفاده از یک معیار تشابه به‌عنوان مثال فاصله اقلیدسی، تشابه بین بردارهای ویژگی را محاسبه کرده و تصاویر مشابه را بازیابی می‌کند. در این سیستم‌ها تصاویر مختلف بر اساس مقدار تشابه به تصویر پرس‌وجو مرتب شده و لیست تصاویر مرتب شده به کاربر ارائه می‌شود [۶].

۲-۲- استخراج ویژگی

ویژگی‌های تصاویر و یا توصیفگرها از محتوای بصری تصاویر استخراج می‌شوند و غالباً به صورت بردار و یا گراف ذخیره می‌شوند و نمایش فشرده‌ای از تصاویر را ارائه می‌دهند. به‌طور کلی ویژگی‌های قابل استخراج از تصاویر به دو دسته اصلی فتومتریک^۳ و هندسی^۴ تقسیم می‌شوند. ویژگی‌های فتومتریک شامل رنگ و بافت تصاویر هستند، و مستقیماً از پیکسل‌های خام تصاویر استخراج می‌شوند [۷]. ویژگی‌هایی همچون گابور، ماتریس هم رخداد^۵ در این دسته قرار می‌گیرند. ویژگی‌های هندسی شامل ویژگی‌های شکل تصاویر هستند. ویژگی‌هایی همچون کاننور^۶، مجموعه نقاط^۷، سطح^۸ در این دسته قرار می‌گیرند [۷].

۲-۴- پیشینه تحقیق

روش‌های بازیابی تصویر بررسی شده در این بخش، روش‌های مبتنی بر محتوا و کیف کلمات بصری هستند.

لیو و همکاران در [۲۰] یک روش بازیابی مبتنی بر مدل کیف کلمات ارائه نموده‌اند. در این روش، از توصیفگرهای محلی مانند SIFT استفاده شده است. چاتورانی و همکاران در [۲۱] یک روش بازیابی محتوای مبتنی بر امضاء ارائه نموده‌اند. این روش یک نمایش تصویر جدید با استفاده از یک رویکرد جدید برای تولید امضاهای تصویر با نام CBIR - ISIG ارائه می‌دهد. امضاهای تصویر توسط اعمال شاخص‌های تصادفی^{۱۱} به کیسه کلمات تولید می‌شوند. روش شاخص تصادفی یک رویکرد کارآمد و مقیاس‌پذیر برای کاهش ابعاد است که هزینه‌های محاسباتی را کاهش می‌دهد. نتایج نشان می‌دهد که امضاهای تصویر سرعت بازیابی را بهبود می‌بخشد و نیاز حافظه را برای ذخیره‌سازی کاهش می‌دهد. در [۲۲] محمد الخوالانی و همکاران یک روش بازیابی استفاده از توصیفگرهای محلی و کیسه کلمات بصری ارائه نموده‌اند. این سیستم برای بهبود فرآیند بازیابی از ترکیب ویژگی‌هایی مانند SIFT, SURF استفاده نموده است. در این روش از الگوریتم خوشه‌بندی k-means برای خوشه‌بندی توصیفگرها و ساخت واژگان بصری استفاده شده است. همچنین از الگوریتم طبقه‌بندی ماشین بردار پشتیبان برای بازیابی تصاویر مشابه با تصویر پرس‌وجو استفاده شده است.

در [۲۳] کاراکاسی و همکاران از ویژگی moment invariants به صورت محلی و تبدیل ویژگی‌های محلی به کیف کلمات برای بازیابی تصاویر استفاده کردند. نتایج روش فوق بر روی دو پایگاه داده UCID و UKBench با توصیفگر محلی SIFT و SURF مقایسه شده است که نشان‌دهنده عملکرد بهتر آن در مقایسه با دیگر روش‌ها می‌باشد. در [۲۴] میراندا و همکاران از روش کیف کلمات بصری مبتنی بر امضاء برای نمایه‌سازی و جستجوی تصویر ارائه نموده‌اند. در این مقاله، SDLC، یک روش بازیابی تصویری است که روش کیف کلمات بصری مبتنی بر امضاء را برای شناسایی کلمات بصری، به جای استفاده از روش‌های خوشه‌ای به کار می‌گیرد. در این روش ویژگی رنگ از تصاویر استخراج شده و هیستوگرام رنگ بر اساس مقادیر آن به صورت نزولی مرتب شده و مقادیر بالاتر از حداقل آستانه به عنوان یک کلمه از یک بلوک از تصویر در نظر گرفته شده است. این روش در مقایسه با روش‌های مبتنی بر خوشه عملکرد مناسبی دارد. در [۲۵] دیمیترفسکی و همکاران بهبودی بر روش کیسه کلمات بصری برای بازیابی تصویر ارائه نموده‌اند. در این روش از درخت خوشه‌بندی پیش‌بینی کننده PTCs^{۱۲}، استفاده شده است همچنین برای افزایش ثبات مدل از جنگل تصادفی PTCs^{۱۳} استفاده کرده‌اند. نتایج نشان می‌دهد که روش پیشنهادی آن‌ها کلمات بصری با قدرت تفکیک‌پذیری برتر را تولید می‌کند و در نتیجه عملکرد بهتر بازیابی را ارائه می‌دهد.

در [۴] میراندا و همکاران روش کیسه کلمات بصری مبتنی بر امضاء با استفاده از دو ویژگی رنگ و بافت غالب^{۱۴} برای بازیابی تصویر ارائه نموده‌اند. یکی از مشکلات این روش این است که در تعداد کم بلوک‌های تصویر، نتایج نامطلوب است و برای رسیدن به نتایج مطلوب تصاویر باید به بلوک‌های زیادی تقسیم شود، از این رو تعداد کلمات منحصر به فرد تولید شده در این حالت بسیار زیاد شده و در نتیجه ماتریس بردارهای وزن TF-IDF دارای ابعاد بسیار بالایی خواهند بود که این امر اثرات جانبی بسیاری از جمله افزایش محاسبات، مشکل حافظه مصرفی را به همراه خواهد داشت و با افزایش سایز تصاویر و تعداد تصاویر در پایگاه داده این مسئله بدتر نیز می‌شود. در [۲۶] سیلوا و همکاران گراف‌های مبتنی بر کیسه کلمات را ارائه نموده‌اند. در این مقاله دو روش BoSG^{۱۵} و BoVG^{۱۶} ارائه شده است که به ترتیب برای نمایش گراف‌ها و تصاویر به کار می‌روند. از روش BoSG برای طبقه‌بندی گراف استفاده شده است. روش BoVG برای طبقه‌بندی تصویر در مجموعه تصاویر استفاده شده است. این چارچوب فرصت‌های بازیابی، طبقه‌بندی

و خوشه‌بندی در مجموعه‌های داده‌های بزرگ را امکان‌پذیر می‌کند. در [۲۷] چینیو و همکاران یک روش بازیابی تصاویر پزشکی با عنوان ICARUS ارائه نمودند که مبتنی بر تکنیک کیف کلمات بصری مبتنی بر امضاء می‌باشد. در این روش با استفاده از ویژگی‌های بافت و رنگ و تکنیک کیف کلمات مبتنی بر امضاء ویژگی‌های تصاویر زخم پوست استخراج شده و به این صورت تشخیص مناسب برای بیماری انجام می‌شود.

