بازیابی تصویر مبتنی بر محتوا با استفاده از همجوشی نتایج چندسطحی

اکبر مقیمیان[۱]، دانشجوی کارشناسی ارشد؛ محرم منصوری زاده[۲]، استادیار؛ میرحسین دزفولیان[۳]، استادیار

۱- دانشکده فنی و مهندسی - گروه کامپیوتر - دانشگاه بوعلی - همدان - ایران – akbarmoghimian@gmail.com
۲- دانشکده فنی و مهندسی - گروه کامپیوتر - دانشگاه بوعلی - همدان - ایران – mansoorm@basu.ac.ir
۳- دانشکده فنی و مهندسی - گروه کامپیوتر - دانشگاه بوعلی - همدان - ایران – dezfoulian@basu.ac.ir

چکیده: بازیابی تصویر مبتنی بر محتوا استفاده از روش‌های بینایی ماشین برای بازیابی تصاویری از یک مجموعه است که به تصویر پرس‌وجو شبیه باشند. چالش اصلی این سیستم‌ها کاهش شکاف معنایی بین ویژگی‌های سطح‌پایین مستخرج از پیکسل و قطعه تصویر و مفاهیم سطح‌بالای موجود در آن است. یکی از روش‌های کاهش این فاصله استفاده از ویژگی‌های سطح‌بالای مستخرج از نواحی و اشیا برای بازیابی است. از طرفی ویژگی‌های سطح‌پایین نیز تمایز خوبی بین خود تصاویر اعمال می‌کنند. بر این اساس انتظار می‌رود استفاده از هر دو دسته ویژگی به نتایج بهتری منجر شود. در این پژوهش بازیابی تصویر در چهار سطح پیکسل، ناحیه، شیء و مفهوم انجام‌شده است و از همجوشی نتایج این سطوح به‌منظور کاهش شکاف معنایی استفاده‌شده است. در سطح پیکسل، از ویژگی‌های SIFT و LBP استفاده‌شده است. در سطح ناحیه، ابتدا تصویر به چند ناحیه افراز و سپس ویژگی‌های رنگ و بافت با استفاده از توصیفگر Hue و فیلتر گابور از هر یک از نواحی تصویر استخراج‌شده است. در سطح شیء از شبکه عصبی کانولوشنی AlexNet برای بازشناسی اشیاء و صحنه‌های درون تصویر و در سطح مفهوم از شبکه عصبی Word2vec برای سنجش شباهت معنایی تصاویر استفاده‌شده است. نتایج بازیابی روی دو پایگاه داده Wang و GHIM نشان‌دهنده بهبود دقت و فراخوانی در بازیابی تصویر است.

واژه‌های کلیدی: بازیابی تصویر مبتنی بر محتوا، همجوشی اطلاعات، ترکیب طبقه‌بندها، AlexNet، Word2vec.

Content Based Image Retrieval by Fusion of Multilevel Results

A. Moghimian[1], MSc Student; M. Mansoorizadeh[2], Assistant Professor; M. H. Dezfoulian[3], Assistant Professor

1- Faculty of Engineering, Bu-Ali Sina University, Hamedan, Iran, Email: akbar.moghimina@gmail.com
2- Faculty of Engineering, Bu-Ali Sina University, Hamedan, Iran, Email: mansoorm@basu.ac.ir
3- Faculty of Engineering, Bu-Ali Sina University, Hamedan, Iran, Email: dezfoulian@basu.ac.ir

Abstract: Content based image retrieval (CBIR) applies machine vision techniques to extract similar images for a given query image. The main challenge of CBIR is the semantic gap between low level pixel and segment based features and high-level concepts in the image. An approach towards reducing this gaps is to use high level region and object based features. However, the low-level features describe image details and enforce between image discriminations. Accordingly, it is expected that the use of both feature types will lead to better results. This paper tries to reduce the mentioned gap by combining decision results at four granularities, namely pixel, region, object, and concept levels. Pixel level retrieval adopts SIFT features and local binary patterns. Region level subsystem partitions the image into a set of segments and extracts their color and texture features using hue descriptor and Gabor filters for subsequent processing. AlexNet convolutional neural network is employed for object based retrieval. Word2vec embedding is used for concept level retrieval that exploits conceptual relations between objects to enhance the retrieval results. Experiments over Wang and GHIM datasets confirm the feasibility of the proposed combination and conclude that it improves overall performance of the retrieval system.

Keywords: Content based Image Retrieval, Information Fusion, Classifier Combination, AlexNet, Word2vec.

تاریخ ارسال مقاله: ۱۳۶۹/۸/۲۶
تاریخ اصلاح مقاله: ۱۳۹۶/۱۲/۱۸و ۱۳۹۷/۳/۱۲
تاریخ پذیرش مقاله: ۱۳۹۷/۴/۳
نام نویسنده مسئول: محرم منصوری زاده
نشانی نویسنده مسئول: ایران - همدان - چهارباغ شهید مصطفی احمدی روشن - دانشگاه بوعلی - دانشکده فنی و مهندسی - گروه کامپیوتر.

۱ -مقدمه

با توجه به رشد روزافزون تصاویر دیجیتال و کاربرد آن‌ها در زمینه‌های مختلف عمومی و تخصصی، در سال‌های اخیر بازیابی و جستجوی تصاویر هم در حوزه‌های پژوهشی و هم در حوزه‌های تجاری اهمیت زیادی را به خود اختصاص داده است. روش‌های ابتدایی بازیابی تصویر بر اساس متن عمل می‌کردند به‌طوری‌که به هر تصویر مجموعه‌ای از کلیدواژه‌ها به‌عنوان برچسب اختصاص داده می‌شد و کاربر می‌توانست بر اساس این کلیدواژه‌ها، در پایگاه داده‌ای برچسب خورده به جستجوی تصاویر موردنظر خود بپردازد [۱، ۲]. ازجمله اشکالات این روش می‌توان به هزینه‌بر بودن، تفاوت در تفسیر تصاویر توسط افراد مختلف و خطای انسانی اشاره کرد. علاوه بر این ممکن است برخی از تصاویر با یک یا چند کلیدواژه قابل توصیف نباشند. برای غلبه بر این مشکلات، روش‌های بازیابی تصویر مبتنی بر محتوا راه‌حل مناسبی هستند. ایده کلی این روش‌ها استفاده از محتوای تصویر به‌جای برچسب‌هاست [۳].

به‌طورکلی سیستم‌های بازیابی تصویر مبتنی بر محتوا دارای دو گام اصلی *نمایه‌سازی*[۱] و *جستجو* هستند. در گام نمایه‌سازی برای هر تصویر موجود در پایگاه داده یک بردار ویژگی که بیان‌کننده ویژگی‌های بصری آن است، محاسبه و در پایگاه داده ویژگی‌ها ذخیره می‌شود. در گام جستجو یک تصویر پرس‌وجو داده‌شده و بردار ویژگی مربوط به آن محاسبه می‌شود. سپس با استفاده از یک معیار شباهت، میزان شباهت بردار مربوط به تصویر پرس‌وجو و بردارهای متناظر با تصاویر پایگاه داده محاسبه می‌شود و با مرتب‌سازی نزولی این لیست، مجموعه تصاویر مشابه با تصویر پرس‌وجو بازیابی می‌شوند [۴].

مهم‌ترین چالش در بازیابی تصاویر بر اساس محتوای آن‌ها اختلاف بین ویژگی‌های سطح‌پایین استخراج‌شده از تصاویر و مفاهیم سطح‌بالای موجود در آن‌هاست که آرنولد اسملدرز و همکارانش در [۵] از آن با عنوان «شکاف معنایی[۲]» نام‌برده‌اند. یک مفهوم سطح‌بالا درکی است که یک انسان از یک تصویر دارد. درواقع دو تصویر ممکن است ازنظر رنگ و بافت شباهت بالایی داشته باشند ولی ازنظر مفهومی نتوان آن‌ها را مشابه دانست [۶].

از مهم‌ترین گام‌هایی که تاکنون برای کاهش شکاف معنایی ارائه‌شده است استفاده از یک گام قطعه‌بندی قبل از مرحله استخراج ویژگی از تصاویر است که به دلیل نزدیک‌تر شدن به سیستم بینایی انسان انجام‌شده است [۷]. همچنین می‌توان به استفاده از روش‌های طبقه‌بندی تصاویر[۳] [۸] اشاره کرد. هدف از طبقه‌بندی، این است که یک تناظر بین ویژگی‌های سطح‌پایین و مفاهیم سطح‌بالا که همان گروه‌های معنایی تصاویر هستند ایجاد شود. از روش‌های یادگیری ماشین نیز برای استفاده از بازخورد ربط[۴] و اعمال نظرات کاربران در مورد میزان ارتباط تصویر پرس‌وجو و تصاویر بازیابی‌شده در فرایند جستجو استفاده می‌شود. [۹] این بازخوردها می‌تواند میزان اهمیت بردارهای ویژگی در تصاویر مختلف را تغییر دهد.

روش‌های قطعه‌بندی به خاطر عملکرد نادقیقی که از سختی ذاتی این مسئله ناشی می‌شود، در بسیاری از موارد آن‌طور که انتظار می‌رود عمل نمی‌کنند.

روش‌های بازخورد ربط نیز نیاز به تعامل با کاربران دارند و برای همه سیستم‌های بازیابی مناسب و قابل‌استفاده نیستند. ممکن است در یک سیستم بینایی ماشین نیاز به بازیابی تصاویر داشته باشیم و اصلاً کاربر انسانی برای ارسال بازخورد ربط وجود نداشته باشد. شاید این دلایل انگیزه استفاده از بازیابی تصویر چندسطحی را افزایش دهد.

در این پژوهش ابتدا استخراج ویژگی از تصاویر در چهار سطح پیکسل، ناحیه، شیء و مفهوم انجام‌شده است. سپس در هر یک از این سطوح با استفاده از یک معیار شباهت یا فاصله مناسب، نتایج بازیابی محاسبه‌شده و از همجوشی این نتایج در سطح تصمیم به‌منظور کاهش شکاف معنایی استفاده‌شده است. دو مزیت در بازیابی تصویر چندسطحی وجود دارد. مزیت اول این است که توصیف تصاویر با استفاده از ویژگی‌های سطح‌بالا به مفاهیم درون ذهن انسان نزدیک‌تر است و انتظار می‌رود به کاهش شکاف معنایی کمک کند. مزیت دوم این است که در مواردی که استفاده از ویژگی‌های سطح‌بالا نمی‌تواند تمایزی بین گروه‌های مختلف تصویر پایگاه داده پیداکرده و نتایج مورد انتظار را تولید کند می‌توان با ترکیب آن با سطوح پایین‌تر، این کمبود را جبران نمود.

