



طبقه‌بندی تصاویر بازیابی شده با استفاده از الگوریتم SIFT و کدگذاری خطی با قید محلی

محسن جابری^۱ و فرزین یغمایی^۲

<http://jmvip.sinaweb.net>

www.ismvip.ir

چکیده

با رشد روز افزون اینترنت و ابزارهای تصویربرداری دیجیتال، اندازه پایگاه داده تصاویر به سرعت در حال بزرگتر شدن است. در چنین شرایطی، نیاز شدیدی به ابزارها و روش‌های کارا برای جستجوی تصاویر دلخواه در پایگاه داده‌های بزرگ به وجود آمده است. استخراج ویژگی اساسی‌ترین قدم در ایجاد یک سامانه بازیابی تصاویر براساس محتواست و نقش بسیار تعیین کننده‌ای در دقت سامانه بازیابی دارد. در این مقاله روشی جدید جهت طبقه‌بندی تصاویر بازیابی شده براساس محتوا ارائه شد. پس از استخراج ویژگی و محاسبه توصیفگرهای مربوط به هر دسته توسط الگوریتم SIFT، الگوریتم TF-IDF توصیفگرهای مناسب را مشخص کرده و از خوشه‌بندی جهت یافتن توصیفگرهای کاندیدای هر دسته استفاده می‌کند. در مرحله بعد از ضرایب بازنمایی توصیفگرهای هر دسته با توجه به نماینده‌های تولید شده از مرحله قبل توسط الگوریتم کدگذاری خطی با قید محلی به عنوان ویژگی استفاده شده است. در نهایت از این ویژگی‌های تولید شده برای طبقه‌بندی تصاویر بازیابی شده استفاده می‌شود. دسته‌بندی که برای ارزیابی سیستم پیشنهادی مورد استفاده قرار گرفته، ماشین یادگیر پیشینه می‌باشد. دقت به دست آمده در این دسته‌بند بر روی پایگاه داده Caltech-101، ۹۸/۵ درصد و بر روی پایگاه داده Flowers-17، ۹۷/۹۰ درصد می‌باشد.

کلیدواژه‌ها

بازیابی تصاویر، الگوریتم TF-IDF، الگوریتم SIFT، کدگذاری خطی با قید محلی، ماشین یادگیر پیشینه

مدیریت و سازماندهی و بازیابی دقیق هستند. در ابتدای دهه‌ی نود با افزایش حجم تصاویر موجود در پایگاه‌های نظیر شبکه جهانی اینترنت و نیز برای غلبه بر مشکلات سیستم‌های مبتنی بر متن، توسعه سیستم‌های بازیابی تصویر مبتنی بر محتوا پیشنهاد شد. که از محتوای معنایی تصویر به جای کلمات کلیدی و توصیفات متنی مربوط به تصویر استفاده می‌شود، در بازیابی مبتنی بر محتوا از مفاهیم دیداری و سطح پایین تصویر نظیر رنگ، بافت و شکل تصویر استفاده می‌شود [۱].

سیستم‌های اولیه این رده برای بازیابی، تصویری را از کاربر درخواست می‌کند تا یک یا چند ویژگی سراسری را انتخاب و محدوده‌ای برای مقدار آنها تعیین کند، سپس عمل بازیابی براساس این اطلاعات انجام می‌شود [۲].

۱ مقدمه

امروزه تصویر یکی از مهم‌ترین ابزارهای برقراری ارتباط محسوب می‌شود. که با رشد و گسترش تکنولوژی‌های عکسبرداری و دوربین‌های دیجیتال داشتن آرشیوهای وسیعی از تصاویر حتی برای افراد معمولی نیز امری طبیعی است. این مجموعه تصاویر که حاوی اطلاعات بسیار متنوع و گوناگون می‌باشند، نیازمند یک

این مقاله در تیرماه ۹۸ دریافت، سومین بازنگری در مردادماه ۹۹ انجام و متعاقباً در همان ماه پذیرفته شد.

^۱ کارشناس ارشد مهندسی برق و کامپیوتر از دانشگاه سمنان

رایانامه: mohsenjaberi@semnan.ac.ir

^۲ دانشکده برق و کامپیوتر دانشگاه سمنان

رایانامه: f_yaghmaee@semnan.ac.ir

در پژوهش مطرح شده از روش فازی جهت شناسایی شیء مورد نظر استفاده شده است. اطلاعات مورد نظر هر شیء از یک تصویر استخراج می‌شود. این اطلاعات اندازه‌گیری نزدیکی از شیء مورد علاقه با اشکال شناخته شده را فراهم می‌کند. برای هر شیء مقادیر عضویت فازی محاسبه شده و به عنوان بردار ویژگی در نظر گرفته می‌شود [۶].

ناحیه برجسته یکی از مهمترین بخش‌های یک تصویر می‌باشد. در هنگام جستجوی تصاویر در مجموعه داده‌های بزرگ، ناحیه برجسته تصویر بیشترین توجه را به خود جلب می‌کند. با این حال برای بهبود دقت بازیابی تصویر، فقط توجه به شیء برجسته موجود در تصویر کافی نیست زیرا پس‌زمینه نیز بر دقت بازیابی تصویر تأثیرگذار است. به همین دلیل جینگ ژنگ و همکارانش، مفهوم جدیدی به نام ناحیه برجسته توسعه یافته (ESR) ارائه داده‌اند. در ابتدا ناحیه برجسته تصویر ورودی با استفاده از الگوریتم کنتراست ناحیه (CR³) شناسایی می‌شود. سپس یک سیستم مرکزیت قطبی ساخته و مرکزیت ناحیه برجسته به عنوان قطب تعیین می‌گردد. مناطق اطراف ناحیه برجسته توسط مناطق همسایه تعیین می‌شود. ترکیب حاصل از ناحیه برجسته و مناطق اطراف آن به عنوان ESR تعریف شده است. محتوای بصری استخراج شده از ESR با استفاده از مدل⁴ Bow براساس ویژگی‌های SIFT، Gabor و HSVH می‌باشد و یک مدل گراف از گره‌های محتوای بصری برای نشان دادن تصویر ورودی ارائه شده است. همچنین یک معیار تشابه جدید با ترکیب شباهت ناحیه برجسته و مناطق اطراف آن با استفاده از وزن تعریف شده است [۷].

ژیچینگ لی و همکاران، جهت بازیابی تصاویر براساس محتوا ابتدا از رویکرد آشکارسازی اشیاء برجسته در تصویر استفاده می‌کنند. سپس از تصویر بخش‌بندی شده ویژگی‌ها استخراج شده و این ویژگی‌ها توسط الگوریتم SVM دسته‌بندی می‌شوند [۸].

با توجه به این که پس‌زمینه موجود در تصاویر یکی از مشکلاتی است که فرآیند استخراج ویژگی و بازیابی را با مشکل مواجه می‌کند رهن اشرف و همکارانش، با استفاده از رویکرد بخش‌بندی تصاویر ابتدا ناحیه شیء مورد نظر جداسازی شده و پس از استخراج ویژگی دسته‌بندی با استفاده از شبکه عصبی انجام می‌شود [۹].

سومیا پرکش رعنا و همکارانش، روشی جدید برای بازیابی تصاویر براساس محتوا بر اساس ترکیب ویژگی‌های رنگی و شکلی ارائه داده‌اند. در این روش از میانگین، واریانس و چولگی به عنوان ممان‌های رنگی استفاده می‌شود و برای استخراج ویژگی‌های شکلی از کانولوشن فیلترهای افقی، عمودی و قطری بر روی تصویر استفاده می‌شود. در نهایت از فاصله اقلیدسی برای ارزیابی و بازیابی تصاویر استفاده می‌شود [۱۰].

در سیستم‌های پیشرفته‌تر، کاربر قادر بود تا درجه اهمیت ویژگی‌های انتخاب شده را تعیین کند. به دلیل ناتوانی در جلب رضایت کاربر، در سیستم‌های نسل بعد قابلیت دریافت تصویر به عنوان الگوی بازیابی فراهم شد. با این امکان، کاربر در کنار قابلیت‌های ذکر شده می‌توانست تصویر یا شمای تصویر مورد نظر خود را به سیستم وارد کرده و تصویر مشابه به آن را در پایگاه داده استخراج نماید.

امروزه با افزایش تصاویر و پیچیدگی ساختار آنها و شباهت تصاویر از کلاس‌های مختلف به همدیگر استفاده‌ی تنها از ویژگی‌های سراسری تصویر در بازیابی نتایج جستجوی خوبی به دست نمی‌آید [۳]. بنابراین استفاده از ویژگی‌های محلی تصویر ضروری شد و استخراج نقاط کلیدی تصاویر کمک زیادی به افزایش دقت تصاویر بازیابی شده می‌کند.