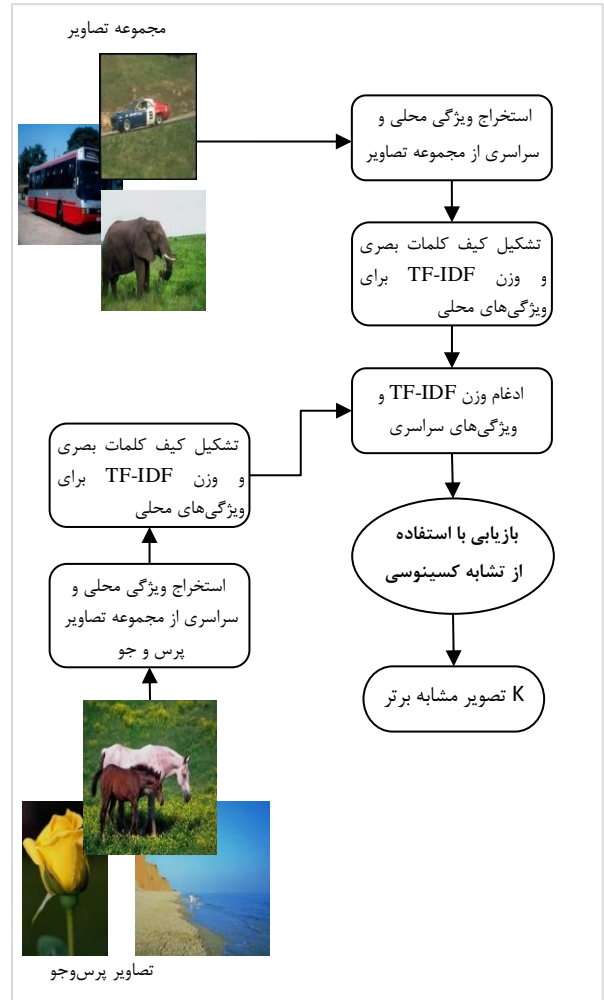
در [۲۸] چینیو و همکاران یک روش بازیابی مبتنی بر تکنیک بهبود یافته کیف کلمات بصری مبتنی بر امضاء با عنوان BOSS^{۱۷} ارائه کردند. در این روش ویژگی‌های محلی از سوپر پیکسل‌ها استخراج می‌شوند که این مورد باعث غلبه بر محدودیت‌های روش کیف کلمات مبتنی بر امضاء اولیه می‌شود. در [۲۹] کبایی و همکاران از ترکیب ویژگی‌های محلی و سراسری برای بازیابی تصویر استفاده کرده‌اند. در این روش یک توصیفگر جدید معرفی شده است که در آن ویژگی‌های سراسری با استفاده از تبدیل موجک یا یکدیگر ترکیب شده و برای ویژگی محلی از توصیفگر SURF و تکنیک کیف کلمات بصری استفاده کرده‌اند. نتایج این تحقیق حاکی از آن است که ترکیب ویژگی‌های سراسری و محلی در بازیابی تصاویر موفق عمل می‌کند. در [۳۰] رشنو و همکاران یک سیستم بازیابی با استفاده از یک روش استخراج ویژگی جدید ارائه کردند، روش استخراج ویژگی جدید شامل نرم^{۱۸} مؤلفه‌های فرکانس پایین در تبدیل موجک و ویژگی‌های رنگ در دامنه‌های RGB و HSV است. همچنین آن‌ها برای یافتن مرتبط‌ترین ویژگی‌ها از الگوریتم بهینه‌سازی کلونی مورچگان^{۱۹} استفاده نموده‌اند. در [۳۱] شریف و همکاران جهت ارائه یک سیستم بازیابی موفق دو ویژگی SIFT^{۲۰} و BRISK^{۲۱} را با استفاده از تکنیک کیف کلمات بصری ترکیب نمودند. در [۳۲] سرور و همکاران برای از بین بردن شکاف معنایی و بهبود فرآیند بازیابی دو ویژگی LIOP^{۲۲} و LBPV^{۲۳} را با استفاده از کیف کلمات ترکیب نمودند. در [۳۳] گیتا و همکاران یک سیستم بازیابی تصویر ارائه کردند که در آن از ترکیب ویژگی‌های محلی و سراسری استفاده شده است. در این مدل ویژگی رنگ و لبه دو ویژگی سراسری استخراج شده از تصاویر هستند و الگوی باینری محلی و SURF نیز به صورت محلی استخراج شده‌اند. در جدول ۱ روش‌های فوق به طور مختصر بیان شدند.

جدول ۱- خلاصه روش‌های معرفی شده

مرجع	ویژگی محلی	ویژگی سراسری	کیف کلمات	پایگاه داده
[۲۱]	SIFT	-	✓	Corel 1K
[۲۲]	SURF, SIFT	-	✓	Filckr logos, ALOI
[۲۳]	moment invariants	-	✓	UKBench, UCID
[۲۴]	Color	-	✓	Corel 1K, Oxford
[۲۵]	SIFT	-	✓	Oxford, Paris
[۴]	Color- LPB	-	✓	Corel 1K, Oxford
[۲۶]	Interest point	-	✓	Caltech, ALOI
[۲۷]	Color- LPB	-	✓	Medical skin image
[۲۷]	Color Histogram	-	✓	Corel, Caltech, Texture
[۲۹]	SURF -Local ternary patterns	color wavelet transform	✓	New-BarkTex, Outex-TC13, Outex-TC14
[۳۰]	Wavelet Feature, DCD, Color Statistic, Color Histogram	-	-	Corel
[۳۱]	SIFT, BRISK	-	✓	Corel-1K, Caltech-256
[۳۲]	LBPV, LIOP	-	✓	Corel-5K, Holidays
[۳۳]	LPB, SURF	Color, Edge	-	Corel1k, medical database

۳- روش پیشنهادی

روش ارائه شده در این مقاله با نام $SDLCTG^4$ ، در دسته روش‌های بازیابی تصویر مبتنی بر محتوا است که از ترکیب تکنیک کیف کلمات بصری مبتنی بر امضا و ویژگی‌های سراسری برای بازیابی موفق تصاویر استفاده می‌کند. در این روش از دو ویژگی محلی رنگ و بافت غالب و سه ویژگی سراسری هیستوگرام جهت لبه، هیستوگرام گرادین جهت دار و گابور برای توصیف کامل تصاویر استفاده شده است. دیاگرام این روش در شکل ۱ ارائه شده است.



شکل ۱- دیاگرام الگوریتم پیشنهادی SDLCTG

۳-۱- پیش‌پردازش و استخراج ویژگی‌های محلی و سراسری

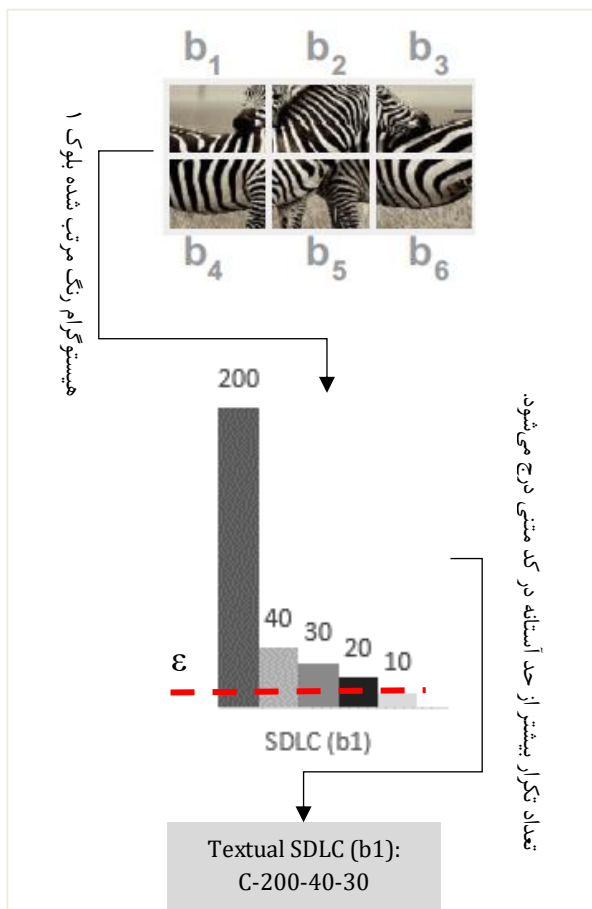
در الگوریتم $SDLCTG$ ، تصاویر برای استخراج دو نوع ویژگی محلی و سراسری پیش‌پردازش می‌شوند. برای استخراج ویژگی‌های محلی تصاویر به بلوک‌های 25×25 تقسیم می‌شوند و هر تصویر به 625 بلوک ثابت تقسیم می‌شود و از هر بلوک دو ویژگی محلی بافت و رنگ غالب استخراج می‌شود. برای استخراج ویژگی‌های سراسری، با توجه به اینکه طول بردار خروجی ویژگی‌هایی همچون HOG و گابور وابسته به سایز تصاویر بوده و متغیر می‌باشد، تصاویر موجود در پایگاه داده با توجه به سایز بزرگ‌ترین تصویر در پایگاه تغییر سایز داده می‌شوند، به‌عنوان مثال در پایگاه داده Oxford کلیه تصاویر به سایز 768×1024 تغییر سایز داده می‌شوند. پس از آماده‌سازی تصاویر، استخراج ویژگی انجام می‌شود.