بیشتر کارهای انجام‌شده در حوزه بازیابی تصویر مبتنی بر محتوا، ویژگی‌های بصری را در سطح پیکسل و ناحیه استخراج می‌کنند. نوآوری اصلی این پژوهش شامل استخراج ویژگی در چهار سطح پیکسل، ناحیه، شیء و مفهوم و استفاده از همجوشی این اطلاعات در سطح تصمیم، برای پیاده‌سازی یک سیستم بازیابی تصویر چندسطحی است. در سطح پیکسل از یکی از روش‌های پیشنهادشده در [۱۰] که نتیجه بهتری دارد استفاده‌شده است. در سطح ناحیه از روش پیشنهادشده در [۱۱] استفاده‌شده است با این تفاوت که این روش با مدل کیفی از لغات[۵] ترکیب‌شده است درحالی‌که در [۱۱] از روش بازخورد ربط و یک معیار فاصله متفاوت استفاده‌شده است. در سطح شیء از روش استخراج ویژگی پیشنهادشده در [۱۲] الهام گرفته‌شده است با این تفاوت که شناسایی‌کننده‌های اشیاء[۶] ماشین بردار پشتیبان[۷] (SVM) هستند و با استفاده از ویژگی‌های عمیق لایه کاملاً متصل دوم یک شبکه عصبی کانولوشنی آموزش داده‌شده‌اند. همچنین در این پژوهش نحوه ساخت بردارهای ویژگی با استفاده از پاسخ به‌دست‌آمده از شناسایی‌کننده‌های اشیاء با روش پیشنهادشده در [۱۲] متفاوت است. برای استخراج ویژگی در سطح مفهوم نیز یک روش جدید پیشنهادشده است که مشابه آن در کارهای پیشین انجام‌نشده است.

ساختار ادامه مقاله بدین‌صورت است: در بخش دوم کارهای شاخص مرتبط در سطوح مختلف معرفی و مرور شده‌اند. در بخش سوم

روش پیشنهادی به همراه جزئیات توضیح داده می‌شود. سپس در بخش چهارم مجموعه تصاویر استفاده‌شده و آزمایش‌های انجام‌شده روی آن‌ها همراه با جزئیات پیاده‌سازی مانند مقادیر پارامترها و همچنین نتایج به‌دست‌آمده ارائه داده خواهد شد. در بخش پنجم نیز به بحث و نتیجه‌گیری پرداخته خواهد شد.

۲ مروری بر کارهای مرتبط

در این بخش کارهای مرتبط با توجه به سطح استخراج ویژگی به چهار سطح پیکسل، ناحیه، شیء و مفهوم تقسیم می‌شوند. این کارهای مرتبط در حوزه‌های بازیابی، طبقه‌بندی و برچسب‌زنی تصاویر[۸] هستند. در روش‌های قدیمی‌تر استخراج ویژگی از همه پیکسل‌های تصویر صورت می‌گرفت و این ویژگی‌های سطح‌پایین غالباً توصیف‌کننده‌های رنگ و بافت بودند [۱۳، ۱۴]. پس‌ازآن به‌منظور دست‌یابی به نتایج بهتر، از روش‌هایی همچون افراز تصویر به بلوک‌های برابر و استخراج ویژگی از هر یک از بلوک‌ها استفاده شد [۱۵]. در کارهای جدیدتر از توصیفگرهای محلی مانند تبدیل ویژگی مستقل از مقیاس[۹] (SIFT) که ویژگی‌ها را از مجموعه‌ای از همسایگان برخی پیکسل‌ها استخراج می‌کنند و روش کیفی از لغات با الهام از بازیابی متن در بازیابی تصویر استفاده‌شده است [۱۰، ۱۶]. در [۱۷] روشی برای طبقه‌بندی نقاشی‌های هنرمندان ایرانی با استفاده از ویژگی‌های هیستوگرام گرادیان جهت‌دار[۱۰] (HOG) و الگوی باینری محلی[۱۱] (LBP) پیشنهادشده است. در پژوهش [۱۷] پس از استخراج هر یک از ویژگی‌ها و الحاق آن‌ها به یکدیگر، یک SVM برای طبقه‌بندی تصاویر آموزش داده می‌شود. در یکی از روش‌های پیشنهادشده در [۱۰] ابتدا ویژگی SIFT از تصاویر پایگاه داده استخراج‌شده و با استفاده از مدل کیسه‌ای از واژه‌های دیداری کتاب کد[۱۲] مربوط به این ویژگی با V1 واژه ساخته‌شده است. سپس تصاویر به سلول‌های ۱۶×۱۶ تقسیم‌شده و هیستوگرام‌های LBP برای هر یک از سلول‌ها محاسبه‌شده است. کتاب کد مربوط به این ویژگی نیز مانند ویژگی SIFT و با V2 واژه ساخته‌شده است. سپس هر دو کتاب کد با یکدیگر ادغام شده‌اند و هر برای هر تصویر یک هیستوگرام با V1+V2 ستون به‌دست‌آمده است. تصاویر نیز با استفاده از معیار شباهت اشتراک هیستوگرام با یکدیگر مقایسه شده‌اند. در [۱۸] از ویژگی الگوی اکسترمم محلی جهت‌دار در بازیابی تصویر استفاده‌شده است. این ویژگی از LBP الهام گرفته‌شده است و برای استخراج اطلاعات لبه‌ها در چهار جهت افقی (صفر درجه)، قطر فرعی (۴۵ درجه)، عمودی (۹۰ درجه) و قطر اصلی (۱۳۵ درجه) است و برای هر پیکسل از تصویر چهار عدد تولید می‌کند. به این شکل که هر پیکسل با دو همسایه مجاورش مقایسه می‌شود و اگر از هردوی آن‌ها بزرگ‌تر یا از هردوی آن‌ها کوچک‌تر بود، یک «یک» و در غیر این صورت یک «صفر» درج می‌شود و با ادامه این روند برای هشت همسایه پیکسل مربوطه یک عدد نه بیتی به دست می‌آید. با تکرار این روش برای هر چهار جهت، چهار عدد نه بیتی به ازای هر پیکسل از تصویر به دست می‌آید. با محاسبه هیستوگرام روی اعداد مربوط به هر یک از جهت‌ها، چهار بردار ویژگی برای هر تصویر محاسبه خواهد شد. در [۱۹] از هیستوگرام‌های فضای رنگی HSV و یک ویژگی بافت محلی که از LBP الهام گرفته‌شده است، در بازیابی تصویر مبتنی بر محتوا استفاده‌شده است. در این پژوهش ابتدا تصویر به فضای رنگی HSV منتقل می‌شود. سپس کانال V تصویر به بلوک‌های ۲×۲ که دارای هم‌پوشانی هستند، تقسیم می‌شود. سپس برای هر یک از این بلوک‌ها طبق هشت الگوی متفاوت با اندازه ۲×۲، یک عدد در نظر گرفته می‌شود. با این روش یک تصویر فیلتر شده به دست می‌آید که ابعاد آن نصف ابعاد تصویر اصلی است. در تصویر اصلی هر پیکسل با هشت همسایه اطرافش مقایسه می‌شود و در صورت نابرابر بودن یک «یک» و در صورت مساوی بودن یک «صفر» به‌جای آن درج می‌شود. با تبدیل این عدد دودویی به دهدهی، برای هر پیکسل از تصویر فیلتر شده یک عدد به دست می‌آید. با محاسبه یک هیستوگرام برای اعداد به‌دست‌آمده، یک بردار ویژگی برای هر تصویر به دست می‌آید که لبه‌های موجود در تصویر را به‌خوبی توصیف می‌کند.

سطح دوم، استخراج ویژگی در سطح ناحیه است. در این سطح ابتدا تصویر با استفاده از یک روش قطعه‌بندی به چند ناحیه افراز می‌شود و در مرحله بعد ویژگی‌های بصری موردنظر از هر یک از نواحی استخراج می‌شوند. در [۷] ابتدا تصویر به بلوک‌های ۴×۴ افراز شده و ویژگی‌های رنگ و بافت برای هر بلوک استخراج‌شده است. سپس با اعمال خوشه‌بندی k-means روی بردارهای ویژگی استخراج‌شده و تقسیم آن‌ها به چند کلاس که هر کلاس متناظر با یک ناحیه از تصویر است، تصویر قطعه‌بندی شده و مرکز هر خوشه به‌عنوان بردار ویژگی مربوط به ناحیه متناظر با آن خوشه در نظر گرفته‌شده است. درنهایت با استفاده از معیار تطبیق یکپارچه نواحی[۱۳]، فاصله دو تصویر به‌دست‌آمده است. در [۲۰] پس از در نظر گرفتن یک مقیاس همسایگی مناسب برای هر پیکسل، بردارهای ویژگی رنگ و بافت برای همه پیکسل‌ها بر اساس همسایگی آن‌ها استخراج‌شده است. سپس به‌وسیله مدل کردن توزیع این ویژگی‌ها با یک مدل مخلوط گاوسی و استفاده از الگوریتم امید ریاضی بیشینه‌سازی[۱۴] (EM)، پیکسل‌ها به چندین ناحیه گروه‌بندی شده‌اند. برای تطبیق رنگ نواحی از فاصله ماهالانوبیس و برای محاسبه فاصله بین بافت آن‌ها از فاصله اقلیدسی استفاده‌شده و برای سنجش نهایی شباهت دو تصویر، این دو فاصله باهم ترکیب‌شده‌اند. در [۱۱] ابتدا تصویر با استفاده از الگوریتم Jseg [۲۱]، به نواحی همگنی از رنگ و بافت افراز شده و گشتاورهای رنگ هر ناحیه به‌عنوان ویژگی رنگ استخراج‌شده‌اند. سپس با در نظر گرفتن درصد مساحت هر ناحیه و بازخورد ربط به‌عنوان ضریب اهمیت آن ناحیه و استفاده از معیار فاصله محرک زمین[۱۵]، فاصله بین دو تصویر محاسبه‌شده است.