شناسایی نقاط کلیدی در یک تصویر و استخراج ویژگی‌های مناسب از جهت تطبیق نقطه کلیدی، با نقاط کلیدی سایر تصاویر یکی از چالش‌های اساسی در پردازش تصویر و بینایی ماشین است. نقاط کلیدی باید نقاطی تکرار پذیر باشند که صرف نظر از زاویه دید، تغییر مقیاس تصویر، چرخش و یا تغییر شدت نور در تصویر قابل شناسایی باشند. همچنین برای تطبیق دو نقطه کلیدی در دو تصویر، علاوه بر این که نقاط مرتبط باید به صورت همزمان بین نقاط کلیدی انتخاب شده در تصویر باشند، باید به گونه‌ای از این نقاط ویژگی استخراج شود تا دو نقطه کلیدی مرتبط با توجه به ویژگی‌هایشان به هم انتساب داده شوند [۴].

در روش‌های سنتی پس از استخراج ویژگی و محاسبه توصیفگرهای مربوط به هر دسته توسط الگوریتم SIFT⁵، از خوشه‌بندی جهت یافتن توصیفگرهای کاندیدای هر دسته استفاده می‌شود. سپس از هیستوگرام تکرار توصیفگرها در هر دسته به عنوان ویژگی برای داده‌های آموزشی و آزمون استفاده شده و عملیات بازیابی و دسته‌بندی انجام می‌شود.

در ادامه بخش‌های مختلف الگوریتم پیشنهادی مورد بررسی قرار می‌گیرد. در بخش دوم مروری بر تحقیقات گذشته انجام می‌شود و در بخش سوم روش پیشنهادی و در بخش چهارم نیز آزمایشات و نتیجه‌گیری مورد بحث قرار می‌گیرد.

۲ مروری بر تحقیقات گذشته

هیستوگرام رنگ و مشتقات آن یکی از مهمترین بردارهای ویژگی برای رنگ می‌باشد. محاسبه هیستوگرام رنگ ساده است و برای به‌دست آوردن ویژگی پراکندگی سراسری و محلی رنگ‌ها در یک تصویر بسیار مؤثر است. به علاوه نسبت به انتقال و چرخش ثابت است و با تغییر اندازه و زاویه دید به کندی تغییر می‌کند [۵].

شانموگودیو و همکارانش روشی جهت بازیابی تصویر مبتنی بر مفهوم با استفاده از فازی‌سازی ارائه کرده است. شناسایی شکل جسم به دلیل عدم اطمینان در مرز جسم مورد نظر پیچیده می‌باشد.

²extended salient region

³Region Contrast

⁴Bag of Words

⁵Scale Invariant Feature Transform (SIFT)

نهایت با ترکیب این روش‌ها عمل بازیابی تصویر را برای بازیابی مبتنی بر محتوا افزایش می‌دهد [۱۶].
چابرا و همکارانش، یک سیستم بازیابی تصویر مبتنی بر محتوا برای استخراج یک بردار ویژگی از یک تصویر و بازیابی مؤثر تصاویر مبتنی بر محتوا پیشنهاد شده است. سیستم پیشنهادی دو تکنیک استخراج ویژگی به نام‌های SIFT^۴ و ORB^۵ را برای استخراج توصیفگرهای یک تصویر در نظر گرفته‌اند. سپس با استفاده از الگوریتم خوشه‌بندی K میانگین، ۳۲ خوشه برای هر آرایه توصیفی تولید می‌کند و در نهایت یک بردار ویژگی ۳۲ بعدی برای هر توصیفگر به دست می‌آورد. با استفاده از الگوریتم کاهش بعد^۶ LPP بردارهای ویژگی را به ۴ و ۸ مؤلفه کاهش می‌دهد. در نهایت با استفاده از سه تکنیک طبقه‌بندی؛ درخت تصمیم‌گیری^۱، MLP^۷ و جنگل تصادفی^۸ کار طبقه‌بندی تصاویر بازیابی شده را انجام می‌دهد [۱۷].

۳ روش پیشنهادی

در این بخش به معرفی روش پیشنهادی می‌پردازیم. در ابتدا ویژگی‌های SIFT و توصیفگرهای مربوط به آنها از تصاویر استخراج می‌شود. سپس توسط الگوریتم TF-IDF^۹ توصیفگرهای مناسب مشخص شده و عملیات خوشه‌بندی جهت یافتن توصیفگرهای کاندیدای هر دسته انجام می‌شود. پس از یافتن توصیفگرهای کاندیدای هر دسته به جای استفاده از هیستوگرام تکرار توصیفگرها در هر دسته از ضرایب بازنمایی توصیفگرهای هر دسته با توجه به نماینده‌های تولید شده از مرحله قبل توسط الگوریتم کدگذاری خطی با قید محلی^{۱۰} به عنوان ویژگی استفاده می‌شود. در نهایت از این ویژگی‌های تولید شده برای بازیابی تصاویر و طبقه‌بندی استفاده خواهد شد. پایگاه داده مورد استفاده در این مقاله Caltech-101^{۱۱} و Flowers-17^{۱۲} می‌باشد. در این مقاله در مرحله تولید ویژگی، دو رویکرد مورد استفاده قرار می‌گیرد.

رویکرد اول که در روش‌های سنتی برای آموزش سیستم به کار می‌برند. در ابتدا برای هر تصویر آموزشی ویژگی‌های SIFT استخراج می‌شود و هیستوگرام تکرار توصیفگرهای ویژگی‌های تصویر در مراکز خوشه‌ها به عنوان ویژگی در نظر گرفته می‌شود. رویکرد دیگری که برای استخراج ویژگی در این مقاله در نظر گرفته شده است استفاده از ضرایب بدست آمده از رویکرد

در روش سانا و همکارانش، از روش فاصله همینگ برای بازیابی تصویر استفاده شده است. ابتدا یک واژگان بصری از توصیفگرها مانند رنگ، بافت، نقاط موردعلاقه ایجاد می‌گردد. سپس یک امضاء برای هر تصویر اختصاص داده می‌شود. استخراج امضاء براساس موجک هار، نقاط موردعلاقه هریس و هیستوگرام رنگی صورت می‌گیرد. سپس با استفاده از الگوریتم PCA عملیات کاهش بعد انجام شده و برای تطبیق تشابه کد باینری استخراجی از فاصله همینگ استفاده می‌شود [۱۱].

یوژنگ کوی و بیوا خیو، از ویژگی‌های رنگی در فضای RGB برای بازیابی تصاویر براساس محتوا استفاده می‌کنند. در این روش برای افزایش دقت بازیابی تصاویر، در ابتدا ناحیه مربوط به اشیاء برجسته در تصویر با استفاده از رویکرد خوشه‌بندی جابه‌جایی میانگین^۱ مشخص می‌شود. پس از مشخص شدن ناحیه پس‌زمینه و ناحیه مربوط به شیء برجسته در تصویر، عملیات استخراج ویژگی انجام خواهد شد. از معایب این روش می‌توان به زمان‌بر بودن فرآیند بخش‌بندی تصویر اشاره کرد [۱۲].

در کار آینده و همکارانش، از ویژگی هیستوگرام رنگ و ضرایب گابور به منظور بازیابی تصاویر استفاده می‌شود. پس از استخراج و ترکیب ویژگی‌ها با هم، از فاصله منهن برای بازیابی تصاویر استفاده می‌شود [۱۳].

فیلیپ مرکل و همکارانش، روشی برای بازیابی تصویر تعاملی ارائه کرده‌اند. رویکرد پیشنهادی با معرفی یک بازیابی تصویر تعاملی هدایت‌گرانه معنایی، شکاف معنایی بین ویژگی‌های سطح پایین تصاویر و مفاهیم معنایی سطح بالا را کاهش می‌دهد. ایده اصلی شامل بازیابی تصاویر بر اساس ویژگی‌های بصری آنها است، تفسیر این تصاویر با مفاهیم آنها، توسط ساختار هستان شناسی که به وسیله کاربر تعریف می‌شود، توسعه می‌یابد [۱۴].

بازیابی تصاویر براساس ویژگی‌های ابتدایی آنها شامل، شکل، بافت، رنگ و اطلاعات مکانی بسیار مهم می‌باشد از همین رو احمد خوجا و همکارانش، رویکردی برای ترکیب اطلاعات رنگی با ویژگی‌های استخراج شده از شکل تصاویر ارائه داده‌اند. برای استخراج ویژگی‌های شکلی از لبه‌های تصویر و نقاط کلیدی شامل گوشه‌ها استفاده می‌شود. نتایج آزمایشات بر روی پایگاه داده‌های مختلف جامعیت این روش را نشان می‌دهد [۱۵].