استخراج ویژگی رنگ محلی غالب: برای استخراج ویژگی محلی رنگ در الگوریتم $SDLCTG$ ، از تکنیک استخراج رنگ محلی غالب استفاده شده است. اگر فرض شود $B_j = \{b_1, b_2, \dots, b_n\}$ مجموعه بلوک‌های بدون همپوشانی در تصویر

I باشند. برای هر بلوک $b_i \in B_j$ کلمه بصری رنگ محلی غالب به‌صورت زیر از رابطه (۱) تشکیل می‌شود:

$$T_C = \langle \hat{C}_1 - \hat{C}_2 - \hat{C}_3 - \dots - \hat{C}_n \rangle \quad (1)$$

در رابطه فوق \hat{C}_k بیانگر شناسه رنگ می‌باشد، به‌طوری‌که $H_i(\hat{C}_k) \geq \varepsilon$ و $\hat{C}_k > \hat{C}_{k+1}$ ، $1 \leq k \leq n-1$ حداقل آستانه برای تعیین فراوانی رنگ می‌باشد و H_i هیستوگرام رنگ بلوک i از تصویر I است. برای استخراج کلمات رنگ غالب محلی، هیستوگرام رنگ هر بلوک از تصویر دریافت شده و کد رنگ‌ها بر اساس تعداد تکرار به‌صورت نزولی مرتب می‌شوند و کدهایی که تعداد تکرارشان بیشتر از حداقل آستانه تعریف شده باشد به یکدیگر با استفاده از "-" پیوست می‌شوند. بنابراین، $SDLCTG$ یک تصویر مجموعه‌ای از کدهای متنی از تمام بلوک‌های تشکیل دهنده یک تصویر است [۲۴]. در شکل ۲ استخراج رنگ محلی غالب نشان داده شده است.



شکل ۲- استخراج ویژگی رنگ محلی غالب [۴]

استخراج ویژگی بافت محلی غالب: برای استخراج بافت محلی غالب از الگوریتم الگوی باینری محلی یا LBP استفاده شده است [۴، ۳۴]. اگر فرض شود $B_j = \{b_1, b_2, \dots, b_n\}$ مجموعه بلوک‌های بدون همپوشانی در تصویر I باشند. برای هر بلوک $b_i \in B_j$ کلمه بصری بافت محلی غالب از رابطه (۲) حاصل می‌شود.

$$T_t = \langle \hat{t}_1 - \hat{t}_2 - \hat{t}_3 - \dots - \hat{t}_n \rangle \quad (2)$$

در رابطه فوق \hat{t}_n بیانگر الگوی بافتی می‌باشد که در بلوک مربوط رخ داده است، به‌طوری‌که $frequency(\hat{t}_i) \geq \delta$ و $\hat{t}_i > \hat{t}_j$ ، $i < j$ حداقل

تشکیل کیف کلمات بصری از ویژگی رنگ و بافت غالب: پس از تشکیل

فایل متنی مربوط به هر تصویر، با استفاده از تکنیک فضای بردار و الگوریتم وزن‌دهی TF-IDF بردار وزن نهایی مربوط به دو ویژگی استخراج شده تشکیل می‌شود. کلمات هر تصویر با در نظر گرفتن فاصله تفکیک می‌شوند. از لیست کلمات تشکیل شده، کلمات منحصربه‌فرد انتخاب می‌شوند و کلمات تکراری حذف می‌شوند. در این مرحله دیکشنری کلمات کل تصاویر که مجموعه‌ای از کلمات منحصربه‌فرد در کل پایگاه داده است، تشکیل می‌شود. با توجه به سایز دیکشنری کلمات، ماتریسی تشکیل می‌شود که سطرها این ماتریس تعداد تصاویر و ستون‌های آن برابر با تعداد کلمات موجود در دیکشنری لغات است. برای هر کلمه در دیکشنری، اگر کلمه i در تصویر j حضور داشته باشد با استفاده از رابطه (۳) وزن TF-IDF آن محاسبه شده و مقدار وزن حاصل شده در سطر j و درایه i درج می‌شود. ماتریس حاصل شده در این مرحله یک ماتریس خلوت^{۲۶} است.

$$W_{tfidf_t} = raw\ count_t \times \log \frac{N}{n_t} \quad (3)$$

در رابطه فوق N تعداد کل تصاویر در پایگاه داده است و n_t بیانگر تصاویری است که در برگزیده کلمه مورد نظر هستند. لازم به ذکر است که در این مرحله دو ماتریس وزن مجزا برای بافت محلی غالب و رنگ محلی غالب تشکیل می‌شود. پس از تشکیل بردار وزن هر یک از تصاویر و تشکیل دو ماتریس وزن خلوت برای کلیه تصاویر موجود در پایگاه داده دو ماتریس فوق به‌صورت خطی با استفاده از رابطه (۴) با یکدیگر ترکیب می‌شوند.

$$SDLCT = \alpha SDLC + (1 - \alpha)SDLT \quad (4)$$

در رابطه بالا α وزنی است که برای ترکیب دو ویژگی بافت و رنگ غالب محلی استفاده می‌شود و مقداری بین $(0,1)$ می‌گیرد. پس از ترکیب دو ویژگی فوق خروجی این مرحله یک ماتریس خلوت متشکل از مقادیر بافت و رنگ محلی می‌باشد.

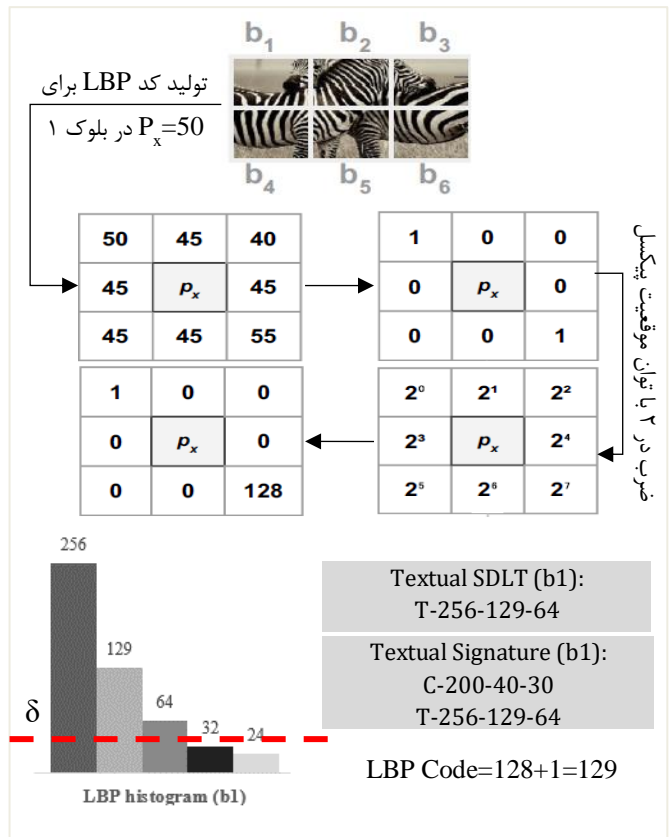
استخراج ویژگی سراسری گابور: در الگوریتم SDLCTG برای استخراج

ویژگی گابور ابتدا مجموعه فیلترهای آن مشخص می‌شود، که در این روش از ۴۰ فیلتر گابور در ۸ جهت و ۵ مقیاس مختلف استفاده شده است. خروجی فیلتر گابور از با توجه به طول عرض تصویر متغیر می‌باشد. خروجی گابور یک بردار $1 \times N$ بعدی است که N ابعاد بردار و وابسته به سایز تصویر است. از آنجاکه پیکسل‌های مجاور در یک تصویر معمولاً بسیار همبسته هستند، این افزونگی اطلاعات را می‌توان با کمینه کردن ویژگی‌های حاصل از فیلترهای گابور، کاهش داد. با این روش سایز بردار خروجی ویژگی گابور نیز با توجه به فاکتور کمینه‌سازی^{۲۷} کاهش می‌یابد [۳۵]. به‌عنوان مثال اگر سایز بردار خروجی برابر با ۴۰۰۰۰۰ باشد و فاکتور کمینه‌سازی برابر با ۴ باشد، خروجی نهایی برابر با $\frac{400000}{16} = 25000$ خواهد بود.