سطح سوم، استخراج ویژگی از تصاویر در سطح شیء است. در این سطح با کمک گرفتن از الگوریتم‌های بازشناسی اشیاء[۱۶] می‌توان یک

تصویر را با استفاده از اشیاء درون توصیف کرد و با این توصیف سطح‌بالا شکاف معنایی را کاسته و به‌دقت سیستم‌های بازیابی و طبقه‌بندی تصاویر افزود. از کارهای انجام‌شده در این سطح می‌توان به [۲۲] اشاره کرد که در آن کاربر یک شیء را با فراهم کردن مجموعه کوچکی از تصاویر شامل آن شیء به‌عنوان تصاویر پرس‌وجو مشخص می‌کند و سیستم بازیابی بر اساس یک مدل احتمالی، تصاویری که شامل آن شیء مخصوص هستند را برمی‌گرداند. همچنین در [۱۲] سیستمی برای طبقه‌بندی تصاویر پیشنهادشده است. این سیستم یک نمایش سطح‌بالا از تصاویر را بر اساس اشیاء درون آن‌ها ارائه می‌دهد و از این نمایش برای طبقه‌بندی اشیاء استفاده می‌کند. روش کار این سیستم به این شکل است که هر تصویر با ۱۲ مقیاس مختلف به تعداد زیادی الگوریتم شناسایی‌کننده شیء خاص‌منظوره و آموزش داده شده (برای ۲۰۰ شیء مختلف) داده می‌شود. سپس برای هر مقیاس و هر شناسایی‌کننده، یک نگاشت پاسخ اولیه از تصویر به دست می‌آید که از آن برای استخراج ویژگی جهت بازشناسی اشیاء مختلف درون تصویر استفاده‌شده است. در [۲۳] هر تصویر به شکل مجموعه‌ای از ویژگی‌ها نمایش داده می‌شود که هر ویژگی مربوط به یک ناحیه از تصویر است که می‌تواند از روش‌های مختلف قطعه‌بندی به دست آید. در مرحله اول الگوریتم EM خوشه‌هایی را در فضای ویژگی پیدا می‌کند که به‌احتمال‌زیاد مربوط به تصاویری هستند که حاوی یک شیء خاص هستند. این کار با استفاده از یک مدل مخلوط گاوسی چند متغیره انجام می‌گیرد که هر مؤلفه گاوسی آن، یک خوشه از بردارهای ویژگی را نشان می‌دهد. این مرحله همچنین دارای یک گام جهت نرمال‌سازی است زیرا هر تصویر می‌تواند شامل تعداد دلخواهی ناحیه باشد. در مرحله دوم یک طبقه‌بندی‌کننده[۱۷] یاد می‌گیرد که هر تصویر که با یک توصیف با طول ثابت نمایش داده شده است، شامل کدام دسته از اشیاء است. این طبقه‌بندی با استفاده از یک شبکه عصبی پرسپترون چندلایه[۱۸] (MLP) با سه لایه صورت می‌گیرد.

سطح چهارم، استخراج ویژگی و توصیف تصاویر در سطح مفهوم است. این سطح بالاترین سطح در بازیابی و طبقه‌بندی تصاویر است. در این سطح هدف اصلی این است که عملکرد و دقت سیستم‌های بازیابی و طبقه‌بندی تصاویر به سیستم بینایی و دقت تشخیص و درک انسان نزدیک‌تر شوند. در سطح مفهوم انتظار می‌رود تصاویر با استفاده از مفاهیم معنایی موجود در ذهن انسان توصیف شوند. در [۲۴] ابتدا کاربر یک تصویر را به‌عنوان پرس‌وجو به سیستم بازیابی ارسال می‌کند. سپس ویژگی‌های سطح‌پایین آن استخراج‌شده و بازیابی بر اساس بردار ویژگی آن تصویر انجام می‌شود. در مرحله بعد کاربر تصاویر بازیابی‌شده مرتبط را با استفاده از فرایند بازخورد ربط مشخص می‌کند. سپس یک خوشه‌بندی روی ویژگی‌های مربوط به تصاویر مرتبط صورت گرفته و مرکز هر خوشه به‌عنوان یک پرس‌وجوی جدید در نظر گرفته می‌شود. بعد از جستجوی پایگاه تصاویر با مرکز هر خوشه، برای هر تصویر پایگاه، کمترین فاصله از مراکز خوشه‌ها به‌عنوان میزان شباهت آن تصویر به بردار پرس‌وجو در نظر گرفته می‌شود. از دیگر کارهای موجود در سطح مفهوم می‌توان به [۲۵] اشاره کرد که هدف آن حل مشکل «تولید تعداد زیادی کلیدواژه نامرتبط» است که در برچسب‌زنی تصاویر با استفاده از روش «آنالیز تناظر بین ویژگی‌های بصری تصویر و کلیدواژه‌ها» به وجود می‌آید. روش پیشنهادشده در [۲۵] به این شکل است که با استفاده از پایگاه دانش WordNet، کلیدواژه‌های نامرتبط هرس شده‌اند. برای تشخیص کلیدواژه‌های نامرتبط با استفاده از WordNet، معیارهای شباهت معنایی مختلفی بین کلمات بررسی شده و نتایج این معیارها باهم ترکیب‌شده‌اند و تصمیم نهایی با استفاده از تئوری ترکیب شواهد دمپستر شفر[۱۹] گرفته‌شده است. در [۲۶] برای برچسب‌زنی تصاویر، ارتباط بین نواحی تصویر به‌صورت توزیع مفاهیم و موضوعات بصری مختلف بین نواحی مدل می‌شود. علاوه بر آن به‌منظور در نظر گرفتن مفهوم کلی تصویر از ویژگی‌های سراسری به‌صورت بردار توزیع موضوعات بصری نیز استفاده می‌شود.

۳ -روش پیشنهادی

هدف اصلی این مقاله بازیابی تصویر چهارسطحی است. روش کار به این شکل است که ابتدا نتایج پیاده‌سازی در هر چهار سطح پیکسل، ناحیه، شیء و مفهوم به‌دست‌آمده است. سپس برای درک تأثیر همجوشی نتایج هر سطح بازیابی با سطوح پایین‌تر، میانگین دقت برای بازیابی تصویر یک‌سطحی (سطح پیکسل)، دوسطحی (همجوشی سطوح پیکسل و ناحیه)، سه‌سطحی (همجوشی سطوح پیکسل، ناحیه، شیء) و چهارسطحی (همجوشی سطوح پیکسل، ناحیه، شیء و مفهوم) محاسبه‌شده است که در بخش آزمایش‌ها و نتایج به بیان و تحلیل آن پرداخته خواهد شد.

شکل ۱ شمای کلی روش پیشنهادی را نشان می‌دهد. در این شکل، تصویر پرس‌وجو به چهار سیستم بازیابی تصویر داده می‌شود و هر یک از این سیستم‌ها یک امتیاز برای هر یک از تصاویر پایگاه داده محاسبه می‌کنند. سپس امتیازهای به‌دست‌آمده از سطوح مختلف با استفاده از یک عمل میانگین باهم ترکیب‌شده و لیست امتیاز نهایی را برای همه تصاویر پایگاه داده تولید می‌کنند. با مرتب‌سازی این لیست امتیاز، شناسه تصاویر پایگاه داده به ترتیب و بر اساس میزان شباهت با تصویر پرس‌وجو به دست می‌آید.

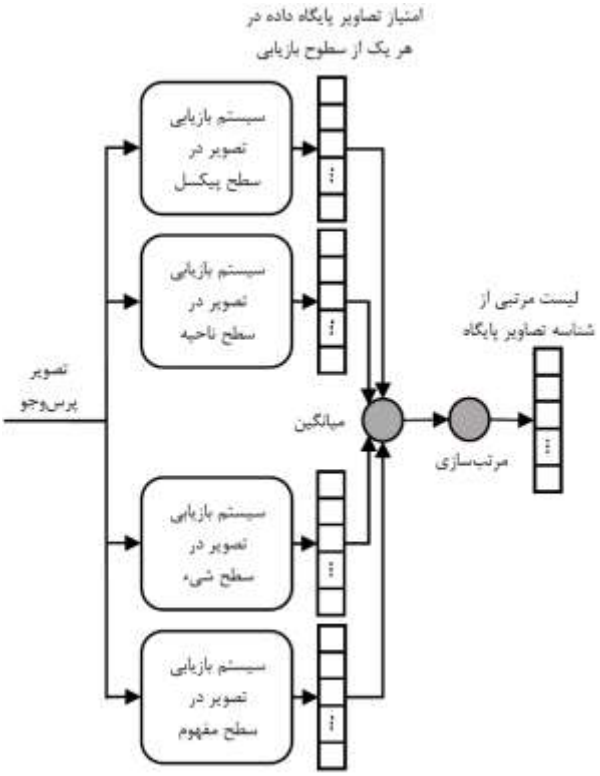
۳ ـ۱ -بازیابی تصویر مبتنی بر محتوا در سطح پیکسل

برای بازیابی تصویر در سطح پیکسل از مدل کیفی از لغات بصری استفاده‌شده است. در این مدل ابتدا ویژگی‌های محلی[۲۰] از تصاویر پایگاه داده استخراج می‌شوند. در مرحله بعد برای ساختن کتاب کد یا فرهنگ لغت، با استفاده از الگوریتم k-means یک خوشه‌بندی در فضای ویژگی انجام می‌شود و مراکز خوشه‌ها به‌عنوان لغات بصری در کتاب کد ذخیره می‌شوند. با اختصاص تعدادی از نزدیک‌ترین لغات بصری از کتاب کد به هر یک از ویژگی‌های محلی، هر تصویر به شکل کیفی از لغات بصری نمایش داده می‌شود. با محاسبه فراوانی نسبی هر

یک از این لغات، برای هر تصویر یک هیستوگرام ساخته می‌شود که به‌عنوان بردار ویژگی سراسری[۲۱] آن تصویر در پایگاه داده ویژگی‌ها ذخیره می‌شود. با استفاده از معیار شباهت اشتراک هیستوگرام (رابطه ۱) می‌توان میزان شباهت دو تصویر را اندازه‌گیری کرد:

$$\mathrm{sim}(A,B)=\sum_{i=1}^{n}\min(a_i,b_i) \qquad (۱)$$

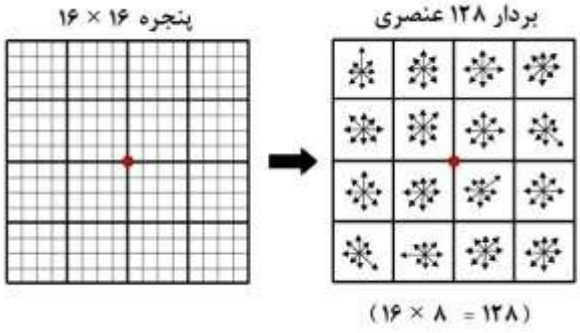
در رابطه ۱، A و B دو هیستوگرام با n ستون هستند و a_i و b_i فراوانی‌های نسبی ستون i-ام در آن‌ها را نشان می‌دهد. در مدل کیفی از لغات بصری معیار شباهت اشتراک هیستوگرام نسبت به سایر معیارها نتایج بهتری تولید می‌کند و به همین دلیل استفاده از آن رایج‌تر است.



شکل ۱: شمای کلی روش پیشنهادی

امروزه باوجود ویژگی‌های گوناگونی که برای توصیف رنگ، شکل و بافت تصاویر پیشنهادشده است، استفاده از ترکیب ویژگی‌های مختلف در پژوهش‌های بینایی ماشین بسیار رایج است. در روش پیشنهادی این پژوهش ابتدا ویژگی‌های SIFT و LBP را از تصاویر استخراج می‌شود و با ساختن کتاب کد با اندازه ۵۰۰ لغت، برای هر یک از این ویژگی‌ها دو هیستوگرام با ۵۰۰ ستون به‌دست‌آمده است و با الحاق این دو هیستوگرام، هر تصویر به هیستوگرامی با ۱۰۰۰ ستون تبدیل شده است. اندازه کتاب کد به‌صورت تجربی تعیین می‌شود. ازآنجایی‌که تعداد ویژگی‌های محلی SIFT و LBP که از هر تصویر استخراج می‌شود تقریباً برابر است، سهم هر دو از ویژگی از کتاب کد یکسان در نظر گرفته می‌شود.