معتمص الصمادی، ویژگی‌های مهم و قوی را از پایگاه داده تصاویر استخراج نموده و آنها را در مخزن ویژگی به شکل بردارهای ویژگی ذخیره می‌کند. این مخزن ویژگی شامل: امضای رنگ، ویژگی‌های شکل و ویژگی‌های بافت است. بر این اساس، یک ارزیابی تشابه ابتکاری با یک الگوریتم فرا ابتکاری (الگوریتم ژنتیک^۲ با تبرید شبیه‌سازی شده^۳) به دست آمده است. و در

⁴ Oriented Fast and Rotated BRIEF

⁵ locality-preserving projection

⁶ decision tree

⁷ multilayer perceptron

⁸ random forest

⁹ Term Frequency Inverse Document Frequency

¹⁰ Locality Constrained Linear Coding

¹¹ www.vision.caltech.edu/Image_Datasets/Caltech101

¹² www.robots.ox.ac.uk/vgg/data/flowers/17/index.html

¹ Mean shift clustering

² genetic algorithm

³ simulating annealing

۳-۱-۱ ساخت کتابچه کد

اولین گام در تکنیک کیسه واژگان، ساخت کتاب کد حاوی کلمات بصری است. به منظور به دست آوردن این کلمات کلیدی (نقاط کلیدی) نیاز به یک آشکارگر ویژگی^۲ است. همچنین نیاز به روشی یکتا برای توصیف هر یک از این نقاط کلیدی است که به عنوان توصیفگر ویژگی^۳ شناخته می شود. در ادبیات موضوع روش های بسیاری را می توان برای یافتن نقاط کلیدی استفاده نمود.

آشکارگر و توصیفگر SIFT یکی از روش های شناخته شده در استخراج و توصیف نقاط کلیدی است. که به دلیل مقاومت آن، در بسیاری از پروژه های بینایی ماشین در سال های اخیر استفاده شده است. الگوریتم SIFT از تابع تفاضل گوسی^۴ استفاده می کند و به همین دلیل ویژگی ها را در فضای مکان، مقیاس جستجو می کند و لذا مقاومت بالاتری را نسبت به سایر روش ها ارائه می کند. در این مقاله برای استخراج نقاط کلیدی از این روش استفاده شده است. خروجی الگوریتم SIFT بر روی یک تصویر مجموعه ای موقعیت های مکانی سطر و ستون روی تصویر به عنوان مکان نقاط کلیدی و یک بردار ویژگی^۵ ۱۲۸ تایی برای توصیف هر نقطه کلیدی است. بنابراین الگوریتم SIFT بردار ویژگی بزرگی تولید می کند. که اگر از همه ویژگی ها استفاده شود در مراحل بعدی هم محاسبات پیچیده تر و سنگین تر می شود و هم حجم کتابچه کد افزایش پیدا می کند. بنابراین در مرحله اول تعداد تکرارهای مربوط به هر توصیفگر را مشخص می کنیم. و با این کار هم میزان اهمیت و یکتا بودن هر کدام از توصیفگرها را مشخص می کنیم و هم ویژگی هایی که از اهمیت بالاتری برخوردار هستند را شناسایی کرده و در مراحل بعدی از آنها استفاده می کنیم. این بیخش از کار را با استفاده از الگوریتم TF-IDF به انجام می رسانیم.

۳-۲ استفاده از رویکرد TF-IDF جهت انتخاب کلمات کد

بهینه

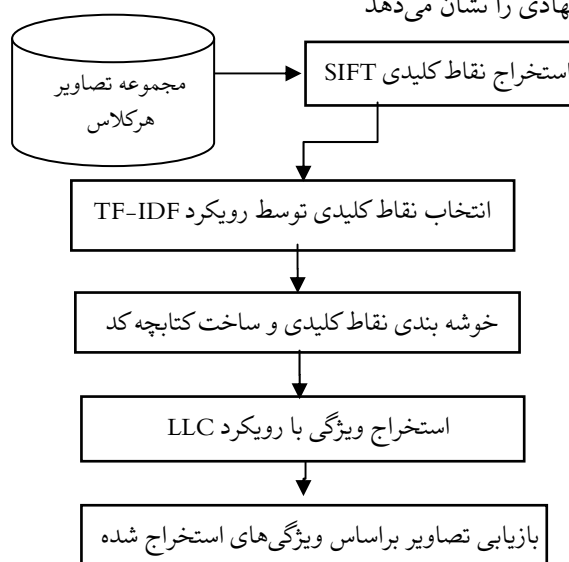
پس از استخراج نقاط کلیدی توسط الگوریتم SIFT و محاسبه توصیفگرهای نقاط کلیدی، برای هر مجموعه از تصاویر، با استفاده از رویکرد TF-IDF کتابچه کد بهینه سازی می شود. در این روش^۶ نشان روش^۶ TF نشان دهنده فرکانس یک توصیفگر در مجموعه ای خاص می باشد و^۱ DF نشان دهنده فرکانس یک توصیفگر در کل مجموعه تصاویر می باشد [۱۸]. برای محاسبه TF می توان از رابطه زیر استفاده کرد.

کدگذاری خطی با قید محلی می باشد. در این روش پس از استخراج ویژگی، توصیفگرهای مربوط به ویژگی ها محاسبه شده و عملیات خوشه بندی روی ویژگی ها انجام می شود تا در صورت هم پوشانی ویژگی ها با یکدیگر، نماینده های مناسب برای هر دسته انتخاب شود. سپس از ضرایب بازنمایی توصیفگرهای هر دسته با توجه به نماینده های تولید شده از مرحله قبل توسط الگوریتم کدگذاری خطی با قید محلی به عنوان ویژگی استفاده می شود. در نهایت از بردارهای ویژگی استخراج شده به عنوان ورودی یک الگوریتم یادگیری ماشین برای طبقه بندی تصاویر ورودی استفاده خواهد شد.

روش دسته بندی مورد استفاده در این مقاله ماشین یادگیر پیشینه می باشد که نتایج حاصل از این روش با محاسبه شباهت بر اساس فاصله اقلیدسی و دسته بندی با استفاده از ماشین بردار پشتیبان مقایسه خواهد شد. در ادامه بخش های مختلف روش پیشنهادی تشریح می شود.

۳-۱ استخراج ویژگی

در بازشناسی الگو و پردازش تصویر، استخراج ویژگی از یک مجموعه از داده های اندازه گیری شده شروع شده و یک مجموعه داده مشتق شده (ویژگی ها) را تولید می کند که این ویژگی ها نمایانگر داده های اولیه اما با ابعاد کمتر و فاقد افزونگی هستند و عملیات بعدی را ساده تر می کنند. در واقع استخراج ویژگی نوعی کاهش ابعاد بشمار می آید. در این مقاله از تکنیک کیسه واژگان بصری^۱ برای استخراج یک بردار ویژگی با طول یکسان از تمامی تصاویر استفاده شده است. شکل ۱ بلوک دیاگرام سیستم پیشنهادی را نشان می دهد



شکل ۱: بلوک دیاگرام سیستم پیشنهادی

^۲Feature Detector

^۳FeatureDescriptor

^۴Difference of Gaussian

^۵term-frequency

^۶document-frequency

^۱Bag of Visual Words

ممکن است سیستم نهایی دقت کافی را نداشته باشد و کتاب‌کد قابلیت ارائه دید کلی از تصاویر مورد نظر نداشته باشد. به منظور حذف تکرارهای احتمالی نقاط کلیدی از خوشه‌بندی استفاده شده است.

فرشاد تاجری‌پور و همکارانش [۱۹]، یک روش جدید برای بازیابی تصاویر دیجیتال از پایگاه‌داده‌های بزرگ ارائه شده‌است. که با استفاده از تکنیک‌های آنالیز بافت، ویژگی‌های متمایز همراه با ویژگی‌های رنگ و شکل را استخراج می‌کنند. الگوهای دودویی محلی رنگی و واریانس محلی به عنوان ویژگی‌های دقیق را از تصویر استخراج می‌کند و با توجه به اندازه پایگاه‌های داده تصویر بزرگ، یک الگوریتم اضافی جدید برای کاهش پیچیدگی محاسباتی پیشنهاد داده است. الگوریتم پیش‌فرایند پیشنهادی میزان عدم شباهت بین تصویر پرس وجو و هر دسته را با توجه به عدم توجه به دسته‌های نزدیک به هم مقایسه می‌کند پس از محاسبه بردارهای ویژگی برای همه تصاویر، آنها را در یک دسته خاص قرار می‌دهند. بردار ویژگی متوسط یک شناسه خوب برای آن دسته خواهد بود. و هر دسته‌ای که از آستانه دورتر باشد، به عنوان یک دسته نزدیک شناخته نمی‌شوند، بنابراین نیازی به محاسبه اندازه‌گیری شباهت بین نمونه‌های آن (تصاویر) و تصویر پرس‌وجو نیست.