استخراج ویژگی سراسری هیستوگرام جهت لبه: برای استخراج این

ویژگی ابتدا تصویر دریافتی با فیلتر گاوسی هموارسازی می‌گردد. سپس از ۵ عملگر تشخیص لبه برای تشخیص لبه در جهت‌های مختلف استفاده می‌شود. در این الگوریتم برای تعیین نهایی لبه‌های تصویر از الگوریتم تشخیص لبه کنی^{۲۸} [۳۶] استفاده می‌گردد؛ به این صورت که در این مرحله فقط نقاطی که توسط الگوریتم کنی به‌عنوان لبه تشخیص داده شده‌اند حفظ می‌شوند و دیگر نقاط برچسب غیر لبه می‌گیرند و حذف می‌شود. الگوریتم EHD خود تصویر را به ۱۶ بخش تقسیم می‌نماید و برای هر بلوک یا بخش از تصویر با استفاده از ۵ عملکرد تشخیص لبه یک هیستوگرام ۵ بخشی تشکیل می‌دهد و سپس بردار ۱۶ بخش از هر تصویر را در کنار هم قرار داده و برداری به طول ۸۰ بعد تشکیل

آستانه برای تعیین فراوانی الگوی بافت است. در این مرحله نیز مشابه با مرحله قبل برای هر یک از بلوک‌های تصویر، هیستوگرام الگوی باینری محلی استخراج می‌شود. کد باینری محلی برای هر پیکسل از بلوک با همسایگی 3×3 محاسبه می‌شود، که در آن مقدار پیکسل مرکزی به‌عنوان آستانه‌ای برای مقایسه با ۸ پیکسل همسایه مورد استفاده قرار می‌گیرد. سپس با توجه به حداقل آستانه تعریف شده برای این مرحله، به پیکسل‌هایی با مقادیر بالاتر و یا برابر با آستانه، مقدار ۱ و به پیکسل‌هایی با مقادیر زیر آستانه، مقدار ۰ تخصیص داده می‌شود. مقادیر صفر و یک حاصل شده با توجه به موقعیت هر پیکسل با توانی از ۲ ضرب می‌شود، که توان اعداد ۲ موقعیت پیکسل مورد نظر می‌باشد. به عبارت ساده پیکسل ۱ با 2^1 ضرب خواهد شد. کدهای تخصیص داده شده به هر پیکسل از هر بلوک از مجموع حاصل ضرب فوق به دست می‌آیند، در نتیجه کدهای حاصل شده برای هر پیکسل از هر تصویر برای تشکیل هیستوگرام با ۲۵۶ ستون^{۲۵} استفاده می‌شود. در این مرحله هیستوگرام الگوی محلی هر بلوک از تصویر دریافت می‌شود، کدهای الگوی بافت بر اساس تعداد تکرار به‌صورت نزولی مرتب می‌شوند و کدهای الگوی بافتی که تعداد تکرارشان بیشتر از حداقل آستانه تعریف شده باشد با استفاده از "-" به یکدیگر پیوست می‌شوند. به‌عنوان مثال اگر بلوک ۱ دارای الگوهای بافت $256, 129, 64$ با تعداد تکرار بیشتر از حداقل آستانه باشد، آنگاه امضاء تشکیل شده برای آن که یک کد متنی می‌باشد برابر با T-256-129-64 خواهد بود. در شکل ۳ استخراج بافت محلی غالب نشان داده شده است. پس از استخراج رنگ و بافت محلی غالب، امضاء نهایی تشکیل شده برای هر بلوک از تصویر، متشکل از دو کد متنی می‌باشد که برای رنگ با حرف C و برای بافت با حرف T شروع می‌شوند. در نهایت هر تصویر با مجموعه‌ای از کدهای متنی مربوط به بلوک‌های آن نمایش داده می‌شود. در مرحله بعد به‌صورت یک سند با هر تصویر برخورد می‌شود و از تکنیک مدل فضای بردار و وزن‌دهی TF-IDF برای تشکیل بردار محلی نهایی هر تصویر استفاده می‌شود. امضاء نهایی بلوک ۱ در شکل ۳ نشان داده شده است.



شکل ۳- استخراج ویژگی بافت محلی غالب [۴]

برای بررسی عملکرد مدل‌های بازیابی تصویر، غالباً از معیار دقت، میانگین متوسط دقت^۱، فراخوان و میانگین f استفاده می‌شود. در این مطالعه نیز از سه معیار دقت و میانگین متوسط دقت و همچنین زمان بازیابی تصاویر برای ارزیابی مدل استفاده شده است که به ترتیب در روابط (۶) و (۷) ارائه شده‌اند [۳۷]. برای ارزیابی الگوریتم از دو پایگاه داده Wang و Oxford استفاده شده است.

$$precision = \frac{relevant\ images \cap retrieved\ images}{retrieved\ images} \quad (6)$$

$$MAP = \frac{\sum_{q=1}^Q AveP(q)}{Q} \quad (7)$$

در رابطه بالا میانگین دقت یا به‌اختصار (AveP) برای هر تصویر پرس‌وجو با در نظر گرفتن میانگین مقادیر دقت در هر تصویر بازیابی شده مرتبط به دست می‌آید. برای مجموعه تصاویر پرس‌وجو Q، میانگین دقت متوسط (MAP) میانگین مقادیر AP برای هر تصویر پرس‌وجو q است.

• پایگاه Wang: زیرمجموعه ۱۰۰۰ تصویر از مجموعه داده‌های تصویر Corel است. این پایگاه داده شامل ۱۰ کلاس و هر کلاس از ۱۰۰ تصویر تشکیل می‌شود. کلاس‌هایی همچون دایناسور، اتوبوس، ساحل، فیل، گل از جمله برجسب کلاس‌های موجود در این پایگاه داده است. این پایگاه داده از آدرس Wang [۳۸] قابل دریافت است.

• پایگاه Oxford: این پایگاه داده مجموعه‌ای از ۵۰۶۲ تصویر مختلف از جمله تصاویر ساختمان‌ها، جمع‌آوری شده از سایت flicker است و از آدرس OX [۳۹] قابل دریافت می‌باشد؛ این مجموعه داده شامل سه کلاس good, ok و Junk است.

۴-۱- بررسی بهترین مقادیر پارامترها

الگوریتم SDLCTG دارای سه پارامتر تعداد بلوک و α متغیر ترکیب خطی دو ویژگی محلی و تعداد تصاویر بازیابی است که تعیین مقدار مناسب برای هر یک از آن‌ها در نتایج بازیابی مؤثر خواهد بود؛ شکل‌های ۴ تا ۶ نتایج آزمون‌هایی است که مقادیر مناسب برای پارامترهای فوق را ارائه می‌دهد. برای بررسی تعداد بلوک، تصاویر به بلوک‌های 4×4 ، 8×8 ، 16×16 ، 25×25 و 36×36 تقسیم‌بندی شده‌اند. برای انجام این آزمون متغیر α برابر با 0.5 تعیین شد و دو حداقل آستانه مربوط به رنگ محلی و بافت محلی به ترتیب برابر با 0.5 و 0.3 تعیین شد. در این آزمون تعداد ۱۰ پرس‌وجو به‌صورت تصادفی از پایگاه داده Wang انتخاب شده و نتایج میانگین دقت ۱۰ تصویر پرس‌وجوی فوق محاسبه شده و در شکل ۴ گزارش شده است. برای یافتن بهترین ترکیب دو ویژگی محلی با استفاده از متغیر α به این متغیر مقادیر 0.03 ، 0.05 ، 0.07 ، 0.09 و 1 تخصیص داده شده و مقدار میانگین دقت برای ۱۰ پرس‌وجوی مختلف برای آن محاسبه شده و در شکل ۵ گزارش شده است.

همان‌طور که در شکل ۴ مشاهده می‌شود بهترین مقدار میانگین دقت در تعداد بلوک 25×25 به‌دست آمده است. در این حالت برای هر تصویر 625 بلوک تشکیل می‌شود و از هر بلوک یک کلمه بصری استخراج می‌گردد که در مجموع تعداد 625 کلمه برای هر ویژگی محلی در هر تصویر حاصل می‌شود. نتایج بالا نشان می‌دهد که تقسیم‌بندی تصویر به بلوک‌های کوچک تا حدی باعث بهبود نتایج می‌شود اما اگر سایز بلوک‌ها بیش از اندازه کوچک شود نتایج روند کاهشی خواهد داشت. به این دلیل که در این حالت الگو باینری محلی در تکنیک بافت محلی غالب، از بخش مورد نظر استخراج نمی‌شود در نتیجه کلمات بصری برای بسیاری از بخش‌های تصویر تشکیل نخواهد شد و در واقع اطلاعات ناقصی از تصویر تشکیل می‌شود. شکل ۵ نیز نشان می‌دهد که بیشترین مقدار میانگین دقت در مقدار α برابر با 0.7 حاصل شده است. این نتایج نشان می‌دهد، در صورتی که

می‌دهد. در واقع خروجی الگوریتم هیستوگرام جهت لبه در روش پیشنهادی برای هر تصویر یک بردار 1×80 بعدی است. خروجی نهایی این الگوریتم برای کلیه تصاویر، ماتریسی است که تعداد سطرهای آن برابر با تعداد تصاویر پایگاه داده و تعداد ستون‌های آن برابر با $80 (5 \times 16)$ است.