ویژگی SIFT نسبت به چرخش، تغییر زاویه دید، بزرگنمایی، نویز، کشیدگی و نور تا حد زیادی غیر حساس است و در توصیف لبه‌ها و اشکال موجود در تصویر کارایی بسیار بالایی دارد. برای استخراج این ویژگی ابتدا نقاط کلیدی[۲۲] تصویر به دست می‌آیند. سپس هر یک از این نقاط با یک بردار ۱۲۸ عنصری توصیف می‌شوند. نقاط کلیدی تصویر نقاطی هستند که در تصاویر به‌دست‌آمده از اختلاف‌های گاوسی[۲۳]، اکسترمم هستند. این اختلاف‌های گاوسی با استفاده از تصاویر فضای مقیاس[۲۴] محاسبه می‌شوند. برای توصیف هر نقطه کلیدی یک پنجره ۱۶×۱۶ اطراف آن در نظر گرفته می‌شود که به ۱۶ بلوک ۴×۴ تقسیم می‌شود. سپس در هر بلوک ۴×۴ مقادیر و جهت‌های گرادیان محاسبه می‌شوند و در یک هیستوگرام با ۸ ستون قرار داده می‌شوند این ستون‌ها متناظر با صفر تا ۴۴ درجه، ۴۵ تا ۸۹ درجه، ۹۰ تا ۱۳۴ درجه و به همین ترتیب تا ۳۵۹ درجه هستند. با تکرار این رویه برای همه ۱۶ بلوک یک بردار ویژگی ۱۲۸ عنصری برای توصیف هر نقطه کلیدی به دست می‌آید. (شکل ۲)

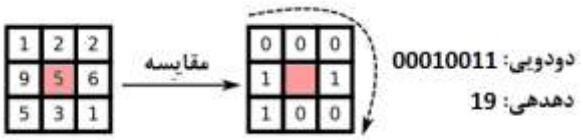


شکل ۲: تبدیل ویژگی مستقل از مقیاس (SIFT)

LBP یک توصیف‌کننده ویژگی بافت تصویر است. این ویژگی در کنار ویژگی SIFT می‌تواند شکل و بافت اشیاء درون تصویر را به‌خوبی توصیف کند. ازاین‌رو انتظار می‌رود ترکیب این دو ویژگی در بازیابی و طبقه‌بندی اشیاء کارایی مناسبی از خود نشان دهد. برای استخراج این ویژگی ابتدا تصویر سیاه‌وسفید به پنجره‌هایی ۱۶ در ۱۶ تقسیم می‌شود و در هر پنجره هر پیکسل با ۸ همسایه اطرافش مقایسه می‌شود. اگر ارزش پیکسل مرکزی از پیکسل همسایه‌اش بزرگ‌تر بود یک «صفر» و در غیر این صورت یک «یک» خواهیم داشت که این رویه منجر به تولید یک عدد دودویی هشت بیتی برای هر پیکسل می‌شود (شکل ۳). با به دست آوردن فراوانی هر یک از این اعداد یک بردار ویژگی ۲۵۶ عنصری برای هر یک از پنجره‌ها خواهیم داشت.

یک نسخه توسعه‌یافته از LBP، الگوی باینری محلی یکنواخت است که به‌منظور کاهش طول بردارهای ویژگی استفاده می‌شود. یک الگوی باینری محلی، یکنواخت است اگر تعداد گذرهای یک به صفر یا صفر به یک آن بیش از دو بار نباشد. برای مثال «۰۰۰۰۰۱۰۰» یک الگوی یکنواخت است و «۰۱۰۱۰۰۱۱» یکنواخت نیست. ازآنجایی‌که تعداد الگوهای یکنواخت ۵۸ تاست، طول بردار ویژگی به ۵۸ عنصر کاهش می‌یابد. البته می‌توان یک ستون هیستوگرام را نیز برای فراوانی

همه الگوهای غیریکنواخت در نظر گرفت که در این صورت طول بردار ویژگی ۵۹ عنصر خواهد بود. در این مقاله از بردارهایی با طول ۵۸ استفاده‌شده است.



شکل ۳: الگوی باینری محلی (LBP)

۳ ۲ -بازیابی تصویر مبتنی بر محتوا در سطح ناحیه

برای بازیابی تصویر در سطح ناحیه نیز از مدل کیفی از لغات بصری استفاده‌شده است. با این تفاوت که ویژگی‌های محلی در سطح ناحیه و از نواحی مختلف تصاویر قطعه‌بندی شده استخراج می‌شوند. در این سطح ابتدا تصاویر با روش قطعه‌بندی Jseg [۲۱] به نواحی همگنی از رنگ و بافت افراز می‌شوند. ایده اصلی الگوریتم Jseg جدا کردن فرایند قطعه‌بندی به دو مرحله است، کوانتیزه‌کردن رنگ و قطعه‌بندی فضایی. در مرحله اول رنگ‌های تصویر به چندین کلاس نماینده کوانتیزه می‌شوند که می‌تواند برای ایجاد تمایز بین نواحی تصویر مورد استفاده قرار گیرد. این کوانتیزه‌سازی در فضای رنگ و بدون در نظر گرفتن توزیع مکانی رنگ‌ها صورت می‌گیرد. سپس مقادیر پیکسل‌های تصویر با برچسب کلاس متناظرشان جایگزین می‌شوند و یک کلاس-نگاشت شکل می‌گیرد. کلاس-نگاشت می‌تواند به‌عنوان یک نوع خاص از ترکیب‌بندی بافت دیده شود. در مرحله دوم قطعه‌بندی مکانی به‌صورت مستقیم روی این کلاس-نگاشت و بدون در نظر گرفتن شباهت رنگ پیکسل‌های متناظر انجام می‌شود. در شکل ۴ نمونه‌هایی از قطعه‌بندی Jseg نشان داده شده است. مزیت این جداسازی دومرحله‌ای واضح است. تحلیل شباهت رنگ‌ها و توزیع مکانی آن‌ها به‌صورت هم‌زمان کار مشکلی است و جدا کردن این دو مرحله اجازه توسعه و استفاده الگوریتم‌های بهتری را در هر یک از مراحل می‌دهد [۲۱].

شکل ۴: تصاویر قطعه‌بندی شده با الگوریتم Jseg

پس از قطعه‌بندی تصویر، ویژگی‌های توصیفگر Hue و گابور به‌عنوان ویژگی‌های رنگ و بافت هر ناحیه استخراج می‌شوند. مطابق بخش ۳-۱ کتاب کد و هیستوگرام با اندازه ۵۰۰ برای هر دو نوع ویژگی به دست می‌آید و با الحاق هیستوگرام‌های هر دو ویژگی، هر تصویر به‌وسیله یک بردار ویژگی سراسری که هیستوگرامی با ۱۰۰۰ ستون است، نمایش داده می‌شود. برای سنجش شباهت نیز از همان معیار اشتراک هیستوگرام استفاده می‌شود. در این سطح نیز تعداد ویژگی‌های محلی Hue و گابور برای هر تصویر دقیقاً یکسان و برابر با تعداد ناحیه‌های آن است. درنتیجه سهم هر دو از ویژگی از کتاب کد یکسان در نظر گرفته می‌شود.

در فضای رنگ HSV، مقادیر Hue در اطراف محور خاکستری ناپایدار می‌شوند. برای این منظور در [۲۷] یک تجزیه‌وتحلیل خطا انجام‌شده است که نشان می‌دهد میزان قطعیت Hue با اشباع[۲۵] رابطه معکوس دارد؛ بنابراین هیستوگرام Hue با وزن دهی هر نمونه از پرده رنگ‌ها با میزان اشباع آن به یک ابزار قوی‌تری برای توصیف رنگ تبدیل می‌شود. در این پژوهش از هیستوگرام‌هایی با ۳۶ ستون به‌عنوان توصیف‌کننده رنگ استفاده‌شده است. به این شکل که ابتدا هر ناحیه به بلوک‌هایی با اندازه ۱۰×۱۰ پیکسل تقسیم‌شده است و متناظر با هر بلوک یک بردار ۳۶ عنصری (طول بردار به‌صورت تجربی به دست می‌آید) با استفاده از توصیفگر Hue به‌دست‌آمده است و درنهایت از میانگین این بردارها برای استخراج ویژگی رنگ هر ناحیه، در قالب یک بردار ویژگی ۳۶ عنصری استفاده‌شده است.

استفاده از فیلتر گابور یکی از رایج‌ترین روش‌های استخراج ویژگی بافت از تصاویر محسوب می‌شود [۱۴]. مهم‌ترین خاصیت این فیلترها مستقل بودن آن‌ها از دوران، مقیاس و انتقال است. این خاصیت فیلترهای گابور به این دلیل است که می‌توان آن‌ها را در مقیاس‌ها و جهت‌های مختلفی تعریف نمود که اصطلاحاً به آن بانکی از فیلترهای گابور گفته می‌شود. استخراج ویژگی بافت از یک تصویر با استفاده از این فیلترها به این صورت است که یک عمل کانولوشن بین تصویر سیاه‌وسفید و یک فیلتر گابور صورت می‌گیرد و انرژی هر پیکسل دریک همسایگی مربعی در اطراف آن محاسبه می‌شود؛ یعنی برای یک تصویر با ابعاد مشخص، فیلتر گابور با استفاده از عمل کانولوشن یک پاسخ تولید می‌کند که تصویر فیلتر شده‌ای هم‌اندازه با تصویر اصلی است. اگر میانگین و واریانس هر یک از نواحی را از برای تصاویر فیلتر شده و به ازای هر یک از فیلترهای بانکی با ۵ مقیاس و ۸ جهت به محاسبه شود، یک بردار ویژگی بافت دارای ۸۰ عنصر برای هر یک از نواحی تصویر به دست می‌آید.

۳ ۳ -بازیابی تصویر مبتنی بر محتوا در سطح شیء

یکی از قوی‌ترین ابزارهای فعلی برای طبقه‌بندی تصاویر و شناسایی اشیاء، شبکه‌های عصبی کانولوشنی است که از الگوی یادگیری عمیق برای آموزش استفاده می‌کند و اولین بار توسط یان لیکن معرفی شد. شبکه عصبی کانولوشنی ساختاری مشابه شبکه عصبی MLP دارد و دارای لایه‌های پنهان متعددی بین ورودی و خروجی است. این شبکه دارای سه نوع لایه است: لایه کانولوشنی، لایه ادغام[۲۶] و لایه کاملاً متصل. همچنین این شبکه با استفاده از تابع فعال‌سازی $f(x)=max(0,x)$ که یک تابع غیرخطی غیرقابل اشباع است و به خروجی همه لایه‌ها اعمال می‌شود، باعث افزایش سرعت آموزش در گرادیان نزولی می‌شود (نسبت به توابع غیرخطی قابل اشباع مانند $f(x)=tanh(x)$). درنهایت آنچه در خروجی نهایی این شبکه‌ها حاصل

می‌شود یک امتیاز نهایی برای هر یک از کلاس‌های موردنظر در مسئله طبقه‌بندی است.