اما در اینجا تمامی نقاط کلیدی ابتدا خوشه‌بندی می‌شوند و سپس تنها از مراکز خوشه‌ها به عنوان نقاط کلیدی نهایی استفاده می‌شود. این روند موجب می‌شود نقاط کلیدی تکراری و شبیه به هم در کتاب‌کد قرار نگیرند و لذا میزان افزونگی کتاب‌کد کاهش یابد. مقدار بهینه و از پیش تعریف‌شده‌ای برای تعداد خوشه‌های استفاده شده در الگوریتم خوشه‌بندی وجود ندارد و با سعی و خطا به دست می‌آید. در این مقاله از الگوریتم بسیار ساده و کارآمد و شناخته شده k-میانگین برای خوشه‌بندی نقاط کلیدی استفاده شده است.

۳-۳-۱ الگوریتم خوشه‌بندی K میانگین

این روش علی‌رغم سادگی آن یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر مانند (خوشه‌بندی فازی) و روشی انحصاری و مسطح محسوب می‌شود.

برای این الگوریتم شکل‌های مختلفی بیان شده است، ولی همه آنها دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند:

- به دست آوردن نقاطی به عنوان مراکز خوشه‌ها این نقاط در واقع همان میانگین نقاط متعلق به هر خوشه هستند.
- نسبت دادن هر نمونه داده به یک خوشه که آن داده کمترین فاصله تا مرکز آن خوشه را دارا باشد.
- در نوع ساده‌ای از این روش ابتدا به تعداد خوشه‌های مورد نیاز نقاطی به صورت تصادفی انتخاب می‌شود. سپس داده‌ها با توجه با میزان نزدیکی (شباهت) به یکی از این خوشه‌ها نسبت داده می‌شوند و بدین ترتیب خوشه‌های جدیدی حاصل می‌شود. با

$$t_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,i}} \quad (1)$$

که احتمال رخداد یک توصیفگر i در مجموعه خاص j می‌باشد. و عبارتست از تقسیم تعداد تکرار یک توصیفگر بر تعداد کل تصاویر موجود در آن مجموعه.

IDF به صورت احتمال رخداد کلمه k در بیش از نصف تصاویر تعریف می‌شود. و برابر است با لگاریتم (تعداد کل محتوا/تعداد محتوا i در آن)

$$IDF = \log\left(\frac{N}{n_i}\right) \quad (2)$$

بنابراین مقدار TF-IDF به صورت زیر می‌باشد.

$$TF - IDF = TF \times I \quad (3)$$

یک عدد از حاصلضرب آنها به دست می‌آید که این عدد نشان دهنده وزن و میزان اهمیت این توصیفگر می‌باشد. هر چه میزان TF-IDF بیشتر باشد بهتر و هر چه این عدد به صفر نزدیک‌تر باشد، نشان دهنده این است که توصیفگر مورد نظر از اهمیت کمی در محتوا برخوردار است. در واقع در این مقاله توصیفگرها را به عنوان همان کلمات در نظر می‌گیریم. تنها مشکل در معادل گرفتن توصیفگرها با کلمات در متن این است که؛ در متن دقیقاً به دنبال همان کلمه هستیم. به عنوان مثال: برای پیدا کردن کلمه سیستم در یک متن، دقیقاً باید مشابه کلمه سیستم در متن وجود داشته باشد تا بتوان آن را به عنوان تکرار کلمه در نظر گرفت. ولی در اینجا الگوریتم SIFT توصیفگرهای که استخراج می‌کند دقیقاً تکرار همدیگر نیستند. بنابراین ما از میزان تشابه ۹۵ درصدی استفاده می‌کنیم. به این صورت که اگر به میزان ۹۵ درصد با هر کدام از توصیفگرها مشابهت داشت. آن را به عنوان تعداد تکرار آن توصیفگر در نظر بگیرد. در حالت کلی الگوریتم مورد استفاده باید فرکانس تکرار هر یک از توصیفگرها در مجموعه خاص خود را ماکزیمم و در سایر مجموعه‌ها، منیمم کند. توصیفگری که این خاصیت را داشته باشد به عنوان نماینده برتر مجموعه انتخاب می‌شود و قابلیت تمایز خوبی را ایجاد می‌کند.

۳-۳ خوشه‌بندی نقاط کلیدی

در تکنیک کیسه کلمات؛ هدف، گرفتن مجموعه‌ای از تصاویر آموزشی و استخراج تعداد زیادی نقطه کلیدی از آنهاست به گونه‌ای که این نقاط یک دید کلی از تصاویر را فراهم کنند. بردارهای توصیفگر این نقاط کلیدی، کتاب‌کد را تشکیل می‌دهند. با این حال اگر از تمامی این توصیفگرها در کتاب‌کد استفاده شود، اندازه کتاب‌کد بسیار بزرگ خواهد شد و همچنین تعداد زیادی داده تکراری را در خود خواهد داشت. بزرگ بودن اندازه کتاب‌کد محاسبات بعدی را پیچیده‌تر و زمان‌گیر خواهد کرد. همچنین در صورتی که از تعدادی کمی از توصیفگرها استفاده شود

$$\arg \min_c \sum_{i=1}^N \|x_i - Bc_i\|^2 + \lambda \|c_i\|_1 \quad (5)$$

در فرمول (5) بخش اول ترم بازسازی و بخش دوم پراکندگی را تشکیل می‌دهد. عامل تنظیم پراکندگی چندین نقش مهم را ایفاء می‌کند.

۱- کتابچه کد Over Complete خواهد بود یعنی $M > N$ می‌باشد یعنی نرم برای اطمینان از اینکه سیستم تعریف شده دارای راه حلی واحد باشد، ضروری است.

۲- پراکندگی این امکان را فراهم می‌کند تا بازنمایی آموزش داده شده، الگوهای برجسته‌ای را از توصیفگرهای محلی ذخیره کند.

از آنجایی که قید محلی بودن نسبت به قید پراکندگی از اهمیت بیشتر برخوردار است الگوریتم LLC قید محلی بودن را به صورت زیر فراهم می‌کند.

$$\min_c \sum_{i=1}^N \|x_i - Bc_i\|^2 + \lambda \|d_i \odot c_i\|^2 \quad (6)$$

s.t. $\mathbf{1}^T c_i = 1$

ورودی توصیفگرها و B مراکز خوشه‌ها و مجهول هست، همان ضریبی است که به دنبال آن هستیم و میزان پراکندگی را نشان می‌دهد. علامت \odot به معنای ضرب عنصر در عنصر می‌باشد $d_i \in \mathbb{R}$ تضمینی جهت محلی بودن تطبیق دهنده می‌باشد که فاصله اقلیدسی هر کدام از توصیفگرها تا مراکز خوشه‌ها را محاسبه می‌کند. که به صورت نسبی با درجه آزادی مختلف برای هر بردار پایه، شباهت آنها را با توصیفگر ورودی فراهم می‌کند. و به صورت زیر تعریف می‌شود.

$$d_i = \exp\left(\frac{\text{dist}(x_i, B)}{\sigma}\right) \quad (7)$$

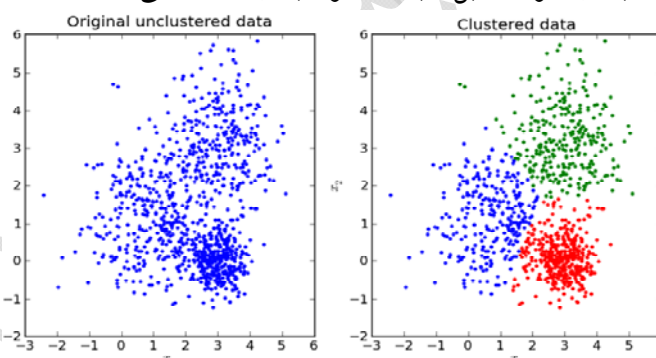
فاصله مورد استفاده برای به دست آوردن شباهت بین بردار پایه و توصیفگر ورودی فاصله اقلیدسی می‌باشد. و برای کاهش بار وزنی می‌باشد، تا میزان محلی بودن را حفظ نماید. برای دستیابی به عملکرد مناسب در دسته‌بندی، فرآیند کدگذاری باید کدهای مشابهی را برای توصیفگرهای مشابه تولید کند. ترم تنظیم محلی بودن که در معادله 6 آورده شده است دارای مزایای زیر می‌باشد.