استخراج ویژگی سراسری هیستوگرام گرادیان جهت‌دار: هیستوگرام گرادیان جهت‌دار اطلاعات شکل محلی را از مناطق یا نقاط در یک تصویر استخراج می‌کند. در واقع توصیفگر HOG، تعداد رخداد جهت گرادیان را در بخش‌های محلی تصویر (پنجره تشخیص^۲) می‌شمارد. این توصیفگر ابتدا تصویر را به مناطق متصل کوچک به نام سلول‌ها تقسیم می‌کند و سپس برای هر سلول یک هیستوگرام جهت گرادیان یا جهت لبه برای پیکسل‌های درون سلول محاسبه می‌کند. سایز پیش‌فرض سلول در الگوریتم HOG 8×8 است. با توجه به جهت گرادیان، هر سلول به طبقه^۳‌هایی تقسیم می‌شود. تعداد طبقه پیش‌فرض در این الگوریتم برابر با ۹ است. پیکسل‌های متعلق به هر سلول، گرادیان وزن داری را به طبقه مرتبط یا سلول اضافه می‌کنند. سپس سلول‌های مجاور گروه‌بندی شده و بلوک‌های بزرگ‌تر را تشکیل می‌دهند. گروه‌بندی سلول‌ها به یک بلوک، اساس گروه‌بندی و نرمال‌سازی نمودار هیستوگرام است. سایز پیش‌فرض بلوک در HOG برابر با 2×2 است. بلوک‌ها در الگوریتم HOG با یکدیگر همپوشانی دارند و هر سلول بیش از یک‌بار در بردار ویژگی نهایی درج می‌شود. گروهی از هیستوگرام‌های نرمال شده، هیستوگرام بلوک را نشان می‌دهد و مجموعه‌ای از این هیستوگرام بلوک توصیفگر هیستوگرام گرادیان جهت‌دار را ارائه می‌دهند. طول بردار نهایی خروجی HOG وابسته به سایز تصاویر است.

پس از استخراج سه ویژگی گابور، هیستوگرام جهت لبه و هیستوگرام گرادیان جهت‌دار از کل تصویر به‌صورت سراسری، ماتریس نهایی تصاویر از پیوست ماتریس وزن TF-IDF دو ویژگی محلی رنگ و بافت غالب و سه ویژگی فوق حاصل می‌شود. هدف از ترکیب ویژگی‌های سراسری با ویژگی‌های محلی بررسی تأثیر ویژگی‌های سراسری در فرآیند بازیابی است. از این‌رو سه ویژگی سراسری مستخرج از تصاویر بدون هیچ‌گونه پیش‌پردازشی با ویژگی‌های محلی پیوست شده‌اند. همچنین با توجه به اینکه این سه ویژگی از کل تصویر استخراج می‌شوند امکان استفاده از تکنیک کیف کلمات همچون ویژگی‌های محلی برای آن‌ها وجود ندارد در نتیجه از اعمال این تکنیک در این سه ویژگی نیز صرف‌نظر گردید. پس از تشکیل ماتریس ویژگی، ماتریس فوق برای محاسبه تشابه تصاویر با استفاده از تشابه کسینوسی استفاده می‌گردد.

۳-۲- محاسبه تشابه و بازیابی تصویر مشابه

آخرین مرحله از الگوریتم SDLCTG بازیابی تصاویر مشابه با تصویر پرس‌وجو با استفاده از تشابه کسینوسی است. در این مرحله تصاویر پرس‌وجو دریافت می‌شود و فاصله بین بردار تصویر پرس‌وجو با بردار کلیه تصاویر موجود در پایگاه داده با استفاده از رابطه (۵) که رابطه تشابه کسینوسی می‌باشد، محاسبه می‌شود. سپس مقادیر فاصله حاصل شده به‌صورت صعودی مرتب‌سازی می‌شود و از لیست شاخص‌های تصاویر مرتب شده به تعداد Rank تعیین شده توسط کاربر، K تصویر اول از لیست فوق به کاربر نمایش داده می‌شود. در رابطه (۵) x و y دو بردار ویژگی متعلق به دو تصویر هستند.

$$d_{\cosine}(x, y) = \frac{x \cdot y}{\|x\| \|y\|} \quad (5)$$

۴- تحلیل و بررسی نتایج روش

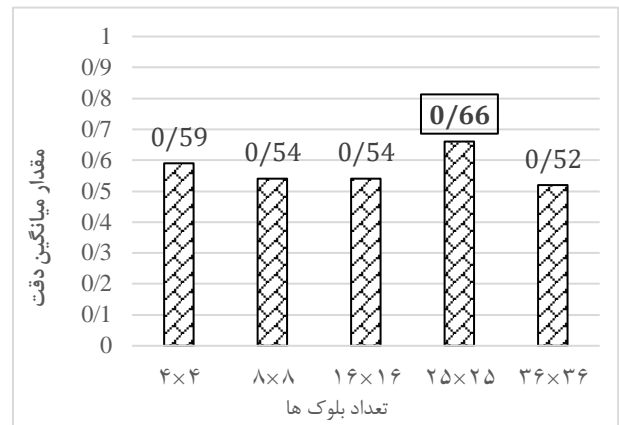
در این بخش معیار ارزیابی و خصوصیات مجموعه داده، و نتایج آزمون‌های مختلف که نشان دهنده عملکرد الگوریتم می‌باشند ارائه خواهد شد.

نتایج بالا نشان می‌دهد که اگرچه تأثیر تعداد تصویر بازبایی شده برای پرس‌وجوهای مختلف متفاوت است، اما به‌طور کلی افزایش تعداد تصاویر بازبایی شده باعث کاهش دقت کلی مدل می‌شود. به این دلیل که با افزایش تعداد تصاویر بازبایی شده، تصاویر با مقدار تشابه کم نیز در لیست نهایی درج می‌شوند که در واقع با تصویر پرس‌وجو دارای کلاس یکسانی نیستند و این مسئله باعث کاهش دقت در هر تصویر و در نتیجه کاهش میانگین دقت در چندین تصویر خواهد شد. در نمودار بالا نیز این مسئله به وضوح مشاهده می‌شود که با افزایش تعداد تصاویر بازبایی شده دقت کلی مدل در هر دو پایگاه داده Wang و Oxford کاهش داشته است.

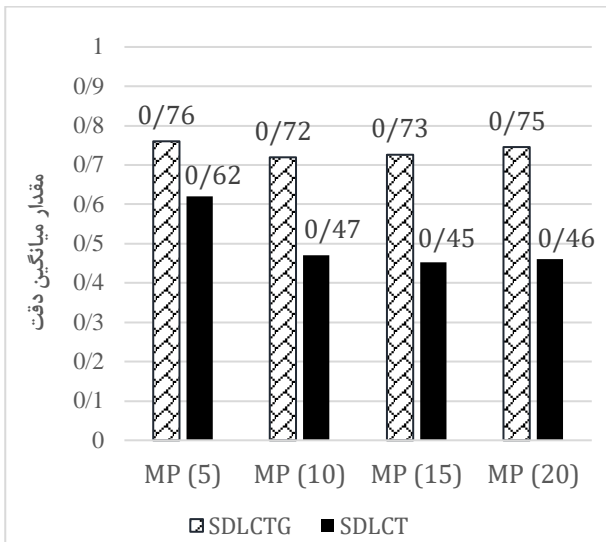
۲-۴- مقایسه نتایج روش‌ها

این آزمایش‌ها عملکرد الگوریتم SDLCTG را با روش پایه [۴] مقایسه می‌کند. در شکل‌های ۷ و ۸ میانگین دقت روش‌ها در تعداد ۵، ۱۰، ۱۵ و ۲۰ پرس‌وجو مختلف در دو پایگاه داده Wang و Oxford گزارش شده است. در کلیه این آزمایش‌ها دقت مدل‌ها در ۱۰ تصویر بازبایی شده (P@10) محاسبه شده و ارائه شده است. همچنین تعداد بلوک برابر با ۲۵×۲۵ و α نیز برابر با ۰/۷ مقداردهی شده‌اند.