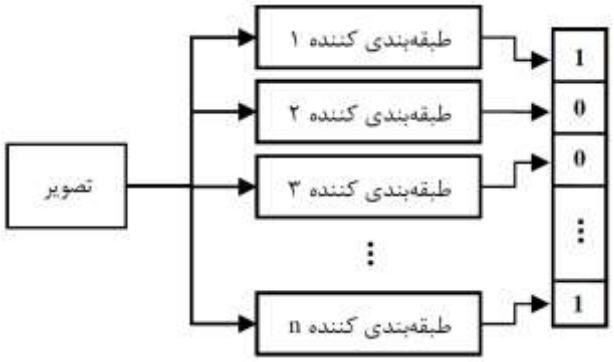
یکی از معروف‌ترین شبکه‌های عصبی کانولوشنی، AlexNet است که یادگیری آن با استفاده از بالغبر ۱۵ میلیون تصویر در قالب حدود ۲۲۰۰۰ گروه صورت گرفته است [۲۸]. این شبکه دارای هشت لایه است. در لایه اول ابتدا کانولوشن و فعال‌سازی، سپس نرمال‌سازی و درنهایت ادغام بیشینه روی تصویر ورودی انجام می‌شود. در لایه دوم نیز مجدداً ابتدا کانولوشن و فعال‌سازی، سپس نرمال‌سازی و درنهایت ادغام بیشینه روی خروجی لایه اول صورت می‌گیرد. در لایه‌های سوم و چهارم تنها کانولوشن و فعال‌سازی انجام می‌شود. لایه پنجم نیز مانند لایه‌های اول و دوم است با این تفاوت که عمل نرمال‌سازی انجام نمی‌شود. لایه‌های ششم تا هشتم نیز لایه‌های کاملاً متصل هستند که به ترتیب ۴۰۹۶، ۴۰۹۶ و ۱۰۰۰ نورون دارند. خروجی لایه هشتم به یک تابع Softmax داده می‌شود تا یک توزیع که همان امتیاز کلاس است را روی ۱۰۰۰ کلاس ایجاد کند. ضمناً آموزش این شبکه روی دو GPU انجام‌شده است و نیمی از وزن‌ها در GPU اول و نیمی دیگر در GPU دوم قرار داده‌شده‌اند تا مدت‌زمان آموزش کوتاه‌تر شود.

مطابق با روشی که در [۲۹، ۳۰] پیشنهادشده است، می‌توان از مقادیر فعال‌سازی هر یک از لایه‌های یک شبکه عصبی کانولوشنی از قبل آموزش داده شده به‌عنوان ویژگی‌های تصویر در مسئله طبقه‌بندی و شناسایی تصاویر و اشیاء استفاده نمود. با ورود یک تصویر به یک شبکه عصبی کانولوشنی، لایه‌های ابتدایی شبکه ویژگی‌های ساده و سطح‌پایین مانند رنگ، بافت و لبه‌ها و لایه‌های عمیق‌تر آن مانند لایه‌های کاملاً متصل، ویژگی‌های پیچیده و سطح‌بالای تصویر مانند چشم، صورت و پنجره‌های یک ساختمان را یاد می‌گیرند. از این ویژگی‌ها می‌توان برای آموزش یک طبقه‌بندی‌کننده استفاده کرد.

در این مقاله از شبکه عصبی آموزش داده شده AlexNet برای استخراج ویژگی استفاده‌شده است. تصاویر مجموعه آموزشی به شبکه عصبی AlexNet داده شده و از خروجی‌های لایه کاملاً متصل ماقبل آخر آن که بردارهای ۴۰۹۶ عنصری هستند به‌عنوان ویژگی‌های سطح‌بالای استخراج‌شده از تصاویر استفاده‌شده است. با این روش ۱۰۰ طبقه‌بندی‌کننده باینری مختلف برای ۱۰۰ گروه از اشیاء با استفاده از SVM و تصاویر ImageNet، آموزش داده شده است. برای آموزش هر SVM که وظیفه بازشناسی یکی از ۱۰۰ گروه از اشیاء را بر عهده دارد از ۱۰۰ تصویر مرتبط و ۱۰۰ تصویر غیر مرتبط استفاده‌شده است. برای توصیف اشیاء درون تصویر و استخراج ویژگی در سطح شیء، تصاویر پایگاه داده به هر یک از ۱۰۰ طبقه‌بندی‌کننده باینری داده شده و برای هر تصویر یک بردار ۱۰۰ عنصری با مقادیر صفر و یک به‌دست آمده است. مطابق با شکل ۵ هر تصویر با برداری باینری که بیانگر کیفی از اشیاء است توصیف می‌شود و با الهام از مدل کیفی از لغات در بازیابی متن از فاصله اقلیدسی به‌عنوان معیار عدم شباهت دو تصویر استفاده‌شده است:

$$d(A,B)=\sqrt{\sum_{i=1}^{n}(a_i^2-b_i^2)} \qquad (۲)$$

که در آن A و B دو بردار و a_i و b_i مؤلفه‌های i-ام آن‌ها هستند.



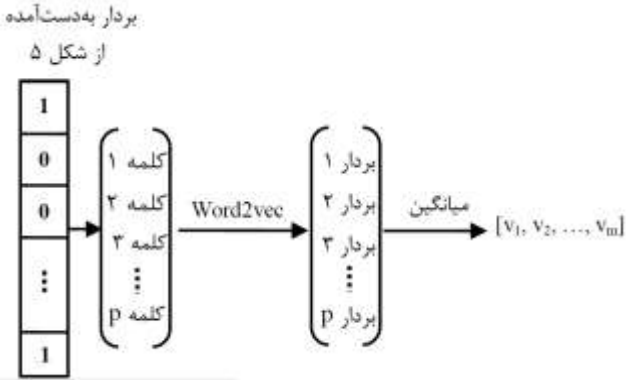
شکل ۵: نمایش برداری استفاده‌شده برای هر تصویر در بازیابی تصویر مبتنی بر محتوا در سطح شیء

۳ ۴ -بازیابی تصویر مبتنی بر محتوا در سطح مفهوم

مدل فضای برداری یکی از پرکاربردترین مدل‌های نمایش برداری لغات و اسناد متنی در پردازش زبان طبیعی، متن‌کاوی و بازیابی اطلاعات است. این مدل دو اشکال اساسی دارد: ۱(برای فرهنگ لغاتی با اندازه N، هر لغت، جمله یا سند با برداری به طول N که هر عنصر آن متناظر با یک لغت است، نمایش داده می‌شود که برای N های بزرگ به‌صرفه نیست. ۲) برای نمایش یک لغت برداری خواهیم داشت که N-1 عنصر آن صفر خواهد بود. برای حل این مشکلات، در سال ۲۰۱۳ روشی به نام Word2vec برای بازنمایی برداری لغات پیشنهاد شد [۳۱]. این روش با استفاده از آموزش یک شبکه عصبی سه لایه به‌وسیله میلیون‌ها لغت منحصربه‌فرد در اسنادی با میلیاردها لغت، بردارهایی منحصربه‌فرد و با یک طول ثابت برای هر کلمه تولید می‌کند. برای دست‌یابی به این هدف از دو مدل کیفی از لغات پیوسته و اسکیپ-گرام[۲۷] در شبکه‌های عصبی استفاده‌شده است. در مدل کیفی از لغات پیوسته، شبکه عصبی با گرفتن متن اطراف یک لغت آن را پیش‌بینی می‌کند ولی در مدل اسکیپ-گرام یک لغت به شبکه عصبی داده شده و شبکه لغات اطراف آن را پیش‌بینی می‌کند. این آموزش تا جایی ادامه می‌یابد که شبکه عصبی به یک همگرایی با کمترین خطای ممکن برسد. پس از پایان آموزش از وزن‌های لایه پنهان شبکه برای بازنمایی برداری هر لغت استفاده می‌شود.

در روش ارائه‌شده این مقاله برای بازیابی تصویر مبتنی بر محتوا در سطح مفهوم از بازنمایی برداری لغات گوگل شده است [۳۲]. این بازنمایی برداری با استفاده از آموزش یک شبکه عصبی سه لایه با ۳۰۰ نورون در لایه پنهان و به‌وسیله حدود ۱۰۰ میلیارد لغت موجود در گوگل نیوز به‌دست‌آمده است. با به دست آوردن بردار نمایش داده شده در شکل ۵ برای هر تصویر، درواقع هر تصویر به شکل کیفی از لغات نمایش داده شده است که در آن هر لغت متناظر باوجود شیء

متناظرش در تصویر است. در این مرحله با استفاده از بازنمایی برداری لغات گوگل، بردارهایی با ۳۰۰ عنصر برای هر لغت به دست می‌آید. سپس از میانگین بردارهای مربوط به یک تصویر، برای توصیف آن تصویر در سطح مفهوم استفاده می‌شود (شکل ۶). درنهایت مطابق با روش‌های بازیابی متن از فاصله اقلیدسی به‌عنوان معیار عدم شباهت دو تصویر در مرحله بازیابی استفاده‌شده است.



شکل ۶: نمایش برداری استفاده‌شده برای هر تصویر در بازیابی تصویر مبتنی بر محتوا در سطح مفهوم

۴ -نتایج تجربی

برای ارزیابی کارایی روش پیشنهادشده در این پژوهش از دو مجموعه تصویر Wang [۳۳] و GHIM [۳۴] استفاده‌شده است. مجموعه تصویر Wang شامل ۱۰۰۰ تصویر در ۱۰ گروه معنایی ۱۰۰ تایی است که ابتدا در [۷] مورد استفاده قرار گرفته است. مجموعه تصویر GHIM شامل ۲۰۰۰ تصویر در ۲۰ گروه معنایی است که در هر گروه ۱۰۰ تصویر قرار دارد. این تصاویر زیرمجموعه‌ای از تصاویری است که در [۳۵] استفاده‌شده است. از مجموعه تصاویر Wang برای مقایسه با روش پیشنهادی در این مقاله با کارهای پیشین استفاده‌شده است. به‌منظور بررسی بیشتر عملکرد بازیابی تصویر چندسطحی و بالا بردن اعتبار نتایج، عملکرد روش پیشنهادی بر روی مجموعه تصاویر GHIM نیز ارزیابی شده است.