بازسازی بهتر³: در Vector Quantization هر توصیفگر توسط یک بردار پایه تکین در کتابچه کد نمایش داده می‌شود. همچنین به دلیل خطاهای بالا در کوانتیزه کردن، VQ برای توصیفگرهای مشابه ممکن است بسیار متفاوت باشد. همچنین VQ ارتباط بین پایه‌های مختلف را نادیده می‌گیرد. اما در روش LLC هر توصیفگر به صورت خیلی دقیق توسط چندین بردار پایه

تکرار همین روال می‌توان در هر تکرار با میانگین‌گیری از داده‌ها مراکز جدیدی برای آنها محاسبه کرد. و به طور تکراری داده‌ها را به خوشه‌های جدید نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر تغییری در داده‌ها حاصل نشود. تابع آورده شده در رابطه 4 به عنوان تابع هدف مطرح است.

$$a = \sum_{i=1}^k \sum_{j=1}^N \|x_i^{(j)} - c_j\|^2 \quad (4)$$

معیارفاصله بین نقاط $\|x_i^{(j)} - c_j\|^2$ و ج مرکزخوشه زام است. در این مقاله حداکثر تکرار برای رسیدن به جواب مطلوب 50 تکرار در نظر گرفته شده است. همچنین تعداد خوشه‌های مورد نظر، 500 عدد در نظر گرفته می‌شود. شکل 2 یک مجموعه داده دوبعدی نمونه را قبل و بعد از خوشه‌بندی را نشان می‌دهد.



شکل 2: داده‌های اولیه (تصویر سمت چپ) و داده‌های خوشه‌بندی شده (تصویر سمت راست)

۳-۴ کدگذاری خطی با قید محلی

در روش‌های سنتی بعد از خوشه‌بندی، هیستوگرام تکرار هرکدام از این توصیفگرها تا مراکز خوشه‌ها را به دست آورده و آنها را جهت بازیابی تصاویر استفاده می‌کند. رویکرد دیگری به جای محاسبه توزیع کلمات کد در تصاویر آموزشی، این است که پس از خوشه‌بندی توصیفگرهای استخراج شده از الگوریتم SIFT، از کدگذاری خطی با قید محلی استفاده کرد. الگوریتم LLC از قیدهای محلی برای تصویر سازی هر توصیفگر در سیستم مختصات محلی استفاده می‌کند و مختصات تصویر سازی شده توسط روش ادغام ماکزیمم¹ برای بازنمایی نهایی ترکیب می‌شوند [20].

با فرض اینکه X توصیفگرهای محلی D بعدی باشند که از یک تصویر استخراج شده‌اند $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}$ و کتابچه کد $D = [b_1, b_2, \dots, b_M] \in \mathbb{R}$ رویکردهای کدگذاری مختلفی وجود دارد که هر توصیفگر را به یک کد M بعدی برای بازنمایی نهایی تصویر تبدیل می‌کند. یکی از این روش‌ها استفاده از کدگذاری پراکنده² می‌باشد. تابع هدف روش کدگذاری پراکنده به صورت زیر می‌باشد.

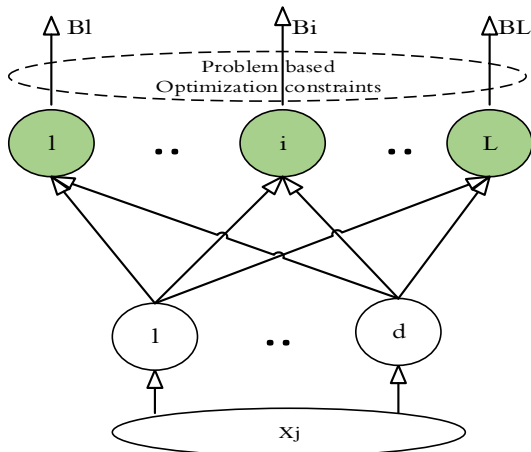
¹Max Pooling

²Sparse Coding

³Better reconstruction

ELM مزایای زیادی نسبت به همتایان خود دارد از جمله این برتری‌ها می‌توان به استفاده آسان، سرعت بالای آموزش، عدم رخداد بیش‌برازش^۴، قدرت تعمیم بالاتر و قابلیت استفاده از توابع فعال‌سازی و توابع هسته بسیار اشاره کرد [۲۱].

شکل ۳ معماری شبکه ELM با L نرون پنهان را نشان می‌دهد که a_i پارامترهای لایه پنهان شامل وزن ورودی به لایه پنهان و بایاس می‌باشند و وزن لایه پنهان L به نرون خروجی است.



شکل ۳: معماری ماشین یادگیر ELM [۲۱]

با داشتن تابع خروجی نرون پنهان $G(a, b, x)$ و جفت‌های آموزشی الگوریتم آموزش شبکه ELM به صورت زیر می‌باشد.

۱- پارامترهای نرون‌های پنهان یعنی وزن‌های لایه ورودی به لایه پنهان و بایاس $L, i = 1, 2, \dots, L$ به صورت تصادفی مقداردهی می‌شوند.

خروجی لایه پنهان یعنی ماتریس H از رابطه زیر به دست می‌آید. که در آن h_i خروجی لایه پنهان برای داده i ام است.

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L} \quad (8)$$

وزن‌های لایه پنهان به لایه خروجی به صورت زیر به دست می‌آید.

$$H\beta = T \quad (9)$$

$$\beta = H^+T$$

که در این رابطه شبه معکوس تعمیم یافته ماتریس H است [۲۱]. از برتری‌های الگوریتم ELM نسبت به شبکه عصبی پیشرو می‌توان به موارد زیر اشاره کرد:

شبکه عصبی نیاز به تنظیم پارامترها و انتخاب ساختار شبکه دارد، به عنوان مثال باید تعداد لایه‌های پنهان، تعداد نرون‌های لایه‌ی پنهان، نوع تابع فعال‌ساز، روش مورد استفاده برای محاسبه‌ی خطای سیستم به عنوان مثال (روش گرادیان نزولی)، آستانه‌ی مورد نیاز برای حداکثر خطای مطلوب به منظور توقف آموزش سیستم و نوع شبکه عصبی مورد استفاده (پرسپترون،

نمایش داده می‌شود و در این روش همبستگی بین توصیفگرهای مشابه با به اشتراک گذاری بردارهای پایه حفظ می‌شود.

پراکنندگی محلی نرم^۱: مشابه روش LLC در روش SC خطای بازسازی توسط استفاده از چندین بردار پایه پایین می‌باشد. با این حال نرم تنظیم با استفاده از نرم SC نمی‌باشد. به دلیل یادگیری Over Complete کتابچه کد در SC، این فرآیند ممکن است بردارهای پایه‌ی کاملاً متفاوتی را برای الگوهای یکسان انتخاب کند تا پراکنندگی را حفظ کند، بنابراین همبستگی بین کدها را از دست می‌دهد. در مقابل تطبیق دهنده محلی در LLC این اطمینان را فراهم می‌کند که الگوهای مشابه کدهای مشابهی داشته باشند

بنابراین ورودی الگوریتم LLC مراکز خوشه‌های به دست آمده از مراحل قبل و توصیفگرهای مربوط به داده‌های آموزش و آزمون هستند. و خروجی مورد انتظار بهترین ضرایب بازسازی برای هر کدام از توصیفگرها تا مراکز خوشه‌ها می‌باشد.

۳-۵ طبقه بند

در این مقاله دو رویکرد استخراج ویژگی مورد استفاده قرار گرفته است. ابتدا با استفاده از تکنیک کیسه واژگان بصری توسط یک بردار هیستوگرام ویژگی‌ها بازنمایی شده‌اند و در ادامه استفاده از ضرایب بدست آمده از رویکرد کدگذاری خطی با قید محلی بررسی شده است. پس از استخراج ویژگی می‌توان از بردارهای ویژگی استخراج شده به عنوان ورودی یک الگوریتم یادگیری ماشین برای طبقه‌بندی تصاویر ورودی استفاده کرد.