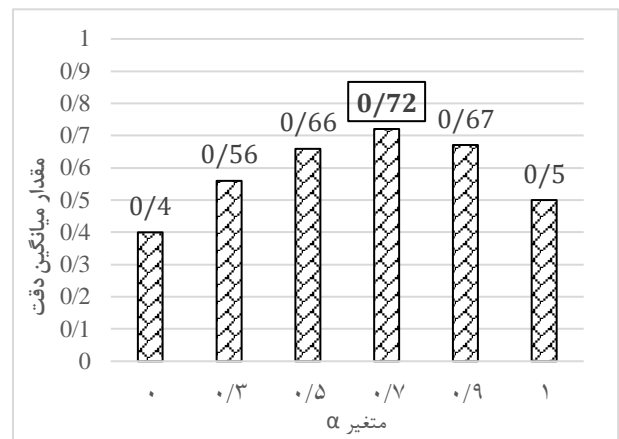
ویژگی رنگ محلی غالب وزن بیشتری در مقایسه با بافت محلی غالب بگیریید نتایج کلی مدل بهتر خواهد شد. در مقدار $\alpha = 0$ فقط بافت محلی غالب با ویژگی‌های سراسری ترکیب شده است که نتایج مناسبی به دست نیامده است و در مقدار $\alpha = 1$ فقط ویژگی رنگ محلی غالب با ویژگی‌های سراسری ترکیب شده است که در مقایسه با حالت قبل نتایج بهتری حاصل شده است. در آخرین آزمایش این بخش تعداد تصاویر مشابه بازبایی شده برای ۵ تصویر پرس‌وجوی مختلف بین ۵ تا ۵۰ متغیر در نظر گرفته شده است. نتایج این آزمایش برای دو پایگاه داده در شکل ۶ گزارش شده است. برای انجام این آزمون متغیر α برابر با ۰/۷ و تعداد بلوک‌های تصاویر ۲۵×۲۵ است.



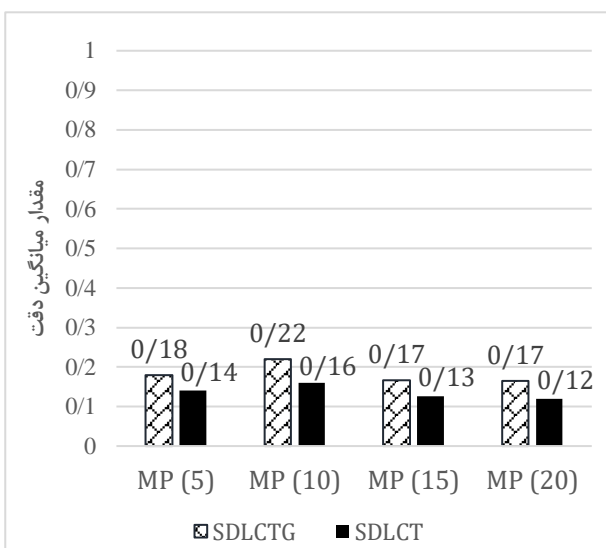
شکل ۴- بررسی تعداد بلوک بر نتایج بازبایی



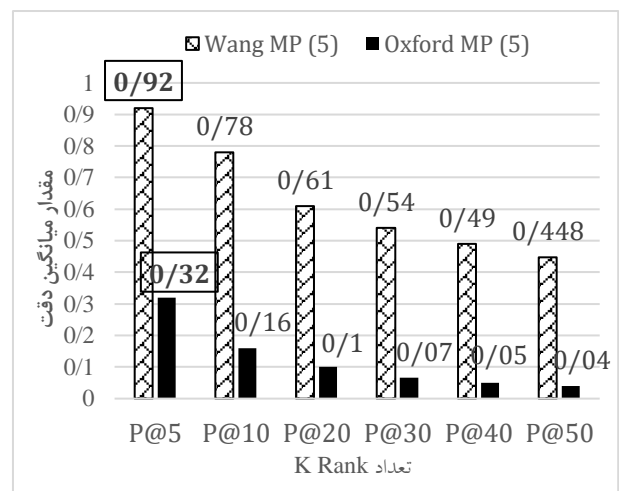
شکل ۷- مقایسه میانگین دقت برای ۵-۱۰-۱۵-۲۰ پرس‌وجو در Wang



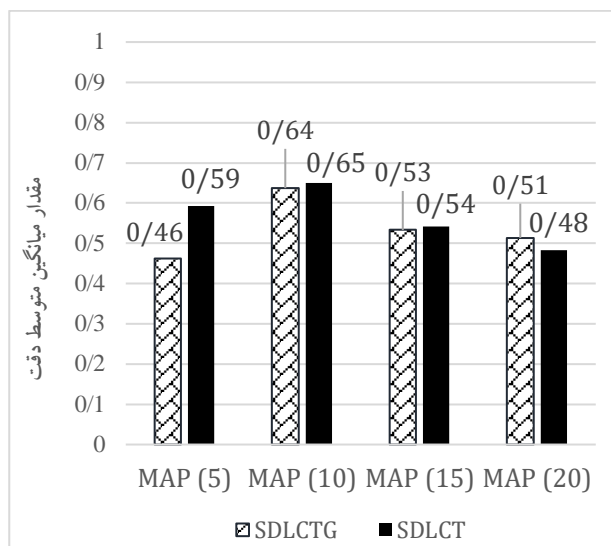
شکل ۵- بررسی متغیر α در ترکیب دو متغیر محلی



شکل ۸- مقایسه میانگین دقت برای ۵-۱۰-۱۵-۲۰ پرس‌وجو در Oxford



شکل ۶- بررسی تأثیر تعداد تصاویر بازبایی شده در دو پایگاه داده



شکل ۱۰- مقایسه میانگین متوسط دقت برای ۵-۱۰-۱۵-۲۰ پرس‌وجو در Oxford

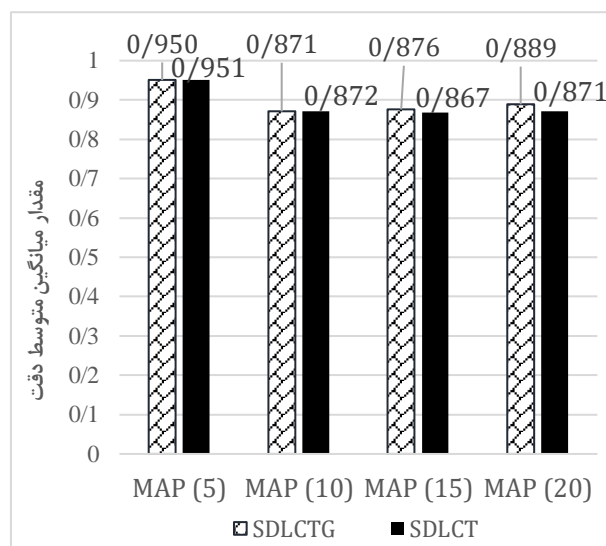
زمان بازیابی الگوریتم SDLCTG در پایگاه داده Wang و Oxford به ترتیب ۱۲،۶۶ و ۵۱،۲۹ ثانیه و زمان بازیابی روش پایه [۴] در دو پایگاه داده فوق به ترتیب ۸،۱۱ و ۴۶،۲۱ ثانیه است که نشان می‌دهد زمان بازیابی الگوریتم پیشنهادی در مقایسه با روش پایه [۴] بیشتر است. علت این نتایج این است که طول بردار ویژگی‌های الگوریتم پیشنهادی SDLCTG در مقایسه با روش پایه بیشتر می‌باشد و این مسئله باعث می‌شود بررسی تشابه بین بردارهای ویژگی‌های تصاویر در روش پیشنهادی در مقایسه با روش پایه بیشتر شود؛ علاوه بر این ماتریس وزن روش پایه، یک ماتریس خلوت و غیر مترامک است که زمان محاسبات در آن کمتر است. در پایگاه داده Oxford زمان هر دو روش نسبت به پایگاه داده Wang بیشتر است. این مسئله مرتبط با سائز تصاویر است، پایگاه داده Oxford دارای تصاویر بزرگی است که در روش پایه بالغ‌بر یک میلیون کلمه بصری از کل تصاویر استخراج می‌شود. در روش پیشنهادی نیز طول بردار خروجی دو ویژگی گابور و هیستوگرام گرادین جهت‌دار وابسته به سائز تصاویر است و با توجه به اینکه تصاویر Oxford بزرگ‌تر از Wang است در نتیجه طول بردار نهایی برای هر تصویر در آن بیشتر می‌شود و این مسئله باعث می‌گردد زمان محاسبه تشابه بین بردارهای ویژگی در پایگاه داده Oxford بیشتر از Wang باشد.