در بازیابی تصویر در سطوح پیکسل و ناحیه اندازه کتاب کد با توجه به کارهای دیگران و به‌صورت تجربی به دست می‌آید که در این پژوهش ۱۰۰۰ در نظر گرفته‌شده است که سهم هر ویژگی ۵۰۰ خوشه است. برای خوشه‌بندی‌هایی با اندازه بزرگ‌تر از ۵۰۰ دقت بازیابی به نسبت زمان‌بر بودن نمایه‌سازی و جستجو افزایش چندانی ندارد. برای ساختن کیفی از لغات بصری در سطوح پیکسل و ناحیه می‌توان به هر بردار ویژگی یک یا چند لغت از کتاب کد را اختصاص داد که بهترین نتیجه در هر دو مجموعه تصویر زمانی به دست می‌آید که به هر بردار ویژگی سه لغت بصری با کمترین فاصله را اختصاص داد.

برای ارزیابی نتایج بازیابی از معیارهای مختلفی استفاده می‌شود که مهم‌ترین آن‌ها دقت[۲۸] (P) است. دقت بازیابی برای یک تصویر پرس‌وجو برابر است با نسبت تعداد تصاویر مرتبط بازیابی‌شده به تعداد کل تصاویر بازیابی‌شده. همچنین می‌توان دقت را بعد از k تصویر بازیابی‌شده با رتبه بالاتر محاسبه کرد و با نماد P@k نمایش داده می‌شود. برای ارزیابی دقت بازیابی در کل مجموعه تصویر، تک‌تک تصاویر پایگاه داده به‌عنوان تصویر پرس‌وجو در نظر گرفته می‌شوند و معیار دقت برای هر یک از آن‌ها محاسبه می‌شود. سپس میانگین دقت برای همه تصاویر پرس‌وجو به دست می‌آید. این معیار میانگین دقت[۲۹] (AP) نام دارد. معیار دیگر فراخوانی[۳۰] (R) نام دارد. فراخوانی برای یک تصویر پرس‌وجو برابر است با نسبت تعداد تصاویر مرتبط بازیابی‌شده به تعداد کل تصاویر مرتبط موجود در پایگاه داده. به همین ترتیب می‌توان فراخوانی را بعد از k تصویر بازیابی‌شده با رتبه بالاتر محاسبه کرد که با نماد R@k نمایش داده می‌شود. میانگین فراخوانی[۳۱] (AR) نیز برای فراخوانی بعد از k تصویر با رتبه بالاتر و محاسبه می‌شود.

معیار دقت، میزان ارتباط تصاویر بازیابی‌شده با تصویر پرس‌وجو را می‌سنجد درحالی‌که معیار فراخوانی تعیین می‌کند که چه کسری از تصاویر مرتبط موجود در پایگاه داده بازیابی شده است. منحنی دقت-فراخوانی (P-R) توازن بین دقت و فراخوانی را در آستانه‌های[۳۲] مختلف اندازه‌گیری می‌کند. هر چه مساحت زیر این منحنی بیشتر باشد، دقت و فراخوانی سیستم بازیابی تصویر بالاتر است و این یعنی سیستم بازیابی علاوه بر دقت بالا، درصد عمده‌ای از تصاویر مرتبط با تصویر پرس‌وجو را نیز برمی‌گرداند. در مواردی که از مجموعه‌ای از تصاویر پرس‌وجو برای ارزیابی کارایی استفاده می‌شود، از معیارهای AP و AR برای ترسیم منحنی P-R استفاده می‌شود.

برای بررسی عملکرد بازیابی تصویر چندسطحی، بازیابی یک‌سطحی بازیابی در سطح پیکسل تعریف می‌شود و در بازیابی دوسطحی از همجوشی نتایج سطوح پیکسل و ناحیه استفاده می‌شود. به همین ترتیب در بازیابی سه‌سطحی، نتایج سطوح پیکسل، ناحیه و شیء و در بازیابی چهارسطحی، نتایج هر چهار سطح را باهم ترکیب می‌شوند.

برای همجوشی نتایج در سطح تصمیم روش‌هایی مانند مجموع وزن‌دار و میانگین امتیازها، مجموع، میانگین، میانه، بیشینه و کمینه رتبه‌ها و همچنین عکس مجموع معکوس رتبه‌ها بسیار مرسوم هستند. در بین همه این روش‌ها مجموع وزن‌دار و میانگین امتیازها بهتر عمل می‌کنند و ازآنجایی‌که در روش مجموع وزن‌دار امتیازها، وزن‌هایی که به‌کاربرده می‌شوند وابسته به پایگاه داده تصاویر است، در این پژوهش از میانگین امتیازها استفاده‌شده است که روش کلی‌تری به‌حساب می‌آید. جدول ۱ میانگین دقت بازیابی روش‌های انجام‌شده در این پژوهش را در کنار چند نمونه از کارهای انجام‌شده قبلی به تصویر کشیده است.

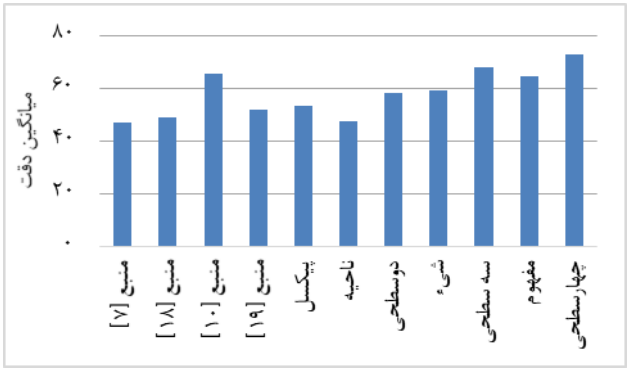
با توجه به این‌که در کارهایی که د جدول ۱ به آن‌ها ارجاع شده است از معیار AP برای ارزیابی کارایی استفاده‌شده است، در این جدول نیز از همین معیار برای مقایسه با کارهای پیشین استفاده‌شده است. در این جدول هر یک از سطرها بیان‌کننده میانگین دقت برای هر یک از گروه‌های معنایی تصاویر است و سطر آخر میانگین دقت را برای کل

مجموعه نشان می‌دهد که به نمودار این نتایج در شکل ۷ قابل‌مشاهده است. جزئیات کارهای پیشین موجود در جدول ۱ و شکل ۷ در بخش ۲ آمده است. در [۱۹] به میانگین دقت مربوط به هر گروه از تصاویر اشاره‌ای نشده است و این معیار فقط برای کل تصاویر گزارش شده است.

جدول ۱: مقایسه میانگین دقت به‌دست‌آمده در هر یک از سطوح بازیابی و کارهای پیشین برای مجموعه تصاویر Wang

گروه	نتایج کارهای پیشین				نتایج روش‌های پیشنهادی در سطوح مختلف بازیابی						
	[۷]	[۱۸]	[۱۰]	[۱۹]	پیکسل	ناحیه	دوسطحی	شیء	سه‌سطحی	مفهوم	چهارسطحی
آفریقا	۴۸	۳۹/۷	۵۷	-	۴۴/۹۳	۴۷	۵۶/۳۵	۲۴/۹۶	۴۷/۷۹	۵۱/۵	**۵۷/۷۳**
ساحل	۳۲	۳۷/۳	۵۸	-	۳۰/۲۹	۳۳/۵۴	۳۶/۸۹	۳۸/۹۷	۳۷/۲	۴۷/۳۱	**۴۷/۶۲**
ساختمان	۳۵	۳۴/۹	۴۳	-	۳۲/۹۵	۳۴/۱۷	۴۲/۲۸	۴۱/۹۲	۴۴/۶۹	۴۹/۰۲	**۵۲/۸۲**
اتوبوس	۳۶	۷۴/۱	۹۳	-	۸۵/۲۵	۴۰/۲۳	۸۵/۶۶	۸۸/۴۹	۹۵/۳۳	۸۸/۶۲	**۹۶/۳۲**
دایناسور	۹۵	۸۸	۹۸	-	۹۶/۸۶	۹۵/۳۹	۹۹	۸۳/۶۹	۹۹/۴۸	۷۶/۴۵	**۹۹/۵۵**
فیل	۳۸	۲۹	۵۸	-	۴۰/۸۶	۴۲/۴۴	۴۳/۳۵	۶۴/۹۸	۶۸/۵۱	۵۰/۷۳	**۶۵/۶۵**
گل	۴۲	۷۰/۸	۸۳	-	۶۸/۶۴	۴۱/۸۱	۶۴/۱۲	۹۰/۳	۹۰/۷۲	۸۵/۹۵	**۹۳/۵۱**
اسب	۷۲	۴۱/۷	۶۸	-	۵۴/۳۱	۶۸/۴۶	۶۱/۰۵	۵۲/۹۱	۷۲/۸۴	۶۶/۰۳	**۷۴/۵۹**
کوهستان	۳۵	۲۹	۴۶	-	۳۰/۸۶	۳۶/۰۹	۳۹/۶۵	۴۴/۷	۴۹/۵۲	۵۵/۸۹	**۶۱/۱۵**
غذا	۳۸	۴۷	۵۳	-	۴۷/۵۷	۳۷/۵۶	۵۳/۱۵	۶۲/۳۴	۷۳/۱۷	۷۵/۱۳	**۸۰/۱۵**
میانگین	**۴۷/۱**	**۴۹**	**۶۵/۷**	**۵۲**	**۵۳/۲۵**	**۴۷/۶۸**	**۵۸/۱۵**	**۵۹/۳۳**	**۶۷/۹۲**	**۶۴/۴۶**	**۷۲/۹**

علاوه بر آن روش‌های بازیابی تصویر در هر یک از سطوح پیکسل، ناحیه، شیء و مفهوم نیز در بخش‌های ۳-۱ الی ۴-۳ توضیح داده شد.



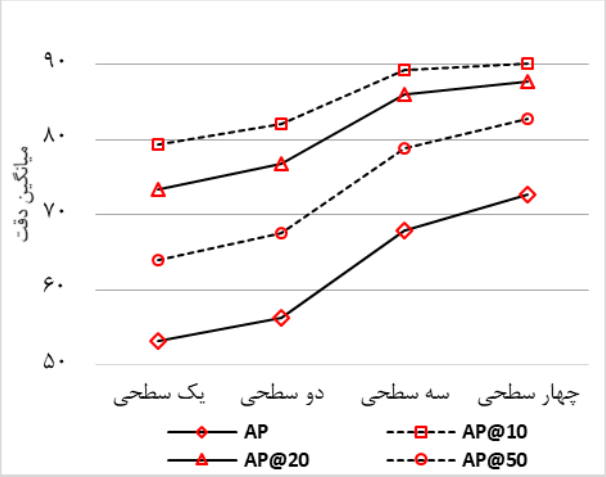
شکل ۷: مقایسه میانگین دقت بازیابی برای مجموعه تصاویر Wang در روش‌های پیشنهادی و کارهای پیشین

برای بررسی دقیق‌تر عملکرد سیستم بازیابی تصویر چندسطحی، بازیابی یک‌سطحی بازیابی در سطح پیکسل تعریف می‌شود و در بازیابی دوسطحی از همجوشی نتایج سطح پیکسل و ناحیه استفاده می‌شود. به همین ترتیب در بازیابی سه‌سطحی، از همجوشی نتایج سطوح پیکسل، ناحیه و شیء و در بازیابی چهارسطحی، از همجوشی نتایج هر چهار سطح استفاده می‌شود. طبق جدول ۱ و شکل ۷ نتایج مربوط به روش‌های چهارسطحی، سه‌سطحی و مفهوم نسبت به سایر روش‌ها بهتر است هرچند که بقیه نتایج نیز تا حد زیادی با کارهای دیگران قابل‌رقابت هستند.