۳-۶ ساختار ماشین یادگیر پیشینه جهت دسته‌بندی و

ارزیابی سیستم

هدف در این قسمت، دسته‌بندی ویژگی‌های استخراج شده از تصاویر با استفاده از ماشین یادگیر پیشینه^۲ است. ماشین یادگیر پیشینه که توسط هونگ و همکارانش ارائه شد [۲۱]، از معماری یک شبکه عصبی تک لایه پیشرو^۳ استفاده می‌کند. وزن‌های لایه ورودی به پنهان به صورت تصادفی انتخاب می‌شوند و وزن‌های خروجی از طریق عملیات شبه معکوس به دست می‌آیند. ELM از قابلیت تعمیم‌پذیری بالاتر و زمان آموزش کمتری نسبت به شبکه عصبی برخوردار است. همچنین ELM نسبت به سایر روش‌ها از مداخله انسانی کمتری برخوردار است. تمامی پارامترهای شبکه به صورت خودکار تعیین می‌شوند که این امر باعث جلوگیری از مداخله انسانی شده و این روش را در کاربردهای برخط کارآمد می‌کند.

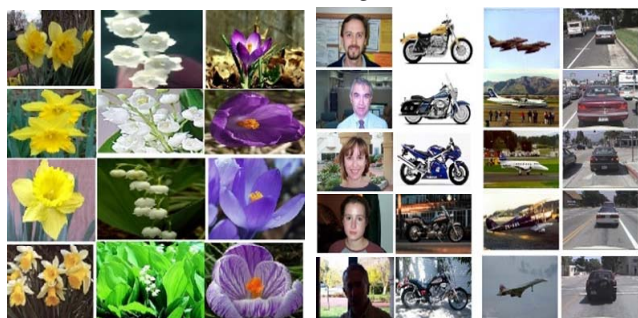
¹Local smooth sparsity

²Extream Learning Machine(ELM)

³Single Layer Feedforward Neural network(SLFN)

⁴Overfitting

خروجی ماشین یادگیر بیشینه نیز به تعداد کلاس‌ها خواهد بود. نمونه‌ای از تصاویر این پایگاه داده‌ها در شکل ۴ نشان داده شده است.



شکل ۴: تصاویر سمت راست نمونه‌ای از پایگاه داده Caltech-101 و سمت چپ پایگاه داده Flowers-17

در آزمایشات انجام شده ویژگی‌های که توسط رویکرد LLC و روش‌های سنتی استخراج شده‌اند برای بازیابی تصاویر براساس محتوا به سه دسته‌بند داده شده است. در جدول ۱ مقایسه روش ارائه شده با ویژگی‌های مختلف و روش‌های دسته‌بندی مختلف آورده شده است.

جدول ۱: بررسی روش‌های مختلف دسته‌بندی برای ویژگی‌های LLC روی پایگاه داده Caltech-101 و Flowers-17

پایگاه داده	ویژگی	روش دسته‌بندی	دقت دسته‌بندی
Caltech-101	ویژگی‌های LLC	ماشین بردار پشتیبان	۹۷/۵%
	ویژگی‌های LLC	معیار فاصله اقلیدسی	۹۳%
	ویژگی‌های LLC	ماشین یادگیر بیشینه	۹۸/۵%
	هیستوگرام تکرار توصیفگرها	ماشین بردار پشتیبان	۹۴%
	هیستوگرام تکرار توصیفگرها	معیار فاصله اقلیدسی	۸۵%
	هیستوگرام تکرار توصیفگرها	ماشین یادگیر بیشینه	۹۵%
Flowers-17	ویژگی‌های LLC	ماشین بردار پشتیبان	۹۵/۸%
	ویژگی‌های LLC	معیار فاصله اقلیدسی	۸۹/۶%
	ویژگی‌های LLC	ماشین یادگیر بیشینه	۹۷/۹%
	هیستوگرام تکرار توصیفگرها	ماشین بردار پشتیبان	۹۱%
	هیستوگرام تکرار توصیفگرها	معیار فاصله اقلیدسی	۸۲%
	هیستوگرام تکرار توصیفگرها	ماشین یادگیر بیشینه	۹۳%

در مسئله دسته‌بندی زمانی که دسترسی به تمامی داده‌های آموزشی وجود داشته باشد، می‌توان به راحتی سیستم را آموزش داد و جامعیت و قابلیت تعمیم نیز بالا خواهد بود. اما در حالتی که دسترسی به تمام داده‌های آموزشی امکان پذیر نباشد، فاصله اقلیدسی می‌تواند معیار مناسبی برای ارزیابی سیستم باشد. البته باید توجه داشت که نتایج به

MLP، RBF، شبکه‌های خودسازمان یافته و غیره) از پیش تعیین شود. در حالی که در ماشین یادگیر بیشینه تنها تعداد نرون‌های لایه پنهان و نوع تابع فعال ساز تعیین می‌شود.

مزیت دیگر ماشین یادگیر بیشینه این است که وزن‌ها و بایاس لایه‌ی ورودی بدون توجه به نوع مسئله به صورت تصادفی تعیین می‌شوند و وابسته به داده‌ها نمی‌باشند.

آموزش ماشین یادگیر بیشینه بسیار سریع تر از شبکه عصبی پیشرو بوده و تنها با محاسبه‌ی وزن‌های لایه‌ی پنهان به لایه‌ی خروجی می‌توان سیستم را مورد ارزیابی قرار داد.

قابلیت تعمیم ماشین یادگیر بیشینه نسبت به شبکه عصبی بیشتر می‌باشد.

ورودی ماشین یادگیر بیشینه همان ضرایبی هستند که با استفاده از الگوریتم LLC به دست آمده و تعداد لایه‌های آن ۵۰۰ عدد و نوع تابع فعال ساز آن سیگموئید^۱ است، و خروجی آن هم به تعداد کلاس‌های مربوط به پایگاه داده خواهد بود.

۴ آزمایشات و نتیجه گیری

همان گونه که در بخش قبل اشاره شد در روش پیشنهادی به صورت خلاصه پس از استخراج نقاط کلیدی SIFT و محاسبه توصیفگرهای نقاط کلیدی، برای هر مجموعه از تصاویر، با استفاده از رویکرد TF-IDF کتابچه کد بهینه سازی می‌شود و در مرحله بعد پس از خوشه‌بندی توصیفگرهای به دست آمده، از الگوریتم کدگذاری خطی با قید محلی برای تولید ضرایب بازسازی استفاده می‌شود که از این ضرایب به عنوان ویژگی‌های نهایی استفاده خواهد شد. پس از تعیین تعداد ویژگی‌های مطلوب برای دسته‌بندی، از دسته‌بند ماشین یادگیر بیشینه و دسته‌بند ماشین بردار پشتیبان و فاصله اقلیدسی برای بازیابی استفاده می‌شود.

الگوریتم پیشنهادی با استفاده از نرم افزار متلب نسخه 9.4 و بر روی سیستمی با مشخصات Intel Core i5 با قدرت پردازش 3.2 GH و با حافظه اصلی 16GH پیاده سازی و اجرا شد.

در این مقاله از دو پایگاه داده Caltech-101 و Flowers-17 استفاده می‌شود. پایگاه داده Caltech-101 شامل ۹۱۴۵ تصویر در ۱۰۱ کلاس شامل تصویر انسان، موتورسیکلت، هواپیما، وسایل نقلیه، حیوانات، گل و غیره می‌باشد، که تعداد تصاویر در هر کلاس متفاوت بوده و بین ۳۱ تا ۸۰۰ عدد می‌باشد. از چالش‌های مطرح در این پایگاه داده می‌توان به این اشاره کرد که از همه نوع تصویری در این پایگاه داده موجود می‌باشد و شدت نور و سایه و فاصله‌های زیاد بعضی تصاویر می‌باشد.

پایگاه داده Flowers-17 شامل ۱۳۶۰ تصویر در ۱۷ کلاس از گل‌های رایج در کشور انگلستان می‌باشد. تصاویر این پایگاه داده در مقیاس‌های بزرگ تهیه شده و دارای چالش تغییرات شدت روشنایی و تغییر زاویه تصویر برداری می‌باشد. همچنین گل‌های موجود در کلاس‌های مختلف شباهت زیادی با یکدیگر دارند. در آزمایشات انجام شده ۷۰ درصد از داده‌ها برای آموزش و ۳۰ درصد برای آزمون در نظر گرفته می‌شود.