۵- نتیجه‌گیری و کارهای آتی

در این مقاله یک الگوریتم بازیابی تصویر با نام SDLCTG ارائه شد. در این روش با استفاده از دو ویژگی رنگ و بافت محلی غالب کیف کلمات بصری مبتنی بر امضا تشکیل گردید، سپس سه ویژگی گابور و هیستوگرام جهت لبه و هیستوگرام گرادین جهت‌دار به‌طور سراسری از کل تصاویر استخراج شد؛ که با ترکیب این دو دسته ویژگی، بردارهای ویژگی نهایی تصاویر تشکیل گردید. در نهایت با استفاده از رابطه تشابه کسینوسی تصاویر مشابه با تصویر پرس‌وجو بازیابی شد. نتایج ارزیابی نشان داد که روش پیشنهادی در پایگاه داده Wang در هر دو معیار دقت و Map در مقایسه با روش پایه [۴] نتایج مطلوب‌تری ارائه می‌دهد. در پایگاه داده Oxford در معیار دقت نتایج روش SDLCTG در مقایسه با روش پایه بهتر بود اما در میانگین متوسط دقت بهبودی چندانی حاصل نشد. زمان بازیابی روش ارائه شده نیز در هر دو پایگاه داده در مقایسه با روش پیشین [۴] بیشتر بود؛ به این دلیل که طول بردارهای ویژگی الگوریتم SDLCTG به دلیل استخراج چندین ویژگی مختلف و مخصوصاً ویژگی‌هایی که طول بردار آن‌ها وابسته به سائز تصاویر است، بیشتر می‌باشد. از این‌رو در راستای بهبود نتایج روش پیشنهادی می‌توان از

بهترین مقدار دقت در بین تصاویر مختلف برای روش پیشنهادی به ترتیب در پایگاه Wang و Oxford ۱ و ۰/۴ و برای روش پایه [۴] به ترتیب ۰/۹ و ۰/۳ است. نتایج ارائه شده در شکل ۷ نشان می‌دهد که تعداد پرس‌وجو تأثیر چندانی در نتایج الگوریتم SDLCTG ندارد. میانگین دقت الگوریتم SDLCTG در ۲۰ تصویر پرس‌وجو برابر با ۰/۷۵ و در ۵ تصویر برابر با ۰/۷۶ می‌باشد که اختلاف چندانی با هم ندارند. نتایج فوق نشان می‌دهد که ترکیب ویژگی‌هایی همچون هیستوگرام جهت لبه و گابور با ویژگی‌های کیف کلمات رنگ و بافت می‌تواند دقت بازیابی را بهبود بخشد. پایگاه داده Oxford دارای سه کلاس good و Junk و Ok است که در رابطه با کیفیت تصویر است، و در واقع نشان دهنده محتوای بصری تصویر نیست. از این‌رو در بسیاری از تصاویر پرس‌وجو این پایگاه، تصاویر بازیابی شده دارای محتوای بصری مشابهی با پرس‌وجو است اما تصاویر به یک کلاس تعلق ندارند. از این‌رو همان‌طور که در شکل ۸ مشاهده می‌شود اگرچه نتایج حاصل شده در پایگاه Oxford در مقایسه با پایگاه داده Wang پایین‌تر است، اما همچنان الگوریتم پیشنهاد شده دارای عملکرد موفق‌تری در مقایسه با روش پایه [۴] است. نتایج بالا نشان می‌دهد که مقدار بهبود میانگین دقت در دو پایگاه داده Wang و Oxford به ترتیب ۲۳،۷۰٪ و ۴،۶۲٪ است که در واقع بیانگر میانگین بهبود ۱۴،۱۶٪ روش در مقایسه با روش پایه [۴] در دو پایگاه است. نتایج فوق حاکی از آن است که ترکیب کیف کلمات بصری مبتنی بر امضا و ویژگی‌های سراسری می‌تواند در بازیابی تصاویر به‌طور مؤثری عمل نماید به این دلیل که دو گروه ویژگی‌های مورد استفاده در این روش به‌خوبی توانسته‌اند دانش مناسبی از تصاویر استخراج کنند.

در شکل‌های ۹ و ۱۰ مقایسه نتایج میانگین متوسط دقت یا MAP گزارش شده است. نتایج در شکل ۹ بیانگر این واقعیت است که جایگاه تصاویر مرتبط بازیابی شده با تصویر پرس‌وجو در دو روش SDLCT و روش پیشنهادی SDLCTG تقریباً مشابه می‌باشد. با توجه به اینکه معیار MAP به رتبه‌بندی حساس است، در صورتی که تصاویر مشابه با تصویر پرس‌وجو در انتهای لیست رتبه‌بندی قرار گیرند نتایج آن نامطلوب و در غیر این صورت مناسب خواهد شد.



شکل ۹- مقایسه میانگین متوسط دقت برای ۵-۱۰-۱۵-۲۰ پرس‌وجو در Wang

در شکل ۱۰ در پایگاه داده Oxford، روش پایه [۴] نتایج بهتری در مقایسه با روش پیشنهادی داشته است. در این پایگاه فقط در مجموعه ۲۰ تصویر پرس‌وجو، روش پیشنهادی نتایج بهتری داشته است.

- [17] A. Bosch, X. Muñoz, and R. Martí, "Which is the best way to organize/classify images by content?," *Image and vision computing*, vol. 25, pp. 778-791, 2007.
- [18] D. Filliat, "A visual bag of words method for interactive qualitative localization and mapping," in *Robotics and Automation, 2007 IEEE International Conference on*, 2007, pp. 3921-3926.
- [19] M. L. A. Vidal, J. M. Cavalcanti, E. S. de Moura, A. S. da Silva, and R. da Silva Torres, "Sorted dominant local color for searching large and heterogeneous image databases," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 1960-1963.
- [20] J. Liu, "Image retrieval based on bag-of-words model," *arXiv preprint arXiv:1304.5168*, 2013.
- [21] N. Chaturani, S. Geva, V. Chandran, and T. Chappell, "Content based image retrieval using signature representation," in *12th Australasian Data Mining Conference*, 2014.
- [22] M. Alkhawli, M. Elmogy, and H. Elbakry, "Content-based image retrieval using local features descriptors and bag-of-visual words," *Int J Adv Comput Sci Appl*, vol. 6, pp. 212-219, 2015.
- [23] E. G. Karakasis, A. Amanatiadis, A. Gasteratos, and S. A. Chatzichristofis, "Image moment invariants as local features for content based image retrieval using the bag-of-visual-words model," *Pattern Recognition Letters*, vol. 55, pp. 22-27, 2015.
- [24] J. M. dos Santos, E. S. de Moura, A. S. da Silva, J. M. B. Cavalcanti, R. da Silva Torres, and M. L. A. Vidal, "A signature-based bag of visual words method for image indexing and search," *Pattern Recognition Letters*, vol. 65, pp. 1-7, 2015.
- [25] I. Dimitrovski, D. Koccev, S. Loskovska, and S. Džeroski, "Improving bag-of-visual-words image retrieval with predictive clustering trees," *Information Sciences*, vol. 329, pp. 851-865, 2016.
- [26] F. B. Silva, R. d. O. Werneck, S. Goldenstein, S. Tabbone, and R. d. S. Torres, "Graph-based bag-of-words for classification," *Pattern Recognition*, vol. 74, pp. 266-285, 2018.
- [27] D. Y. T. Chino, L. C. Scabora, M. T. Cazzolato, A. E. S. Jorge, C. Traina, and A. J. M. Traina, "ICARUS: retrieving skin ulcer images through bag-of-signatures," in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, 2018, pp. 82-87.
- [28] D. Y. Chino, L. C. Scabora, C. Traina Jr, and A. J. Traina, "Boss: Image retrieval using bag-of-superpixels signatures," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, 2018, pp. 309-312.
- [29] L. Kabbai, M. Abdellaoui, and A. Douik, "Image classification by combining local and global features," *The Visual Computer*, vol. 35, pp. 679-693, 2019.
- [30] A. Rashno and E. Rashno, "Content-based image retrieval system with most relevant features among wavelet and color features," *arXiv preprint arXiv:1902.02059*, 2019.
- [31] U. Sharif, Z. Mehmood, T. Mahmood, M. A. Javid, A. Rehman, and T. Saba, "Scene analysis and search using local features and support vector machine for effective content-based image retrieval," *Artificial Intelligence Review*, vol. 52, pp. 901-925, 2019.
- [32] A. Sarwar, Z. Mehmood, T. Saba, K. A. Qazi, A. Adnan, and H. Jamal, "A novel method for content-based image retrieval to improve the effectiveness of the bag-of-words model using a support vector machine," *Journal of Information Science*, vol. 45, pp. 117-135, 2019.
- [33] V. Geetha, V. Anbumani, S. Sasikala, and L. Murali, "Efficient hybrid multi-level matching with diverse set of features for image retrieval," *Soft Computing*, pp. 1-22, 2020.
- [34] S. Z. Jumani, F. Ali, I. A. Kandhro, Q. A. Lakhan, U. Ali, M. W. Haroon, and S. Ahmed, "Facial Emotion Identification Based on Local Binary Pattern Feature Detector," *Indian Journal of Science and Technology*, vol. 12, p. 28, 2019.
- [35] M. Haghghat, S. Zonouz, and M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," *Expert Systems with Applications*, vol. 42, pp. 7905-7916, 2015.
- [36] B. Green, "Canny edge detection tutorial," *Retrieved: March*, vol. 6, p. 2005, 2002.
- [37] A. Latif, A. Rasheed, U. Sajid, J. Ahmed, N. Ali, N. I. Ratyal, B. Zafar, S.H. Dar, M. Sajid, and T. Khalil, "Content-based image retrieval and feature extraction: a comprehensive review," *Mathematical Problems in Engineering*, vol. 2019, 2019.
- [38] "Wang," ed. <http://wang.ist.psu.edu/docs/related/>.
- [39] "OXBuilding," ed. <http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/>.
- روش‌های کاهش ابعاد همچون PCA^{۳۲} برای کاهش زمان بازیابی استفاده نمود. همچنین استفاده از روش‌های استخراج قواعد انجمنی از دیگر روش‌های مناسب جهت کاهش ابعاد ویژگی‌ها و انتخاب ویژگی‌های بهینه می‌باشد؛ استفاده از این تکنیک نیز می‌تواند باعث بهبود دقت و همچنین زمان بازیابی شود. به‌عنوان آخرین پیشنهاد، برای استخراج ویژگی‌های مناسب می‌توان از شبکه‌های عصبی عمیق کانولوشن پیش آموزش دیده همچون ResNet و Inception Net استفاده نمود. با توجه به اینکه این شبکه‌های بر روی تصاویر ImageNet آموزش دیده‌اند انتظار می‌رود نتایج ویژگی‌های استخراج شده توسط آن‌ها بتواند به خوبی دقت مدل را بهبود ببخشد.