همان‌طورکه در بخش ۳ بیان شد، در سطوح پیکسل و ناحیه از معیار شباهت کسینوسی و در سطوح شیء و مفهوم از معیار فاصله اقلیدسی استفاده‌شده است. همجوشی در سطح تصمیم به این معنی است که از هر چهار معیار برای محاسبه امتیاز نهایی تصاویر پایگاه داده استفاده شود. برای همجوشی نتایج به‌دست‌آمده ابتدا شباهت‌ها و فاصله‌های به‌دست‌آمده در هر سطح با استفاده از رابطه ۳ بین صفر و یک نرمال‌سازی می‌شوند. سپس با کم کردن مقادیر نرمال شده فاصله از یک، این مقادیر به معیار شباهت تبدیل می‌شوند. درنهایت از میانگین شباهت‌های محاسبه‌شده برای هر یک از سطوح، به‌عنوان امتیازهای هر یک از تصاویر پایگاه داده استفاده می‌شود.

$$X_{[0,1]} = \frac{X - Min(X)}{Max(X) - Min(X)} \qquad (۳)$$

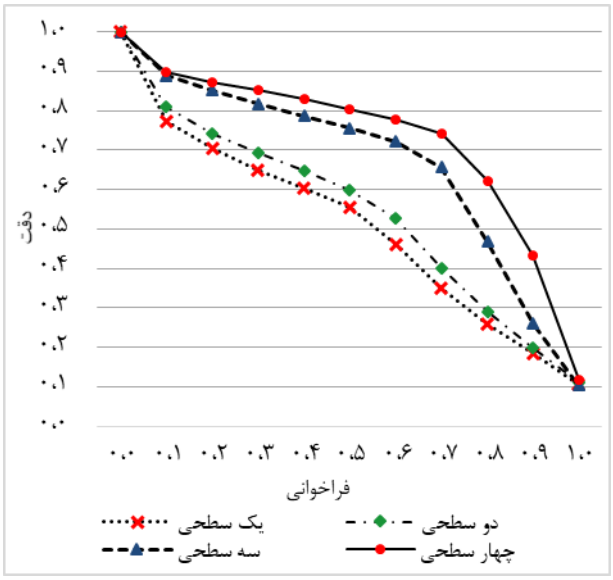
شکل ۸ عملکرد بازیابی تصویر چندسطحی را از دیدگاه دیگری توصیف می‌کند. در این شکل تأثیر همجوشی نتایج هر سطح با سطوح پایین‌تر از خودش به‌خوبی به تصویر کشیده شده است.



شکل ۸: تأثیر بازیابی تصویر چندسطحی در میانگین دقت بازیابی برای مجموعه تصاویر Wang

در شکل ۸ میانگین دقت برای ۱۰، ۲۰ و ۵۰ تصویر بازیابی‌شده با رتبه بالاتر و همچنین میانگین دقت برحسب بازیابی کل تصاویر مرتبط در هر یک از روش‌ها نشان داده شده است. میانگین دقت بازیابی تصویر یک‌سطحی برابر با ۵۳/۲۵ درصد است که در بازیابی دوسطحی به ۵۸/۱۵ درصد افزایش پیدا می‌کند. این معیار در بازیابی سه‌سطحی و چهارسطحی به ۶۷/۹۲ و ۷۲/۹ درصد می‌رسد.

در شکل ۹ منحنی P-R برای مجموعه تصاویر Wang ترسیم شده است. همان‌طورکه در این شکل نشان داده شده است با اضافه کردن سطوح مختلف بازیابی مساحت زیر نمودار P-R در هر دو بعد نمودار به‌خوبی افزایش پیدا می‌کند. درنتیجه در روش پیشنهادی در این پژوهش توازن خوبی بین دو معیار دقت و فراخوانی برقرار است. علاوه بر آن در روش‌های سه‌سطحی و چهارسطحی این دو معیار به میزان قابل‌توجهی افزایش یافته‌اند.



شکل ۹: منحنی P-R برای مجموعه تصاویر Wang

در شکل ۱۰ نمونه‌ای از تصاویر بازیابی‌شده برای شش تصویر پرس‌وجو از شش گروه معنایی مختلف از مجموعه تصاویر Wang به روش چهارسطحی نمایش داده شده است که در آن تصاویر ستون سمت چپ، تصاویر پرس‌وجو هستند.

نکته‌ای که در این نتایج به چشم می‌خورد این است که میانگین دقت متوسط برای ۱۰ و ۲۰ تصویر بازیابی با رتبه بالاتر بسیار بالاست. اهمیت این موضوع در این است که کاربران به نتایجی که در صدر لیست رتبه‌بندی قرار گرفته‌اند اهمیت بیشتری می‌دهند و اگر نیازهای آن‌ها در نتایج بازیابی‌شده در رتبه‌های بالا برآورده شود معمولاً سراغ سایر نتایج نمی‌روند.

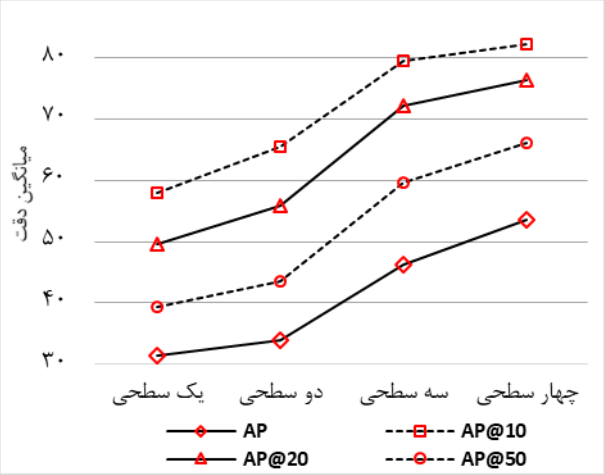
به‌منظور اعتبار بخشیدن بیشتر به نتایج این پژوهش، در ادامه همه معیارهایی که برای پایگاه داده Wang بیان شد، برای مجموعه تصاویر GHIM نیز محاسبه‌شده است. در جدول ۲ میانگین دقت برای ۱۰، ۲۰، ۵۰ تصویر بازیابی‌شده با رتبه بالاتر و همچنین میانگین دقت برای کل تصاویر مرتبط در هر گروه معنایی، در مجموعه تصاویر GHIM نمایش داده شده است. در این جدول معیارهای ارزیابی برای هر یک از سطوح پیکسل، ناحیه، شیء و مفهوم و همچنین روش‌های دوسطحی، سه‌سطحی و چهارسطحی آمده است. در این پایگاه داده، میانگین دقت در بازیابی یک‌سطحی برابر است با ۳۱/۳۳ درصد که در بازیابی دوسطحی، سه‌سطحی و چهارسطحی به ترتیب به ۳۳/۹۷، ۴۶/۳ و ۵۳/۴۷ درصد افزایش یافته است. تأثیر همجوشی نتایج بازیابی در هر سطح با سطوح پایین‌تر در شکل ۱۱ قابل‌مشاهده است.

شکل ۱۰: چند نمونه از تصاویر پرس‌وجو و نتایج به‌دست‌آمده به روش چهارسطحی در پایگاه داده Wang که در آن اولین تصویر سمت چپ پرس‌وجو و چهار تصویر بعدی تصاویر بازیابی‌شده هستند.

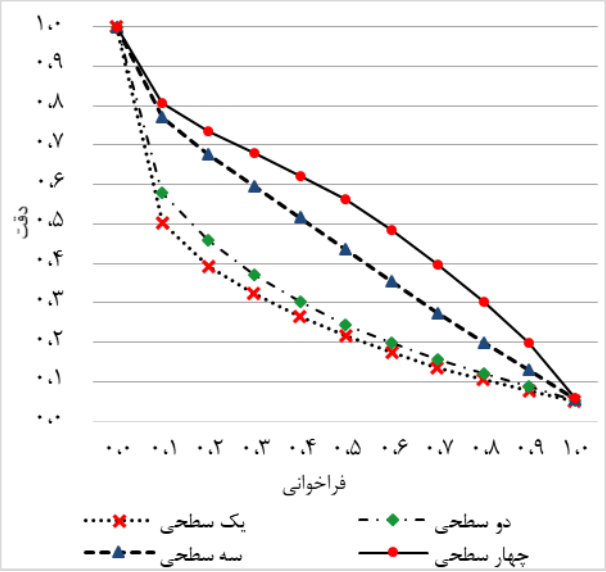
جدول ۲: نتایج به‌دست‌آمده در هر یک از سطوح بازیابی برای مجموعه تصاویر GHIM

روش	معیار ارزیابی			
	AP	AP@50	AP@20	AP@10
پیکسل	۳۱/۳۳	۳۹/۲۸	۴۹/۶۸	۵۸/۰۳
ناحیه	۲۱/۷۱	۲۷/۴۲	۳۶/۷۸	۴۵/۴۵
شیء	۴۰/۸۶	۵۳/۴۱	۶۴	۶۹/۵۲
مفهوم	۴۲/۴۵	۵۲/۷۲	۶۲/۷۳	۶۹/۱۴
دوسطحی	۳۳/۹۷	۴۳/۵۷	۵۵/۹	۶۵/۴۸
سه‌سطحی	۴۶/۳	۵۹/۶	۷۲/۲۵	۷۹/۴۶
چهارسطحی	۵۳/۴۷	۶۶/۱۴	۷۶/۴۱	۸۲/۲۳

در شکل ۱۲ منحنی P-R برای مجموعه تصاویر GHIM ترسیم شده است که نشان‌دهنده عملکرد قابل‌قبول سیستم در دقت و فراخوانی است. همان‌طورکه در این منحنی مشخص است در بازیابی دوسطحی مساحت زیر منحنی نسبت به سطح پیکسل دارای یک افزایش نسبی است. درحالی‌که این مساحت در روش‌های سه‌سطحی و چهارسطحی افزایش زیادی داشته است و این به معنی بهبود قابل قبولی در دقت و فراخوانی و همچنین توازن بین این دو معیار است. در شکل ۱۳ نیز نمونه‌ای از تصاویر بازیابی‌شده برای شش تصویر پرس و جوی مختلف از مجموعه تصاویر GHIM به روش چهارسطحی نمایش داده شده است. در این شکل اولین تصویر هر سطر از سمت چپ تصویر پرس‌وجو و چهار تصویر بعدی این سطر پاسخ به‌دست‌آمده با بازیابی چندسطحی هستند. آنچه به‌وضوح می‌توان در این شکل دید؛ بازیابی تصاویری است که شباهت مفهومی و ساختاری زیادی به تصویر پرس‌وجو دارند اما هیستوگرام رنگی آن‌ها به مقدار قابل‌توجهی با تصویر پرس‌وجو تفاوت دارد.