¹ Sigmoidal

	1	2	3	4	
1	2 25.0%	0 0.0	0 0.0	1 1.0	96.0 4.0
2	0 0.0	2 25.0	0 0.0	0 0.0	100 0.0
3	0 0.0	0 0.0	2 24.0	0 0.0	100 0.0
4	0 0.0	0 0.0	1 1.0	2 24.0%	95.8 4.2
	100 0.0	100 0.0	95.8 4.2	95.8% 4.2	97.9 2.1
	1	2	3	4	

شکل ۶: ماتریس درهم ریختگی ضرایب LLC توسط ماشین یادگیر بیشینه برای ۴ کلاس از پایگاه داده 17-Flowers

در شکل ۶ ماتریس درهم ریختگی مربوط به دسته بندی ضرایب LLC توسط ماشین یادگیر بیشینه برای چهار کلاس از پایگاه داده 17-Flowers آورده شده است. همان طور که از ماتریس شکل ۶ نیز مشخص است در کلاس اول و کلاس دوم، خطای رخ نداده است و تمامی داده های هر دو کلاس به درستی دسته بندی شده اند. در کلاس سوم ۱ داده به اشتباه در کلاس ۴ قرار گرفته است. در کلاس چهارم ۱ خطا رخ داده است که به اشتباه در کلاس ۱ قرار گرفته است. در نهایت دقت سیستم ۹۷/۹ درصد به دست آمده است.

جدول ۲ حاصل مقایسه روش پیشنهادی و روش های ارائه شده توسط مقالاتی است که روش خود را بر روی پایگاه داده 101-Caltech انجام داده اند.

جهت بازیابی تصاویر براساس محتوا ابتدا از رویکرد آشکارسازی اشیاء برجسته در تصویر استفاده می شود. سپس از تصویر بخش بندی شده ویژگی استخراج شده و این ویژگی ها توسط الگوریتم SVM دسته بندی می شوند. با توجه به این که پس زمینه موجود در تصاویر یکی از مشکلاتی است که فرآیند استخراج ویژگی و بازیابی را با مشکل مواجه می کند [۸].

با استفاده از رویکرد بخش بندی تصاویر ابتدا ناحیه شیء مورد نظر جدا سازی شده و پس از استخراج ویژگی دسته بندی با استفاده از شبکه عصبی انجام می شود [۹].

روشی جدید برای بازیابی تصاویر براساس محتوا بر اساس ترکیب ویژگی های رنگی و شکلی ارائه شده است. در این روش از میانگین، واریانس و چولگی به عنوان ممان های رنگی استفاده می شود و برای استخراج ویژگی های شکلی از کانولوشن فیلترهای افقی، عمودی و قطری بر روی تصویر استفاده می شود. در نهایت از فاصله اقلیدسی برای ارزیابی و بازیابی تصاویر استفاده می شود [۱۰].

دست آمده توسط معیار شباهت فاصله اقلیدسی نسبت به سیستم های هوشمند دسته بندی ضعیف تر خواهد بود.

با توجه به نتایج می توان به این نتیجه رسید که ویژگی های به دست آمده توسط ضرایب LLC نسبت به ویژگی های سنتی (هیستوگرام تکرار توصیفگرها) نتیجه بهتری را در بازیابی تصاویر فراهم می کند. همچنین ماشین یادگیر بیشینه نسبت به دیگر دسته بندیها نتیجه بهتری را در دسته بندی فراهم می کند.

ضمناً از آنجایی که داده های پایگاه داده 17-Flowers خیلی به هم شبیه می باشد، دقت به دست آمده نسبت به نتایج حاصل از پایگاه داده 101-Caltech کمتر است.

همچنین جهت نشان دادن جزئیات دسته بندی در شکل ۵ ماتریس درهم ریختگی مربوط به دسته بندی ضرایب LLC توسط ماشین یادگیر بیشینه برای چهار کلاس از پایگاه داده 101-Caltech آورده شده است. در این ماتریس محور افقی خروجی حقیقی یا برجسته کلاس ها است و محور عمودی خروجی تخمین زده شده توسط دسته بندی را نشان می دهد.

	1	2	3	4	
1	49 24.5%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	50 25.0%	0 0.0%	1 0.5%	98.0% 2.0%
3	0 0.0%	0 0.0%	48 24.0%	0 0.0%	100% 0.0%
4	1 0.5%	0 0.0%	2 1.0%	49 24.5%	94.2% 5.8%
	98.0% 2.0%	100% 0.0%	96.0% 4.0%	98.0% 2.0%	98.50% 2.0%
	1	2	3	4	

شکل ۵: ماتریس درهم ریختگی ضرایب LLC توسط ماشین یادگیر بیشینه برای ۴ کلاس از پایگاه داده 101-Caltech

همان طور که در شکل ۵ می بینید، در کلاس اول یک مورد داده متعلق به کلاس ۱ در کلاس ۴ دسته بندی شده است که منجر به خطا شده است. همچنین ۲ مورد از داده های کلاس ۳ در کلاس چهارم دسته بندی شده است و یک داده متعلق به کلاس ۴ در کلاس دوم قرار گرفته است و در کلاس دوم خطایی رخ نداده است و تمامی ۵۰ داده این کلاس به درستی دسته بندی شده اند. در نهایت دقت سیستم که در مربع آبی رنگ آورده شده است ۹۸/۵۰ درصد می باشد.

جدول ۳ حاصل مقایسه روش پیشنهادی و روش‌های ارائه شده توسط مقالاتی است که روش خود را بر روی پایگاه داده Flowers-17 انجام داده‌اند.

یک بازیابی تصویر تعاملی هدایت‌گرانه معنایی، شکاف معنایی بین ویژگی‌های سطح پایین تصاویر و مفاهیم معنایی سطح بالا را کاهش می‌دهد. ایده اصلی شامل بازیابی تصاویر بر اساس ویژگی‌های بصری آنها است، تفسیر این تصاویر با مفاهیم آنها، توسط علم هستان‌شناسی که به وسیله کاربر تعریف می‌شود، توسعه می‌یابد. بازیابی تصاویر بر اساس ویژگی‌های ابتدایی آنها شامل، شکل، بافت، رنگ و اطلاعات مکانی بسیار مهم می‌باشد [۱۴].

رویکردی برای ترکیب اطلاعات رنگی با ویژگی‌های استخراج شده از شکل تصاویر ارائه شده است. برای استخراج ویژگی‌های شکلی از لبه‌های تصویر و نقاط کلیدی شامل گوشه‌ها استفاده می‌شود. نتایج آزمایشات بر روی پایگاه داده‌های مختلف جامعیت این روش را نشان می‌دهد [۱۵].

جدول ۳: مقایسه روش پیشنهادی و مقالات پایگاه داده Flowers-17

مرجع	روش	دقت بازیابی تصاویر
[14] F. Gonçaves 2019	بازیابی تصویر تعاملی	۹۳/۹۷%
[15] K. T. Ahmed 2019	استخراج ویژگی‌های شکلی از لبه‌های تصویر و نقاط کلیدی شامل گوشه‌ها	۹۵%
روش پیشنهادی	کدگذاری خطی با قید محلی	۹۷/۹%

نتایج آزمایشات نشان می‌دهد، ویژگی‌های استخراج شده از تصاویر، در دقت بازیابی و دسته‌بندی تأثیر بسزایی دارد، زیرا در صورتی که ویژگی‌های استخراج شده مناسب باشند، دقت بازیابی و دسته‌بندی افزایش می‌یابد.

۵ جمع بندی

در مباحث بازیابی تصاویر بر اساس محتوا، الگوهای مختلف، اشیاء متفاوت و بافت‌های متنوعی در تصویر وجود دارد. لذا فرآیند استخراج ویژگی بسیار حائز اهمیت می‌باشد. در روش‌های سنتی پس از استخراج ویژگی و محاسبه توصیفگرهای مربوط به هر دسته توسط الگوریتم SIFT، سپس از هیستوگرام تکرار توصیفگرها در هر دسته به عنوان ویژگی برای داده‌های آموزشی و آزمون استفاده شده و عملیات بازیابی و دسته‌بندی انجام می‌شد. در این مقاله پس از استخراج نقاط کلیدی SIFT و محاسبه توصیفگرهای نقاط کلیدی، برای هر مجموعه از تصاویر، با استفاده از رویکرد TF-IDF کتابچه کد بهینه سازی می‌شود و عملیات خوشه‌بندی بر روی توصیفگرها انجام می‌شود. که این

از ویژگی‌های رنگی در فضای RGB برای بازیابی تصاویر بر اساس محتوا استفاده شده است. در این روش برای افزایش دقت بازیابی تصاویر، در ابتدا ناحیه مربوط به اشیاء برجسته در تصویر با استفاده از رویکرد خوشه‌بندی جابه‌جایی میانگین مشخص می‌شود. پس از مشخص شدن ناحیه پس زمینه و ناحیه مربوط به شیء برجسته در تصویر، عملیات استخراج ویژگی انجام خواهد شد. از معایب این روش می‌توان به زمان‌بر بودن فرآیند بخش‌بندی تصویر اشاره کرد [۱۲].