۶- مراجع

- [1] M. Rehman, M. Iqbal, M. Sharif, and M. Raza, "Content based image retrieval: survey," *World Applied Sciences Journal*, vol. 19, pp. 404-412, 2012.
- [2] A. Sandhu and A. Kochhar, "Content based image retrieval using texture, color and shape for image analysis," *International Journal of Computers & Technology*, vol. 3, pp. 149-152, 2012.
- [3] T. W. Cai, J. Kim, and D. D. Feng, "Content-based medical image retrieval," in *Biomedical information technology*, ed: Elsevier, 2008, pp. 83-113.
- [4] J. M. Dos Santos, E. S. De Moura, A. S. Da Silva, and R. da Silva Torres, "Color and texture applied to a signature-based bag of visual words method for image retrieval," *Multimedia Tools and Applications*, vol. 76, pp. 16855-16872, 2017.
- [5] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Past, present, and future," in *Journal of Visual Communication and Image Representation*, 1997.
- [6] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data," *Journal of digital imaging*, vol. 26, pp. 1025-1039, 2013.
- [7] C. B. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: current status and future directions," *Journal of digital imaging*, vol. 24, pp. 208-222, 2011.
- [8] T. Mei, Y. Wang, X.-S. Hua, S. Gong, and S. Li, "Coherent image annotation by learning semantic distance," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8.
- [9] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, pp. 1075-1088, 2003.
- [10] D. M. Blei and M. I. Jordan, "Modeling annotated data," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, 2003, pp. 127-134.
- [11] Z.-J. Zha, X.-S. Hua, T. Mei, J. Wang, G.-J. Qi, and Z. Wang, "Joint multi-label multi-instance learning for image classification," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8.
- [12] S. Zhang, J. Huang, Y. Huang, Y. Yu, H. Li, and D. N. Metaxas, "Automatic image annotation using group sparsity," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 3312-3319.
- [13] J. Tang, H. Li, G.-J. Qi, and T.-S. Chua, "Image annotation by graph-based inference with integrated multiple/single instance representations," *IEEE Transactions on Multimedia*, vol. 12, pp. 131-141, 2010.
- [14] Y. Wang, T. Mei, S. Gong, and X.-S. Hua, "Combining global, regional and contextual features for automatic image annotation," *Pattern Recognition*, vol. 42, pp. 259-266, 2009.
- [15] D. Aldavert, M. Rusiñol, R. Toledo, and J. Lladós, "A study of Bag-of-Visual-Words representations for handwritten keyword spotting," *International Journal on Document Analysis and Recognition (IJ DAR)*, vol. 18, pp. 223-234, 2015.
- [16] T. Deselaers, L. Pimenidis, and H. Ney, "Bag-of-visual-words models for adult image classification and filtering," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 2008, pp. 1-4.

زهرا هادیان برسیانی دانشجوی کارشناسی ارشد
مهندسی کامپیوتر گرایش هوش مصنوعی در دانشگاه آزاد
اسلامی واحد اصفهان (خوراسگان)، می باشد.
آدرس پست الکترونیکی ایشان عبارت است از:
maryamhadian72@gmail.com



فرساد زمانی بروجنی در سال ۱۳۹۲ با معدل ۱۹,۳۷ از
دانشگاه یو پی ام مالزی فارغ التحصیل شد و در همان سال
موفق به کسب مدال نقره در مسابقات کشوری
دانشگاه های مالزی از وزارت علوم این کشور شد. او
هم اکنون به عنوان استادیار تمام وقت و مدیرکل آموزش با
دانشگاه آزاد اسلامی واحد اصفهان (خوراسگان) همکاری دارد.



آدرس پست الکترونیکی ایشان عبارت است از:

f.zamani@khuisf.ac.ir

-
- 1 Content-based Image Retrieval
 - 2 Text-based Image Retrieval
 - 3 Photometric
 - 4 Geometric
 - 5 Texture Co-Occurrence Matrix
 - 6 Contours
 - 7 Point sets
 - 8 Surfaces
 - 9 Bag of Visual Words (BoVW)
 - 10 Signature-Based Bag of Visual Words (S-BoVW)
 - 11 Random Indexing
 - 12 Predictive Clustering Trees
 - 13 Random Forests of Predictive Clustering Trees
 - 14 Local Color and texture Dominant
 - 15 Bag of Singleton Graphs
 - 16 Bag of Visual Graphs
 - 17 Bag-of-Super pixel Signatures
 - 18 Norm
 - 19 Ant Colony Optimization (ACO)
 - 20 Scale - Invariant Feature Transform
 - 21 Binary Robust Invariant Scalable Key points
 - 22 Local Intensity Order Pattern
 - 23 Local Binary Pattern Variance
 - 24 Sorted Dominant Local Color and Texture with Global Features
 - 25 Bin
 - 26 Sparse
 - 27 Down Sampling
 - 28 Canny
 - 29 Detection window
 - 30 bin
 - 31 Mean Average Precision
 - 32 Principle Component Analysis

Improving RGB Image Retrieval Using Color, Texture and Shape Applied To a Signature-Based Bag of Visual Words Method

Zahra Hadian Bersiani, Farsad Zamani Boroujeni

Faculty of Engineering, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran

Abstract

Today, producing and storing digital images has led to emerging of content-based image retrieval systems. Among many types of image retrieval systems, those systems that work based on the content-based methods provide comprehensive functionality and exhibit higher performance results. However, one of the factors that impairs the success of these systems is extracting inappropriate features from the images. That means, extracting incomplete and inconsistent features often result in significant reduction in the performance of the system. One of the methods that recently attracted the attention of researchers is called Signature-Based Bag of Visual Words which quantizes the features without using a clustering technique. Even though its superiority to Bag-of-Words is reported in recent studies, it requires to partition images to small patches to obtain a set of local features for constructing vocabulary words in different images. In other words, in Signature-Based Bag-of-Words, global features are ignored making the method unsuitable for many types of images. Therefore, our proposed method uses a combination of three efficient global features including Gabor descriptor, edge histogram descriptor (EHD) and histogram of oriented gradients (HOG) as well as two texture based features to obtain sufficient image descriptors and improve the accuracy of image retrieval. The results of evaluating the method on Wang and Oxford databases show that the proposed method outperforms the traditional techniques in terms of precision of the retrieval.

Keywords: Content-based retrieval, Signature-Based Bag of Words, Edge Histogram Descriptor, Histogram of Oriented Gradients, Gabor descriptor.