شکل ۱۱: تأثیر بازیابی تصویر چندسطحی در میانگین دقت بازیابی برای مجموعه تصاویر GHIM



شکل ۱۲: منحنی P-R برای مجموعه تصاویرGHIM

۵ -نتیجه‌گیری

در این پژوهش چهار روش استخراج ویژگی از تصاویر در چهار سطح پیکسل، ناحیه، شیء و مفهوم ارائه شد. سطح پیکسل از ویژگی‌های بصری خام تصویر جهت بازیابی استفاده می‌کند و دارای بار مفهومی کمتری است. سطح ناحیه تلاش می‌کند با استخراج ویژگی از نواحی همگن در رنگ و بافت، به سیستم بینایی انسان نزدیک‌تر شود. در سطح شیء از ویژگی‌های سطح‌بالا استفاده می‌شود که باعث کاهش شکاف معنایی می‌شود. سطح مفهوم نیز با استفاده از بازنمایی برداری کلمات متناظر با اشیاء درون تصویر که با یک شبکه عصبی که به‌وسیله مجموعه بزرگی از اسناد متنی تولیدشده توسط انسان آموزش داده شده است، به درک انسان از تصویر نزدیک‌تر می‌شود.

شکل ۱۳: چند نمونه از تصاویر پرس‌وجو و نتایج به‌دست‌آمده به روش چهارسطحی در پایگاه داده GHIM که در آن اولین تصویر سمت چپ پرس‌وجو و چهار تصویر بعدی تصاویر بازیابی‌شده هستند.

در سیستم بازیابی تصویر چندسطحی پیشنهادشده در این پژوهش با استفاده از میانگین امتیازهای به‌دست‌آمده از معیارهای مختلف در هر سطح، یک همجوشی مناسب در سطح تصمیم بین نتایج ویژگی‌های هر یک از چهار سطح به وجود می‌آید. استفاده از ویژگی‌های سطح‌بالا باعث کاهش شکاف معنایی شده و به تصاویری که در سطوح پایین‌تر امتیاز کمی به دست آورده و بازیابی نمی‌شوند، امتیاز بیشتری اختصاص می‌دهد. برخی از تصاویر (مانند دو تصویر از گل مریم و گل رز) ممکن است با استفاده از ویژگی‌های سطح‌بالا قابل‌تفکیک نباشند ولی در پایگاه داده در دو گروه معنایی مختلف قرار گرفته باشند. در این موارد استفاده از ویژگی‌های سطح‌پایین که جزئیات بیشتری را در نظر می‌گیرند می‌تواند به بهبود معیارهای بازیابی کمک کند.

نتایج به‌دست‌آمده در بخش ۴ نشان‌دهنده عملکرد مناسب روش پیشنهادی بر روی دو پایگاه داده شامل ۱۰۰۰ و ۲۰۰۰ تصویر است. همچنین با توجه به منحنی‌های P-R می‌توان گفت در بازیابی تصویر با تعداد سطوح بیشتر، میانگین دقت و فراخوانی برای هر دو پایگاه داده به شکل معنی‌داری بهبود می‌یابد. نتایج به‌دست‌آمده از روش پیشنهادشده در سطح مفهوم نشان می‌دهد که می‌توان از همبستگی مفهومی کلمات متناظر با اشیاء و صحنه‌های درون تصاویر، به‌عنوان معیاری برای سنجش شباهت آن‌ها استفاده کرد. در سیستم‌های جستجوی تصویر مبتنی بر محتوا نتایج بر اساس معیارهای شباهت یا فاصله امتیازدهی و رتبه‌بندی می‌شوند. در این سیستم‌ها همواره دقت نتایج با رتبه بالاتر از اهمیت بیش‌تری برخوردار است. روش پیشنهادشده در این پژوهش میانگین دقتی در حدود ۹۰ و ۸۲ درصد برای دو مجموعه تصویر مختلف که دارای گروه‌های معنایی متفاوتی از تصاویر هستند، به دست آورده است که نشان‌دهنده عملکرد قابل آن در این زمینه است.

مراجع

[1] A. Cawkell, "Imaging systems and picture collection management: a review," *Information Services & Use*, vol. 12, no. 4, pp. 301-325, 1992.

[2] P. G. Enser and C. G. McGregor, Analysis of visual information retrieval queries, British Library Board London, 1993.

[3] V. N. Gudivada and V. V. Raghavan, "Content based image retrieval systems," *Computer*, vol. 28, no. 9, pp. 18-22, 1995.

[4] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu and R. Zabih, "Image indexing using color correlograms," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 762-768, 1997.

[5] A. W. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 12, pp. 1349-1380, 2000.

[6] R. Datta, D. Joshi, J. Li and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys (Csur)*, vol. 40, no. 2, p. 5, 2008.

[7] J. Z. Wang, J. Li and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 9, pp. 947-963, 2001.

[8] G. Carneiro, A. B. Chan, P. J. Moreno and N. Vasconcelos, "Supervised learning of semantic classes for image annotation and retrieval," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 3, pp. 394-410, 2007.

[9] J.-H. Su, W.-J. Huang, S. Y. Philip and V. S. Tseng, "Efficient relevance feedback for content-based image retrieval by mining user navigation patterns," *IEEE transactions on knowledge and data engineering*, vol. 23, no. 3, pp. 360-372, 2011.

[10] J. Yu, Z. Qin, T. Wan and X. Zhang, "Feature integration analysis of bag-of-features model for image retrieval," *Neurocomputing*, vol. 120, pp. 355-364, 2013.

[11] F. Jing, M. Li, H.-J. Zhang and B. Zhang, "An efficient and effective region-based image retrieval framework," *IEEE Transactions on Image Processing*, vol. 13, no. 5, pp. 699-709, 2004.

[12] L.-J. Li, H. Su, L. Fei-Fei and E. P. Xing, "Object bank: A high-level image representation for scene classification & semantic feature sparsification," in *Advances in neural information processing systems*, pp. 1378-1386, 2010.

[13] G. Pass, R. Zabih and J. Miller, "Comparing images using color coherence vectors," in *Proceedings of the fourth ACM international conference on Multimedia*, pp. 65-73, 1997.

[14] Z.-C. Huang, P. P. Chan, W. W. Ng and D. S. Yeung, "Content-based image retrieval using color moment and Gabor texture feature," in *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, vol. 2, pp. 719-724, 2010.

[15] C. Dagli and T. S. Huang, "A framework for grid-based image retrieval," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2, pp. 1021-1024, 2004.

[16] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," IEEE, p. 1470, 2003.

[۱۷] ساناز کشوری و عبداله چاله‌چاله، «طبقه‌بندی سبک نقاشی هنرمندان با استفاده از هیستوگرام گرادیان جهت‌دار و الگوی باینری محلی»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۷، شماره ۳، ۱۳۹۶.

[18] S. Murala, R. Maheshwari and R. Balasubramanian, "Directional local extrema patterns: a new descriptor for content based image retrieval," *International journal of multimedia information retrieval*, vol. 1, no. 3, pp. 191-203, 2012.

[19] A. Bala and T. Kaur, "Local texton XOR patterns: A new feature descriptor for content-based image retrieval," *Engineering Science and Technology, an International Journal*, vol. 19, no. 1, pp. 101-112, 2016.

[20] C. Carson, S. Belongie, H. Greenspan and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026-1038, 2002.

[21] Y. Deng and B. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE*

transactions on pattern analysis and machine intelligence, vol. 23, no. 8, pp. 800-810, 2001.

[22] D. Hoiem, R. Sukthankar, H. Schneiderman and L. Huston, "Object-based image retrieval using the statistical structure of images," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, pp. II, 2004.

[23] Y. Li, Object and concept recognition for content-based image retrieval, Citeseer, 2005.

[۲۴] اسما شمسی گوشکی، سعید سریزدی، حسین نظام‌آبادی‌پور و محمد شهرام معین، «روشی جدید در بازخورد ربط برای بازیابی تصویر بر اساس محتوا به شیوه چند پرسشی»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۰، شماره ۲، ۱۳۸۹.

[25] Y. Jin, L. Khan, L. Wang and M. Awad, "Image annotations by combining multiple evidence & wordnet," in *Proceedings of the 13th annual ACM international conference on Multimedia*, pp. 706-715, 2005.

[۲۶] هنگامه دلجویی و امیرمسعود افتخاری‌مقدم، «حاشیه‌نویسی خودکار تصویر با استفاده از ارتباط معنایی بین نواحی مبتنی بر تئوری تصمیم چندشرطی»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۲، شماره ۲، ۱۳۹۱.

[27] J. Van De Weijer and C. Schmid, "Coloring local feature extraction," in *European conference on computer vision*, Springer, pp. 334-348, 2006.

[28] A. a. S. I. a. H. G. E. Krizhevsky, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097-1105, 2012.

[29] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," in *International conference on machine learning*, pp. 647-655, 2014.

[30] K. a. Z. A. Simonyan, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[31] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, pp. 3111-3119, 2013.

[32] Tool for computing continuous distributed representations of words: https://code.google.com/archive/p/word2vec/

[33] Wang Image Database: http://wang.ist.psu.edu/docs/related/

[34] GHIM Image Database: http://www.ci.gxnu.edu.cn/cbir/Dataset.aspx

[35] G.-H. Liu, J.-Y. Yang and Z. Li, "Content-based image retrieval using computational visual attention model," *pattern recognition*, vol. 48, no. 5, pp. 2554-2566, 2015.

زیرنویس‌ها

[1] Indexing
[2] Semantic Gap
[3] Image Classification
[4] Relevance Feedback
[5] Bag of Words
[6] Object Detectors
[7] Support Vector Machines
[8] Image Annotation
[9] SIFT: Scale Invariant Feature Transform
[10] Histogram of Oriented Gradients
[11] Local Binary Patterns
[12] Codebook
[13] Integrated Region Matching
[14] Expectation Maximization
[15] Earth Mover's Distance
[16] Object Recognition
[17] Classifier
[18] Multilayer Perceptron
[19] Dempster Shafer
[20] Local Features
[21] Global Features
[22] Key Points
[23] Difference of Gaussian
[24] Scale Space
[25] Saturation
[26] Pooling
[27] Skip-gram
[28] Precision
[29] Average Precision (AP)
[30] Recall
[31] Average Recall
[32] Thresholds