از ویژگی هیستوگرام رنگ و ضرایب گابور به منظور بازیابی تصاویر استفاده می‌شود. پس از استخراج و ترکیب ویژگی‌ها با هم، از فاصله منتهن برای بازیابی تصاویر استفاده می‌شود. که یکی از مزیت‌های این روش ترکیب ویژگی‌های رنگی و بافتی است. ولی توجهی به تصاویر سطح خاکستری ندارد [۱۳].

روشی جدید برای بازیابی تصاویر بر اساس محتوا ارائه شده است. در این روش از ترکیب ویژگی‌های سطح پایین، رنگ، بافت و شکلی استفاده می‌شود. برای استخراج ویژگی‌های رنگی از هیستوگرام رنگ استفاده می‌شود، برای ویژگی‌های شکلی از هیستوگرام لبه‌های تصویر استفاده می‌شود و برای استخراج ویژگی‌های بافتی از فیلتر گابور به همراه ماتریس هم‌رخدادی استفاده می‌شود. قبل از ترکیب ویژگی‌ها هر یک از این ویژگی‌ها به تنهایی برای بازیابی مورد استفاده قرار می‌گیرند تا بر اساس نتایج به دست آمده یک امتیاز به آنها اختصاص داده شود، در مرحله بعد بر اساس امتیازات به دست آمده عملیات ترکیب انجام می‌شود [۲۲].

جدول ۲: مقایسه روش پیشنهادی و مقالات پایگاه داده Caltech-101

مرجع	روش	دقت بازیابی تصاویر
[7] Zhang 2017	استفاده از نواحی برجسته	۸۹/۲%
[8] Z. Li, J. Long 2016	آشکارسازی اشیاء برجسته در تصویر	۹۱%
[9] R. Ashraf, K. Bashir 2015	جداسازی ناحیه اشیاء	۸۲%
[10] S. P. Rana, M. 2019	ترکیب ویژگی‌های رنگی و شکلی	۶۸%
[11] S. Fakhfakh 2019	استفاده از رنگ، بافت، نقاط کلیدی	۷۹%
[12] Y. Cui 2010	ویژگی‌های رنگی، خوشه بندی جابه‌جایی میانگین	۸۹%
[13] A. Anandh, 2016	هیستوگرام رنگ و ضرایب گابور	۸۰%
[22] H. Zhou 2010	ترکیب ویژگی‌های سطح پایین، رنگ، بافت و شکلی	۹۱/۵%
روش پیشنهادی	کدگذاری خطی با قید محلی	۹۸/۵%

- [7] J. Zhang, S. Feng, D. Li, Y. Gao, Z. Chen, and Y. Yuan, "Image retrieval using the extended salient region", *Information Sciences*, vol. 399, pp. 154-182, 2017.
- [8] Z. Li, J. Long, and C. Dong, "Visual Attention Model and Relevant Feedback based Image Retrieval", *Electronic Imaging*, vol. pp. 1-9, 2016.
- [9] R. Ashraf, K. Bashir, A. Irtaza, and M. T. Mahmood, "Content based imageretrieval using embedded neural networks with bandletized regions", *Entropy*, vol. 17, pp. 3552-3580, 2015.
- [10] S. P. Rana, M. Dey, and P. Siarry, "Boosting content based image retrieval performance through integration of parametric & nonparametric approaches", *Journal of Visual Communication and Image Representation*, vol. 58, pp. 205-219, 2019.
- [11] S. Fakhfakh, M. Tmar, and W. Mahdi, "Image Retrieval Based on Using Hamming Distance", *Procedia Computer Science*, vol. 73, pp. 320-327, 2015.
- [12] Y. Cui and B. Xiao, "Image-based accurate object retrieval combined with color invariant", in *Intelligent Computation Technology and Automation (ICICTA)*, 2010 International Conference on, pp. 478-481, 2010.
- [13] A. Anandh, K. Mala, and S. Suganya, "Content based image retrieval system based on semantic information using color, texture and shape features", in *Computing Technologies and Intelligent Data Engineering (ICCTIDE)*, International Conference on, pp. 1-8, 2016.
- [14] F. M. F. Gonçalves, I. R. Guilherme, and D. C. G. Pedronette, "Semantic guided interactive image retrieval for plant identification," *Expert Systems with Applications*, vol.91, PP.12-26, 2019.
- [15] K. T. Ahmed, S. Ummesafi, and A. Iqbal, "Content based image retrieval using image features information fusion," *Information Fusion*, vol. 51, PP. 76-99, 2019.
- [16] Alsmadi, M.K., "Content-Based Image Retrieval Using Color, Shape and Texture Descriptors and Features", *Arab J Sci Eng*, vol 45, pp. 3317-3330 2020.
- [17] P. Chhabra, N.K Garg, & M. Kumar, "Content-based image retrieval system using ORB and SIFT features", *Neural Comput & Applic*, vol. 32, pp. 2725-2733, 2020.

عملیات خوشه‌بندی موجب می‌شود تا هم‌پوشانی ویژگی‌ها با یکدیگر از بین برود و قابلیت تمایز مناسبی ایجاد شود. پس از یافتن توصیفگرهای کاندیدای هر دسته به جای استفاده از هیستوگرام تکرار توصیفگرها در هر دسته از ضرایب بازنمایی توصیفگرهای هر دسته با توجه به نماینده‌های تولید شده از مرحله قبل توسط الگوریتم LLC به عنوان ویژگی استفاده شد. در نهایت از این ویژگی‌های تولید شده برای طبقه‌بندی تصاویر بازیابی شده استفاده شد. نتایج آزمایشات برتری روش ارائه شده را نسبت به الگوریتم‌های مطرح نشان می‌دهد.

برای ادامه این پژوهش می‌توان بر روی بخش استخراج ویژگی جهت بازنمایی تصاویر تمرکز کرد. با توجه به این که در روش SIFT پس از استخراج نقاط کلیدی، فرآیند محاسبه توصیفگرها زمان‌بر می‌باشد، می‌توان از الگوریتم‌های دیگری مانند الگوریتم SURF که تعداد توصیفگرهای کمتری دارد، استفاده کرد. برای کارکرد بهتر و کارسازتر، از ترکیب دو ویژگی (الگوی دودویی محلی و SIFT با سیستم فازی خودکار) یا ترکیب ویژگی‌های محلی و سراسری تصویر استفاده کرد.

رویکرد دیگر استفاده از الگوریتم‌های تکاملی بجای رویکرد کدگذاری خطی باقیمانده محلی می‌باشد یعنی بجای محاسبه ضرایب هر کدام از توصیفگرها تا مراکز خوشه‌ها و اختصاص ضریب به آنها از الگوریتم‌های مطرح در این زمینه همانند (ذرات تجمعی) استفاده نمود.

مراجع

- [1] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, et al, "Query by image and video content: The QBIC system", *computer*, vol. 28, pp. 23-32, 1995.
- [2] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues", *Journal of visual communication and image representation*, vol.10, pp. 39-62, 1999.
- [3] N. Sebe and M. S. Lew, "Salient Points for Content-Based Retrieval", in *BMVC*, pp. 401-410, 2001.
- [4] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the HSV color space for image retrieval", in *Proceedings. International Conference on Image Processing*, pp. II-II, 2002.
- [5] G-H. Liu and J-Y. Yang, "Content-based image retrieval using color difference histogram", *Pattern Recognition*, vol. 46, pp. 188-198, 2013.
- [6] P. Shanmugavadivu, P. Sumathy, and A. Vadivel, "FOSIR: Fuzzy-Object-Shape for Image Retrieval applications", *Neurocomputing*, VOL.171, pp. 719-735, 2016.

- [18] C. Moulin, C. Barat, and C. Ducottet, "Fusion of tf-idf weighted bag of visual features for image classification", in Content-Based Multimedia Indexing (CBMI), 2010 International Workshop on, pp. 1-6, 2010.
- [19] F. Tajeripour, M. Saberi, and S. Fekri, "Developing a Novel Approach for Content Based Image Retrieval Using Modified Local binary patterns and Morphological Transform", in Arab Journal of Information technology, vol. 12, no.6, 2015.
- [20] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification", in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp. 3360-3367, 2010.
- [21] G.-B. Huang, "An insight into extreme learning machines: random neurons, random features and kernels", Cognitive Computation, pp. 1-15, 2017.
- [22] H. Zhou, G. Schaefer, M. E. Celebi, H. Iyatomi, K.-A. Norton, T. Liu, et al., "Skin lesion segmentation using an improved snake model", in Engineering in Medicine and Biology Society (EMBC), 2010, pp. 1974-1977, 2010.

پایگاه داده برای چاپ