

## حاشیه‌نویسی خودکار تصاویر با استفاده از واژه‌های بصری

فرزین یغمایی<sup>۱</sup>، استادیار، وفا میهمی<sup>۲</sup>، دانشجوی دکتری، علیرضا نوحی<sup>۳</sup>، کارشناسی ارشد

۱- دانشکده مهندسی برق و کامپیوتر - دانشگاه سمنان - سمنان - ایران - f\_yaghmaee@semnan.ac.ir

۲- دانشکده مهندسی برق و کامپیوتر - دانشگاه سمنان - سمنان - ایران - maihami@semnan.ac.ir

۳- دانشکده مهندسی برق و کامپیوتر - دانشگاه سمنان - سمنان - ایران - noohi@semnan.ac.ir

**چکیده:** افزایش تصاویر موجود در کاربردهای مختلف زندگی، از جمله اینترنت سبب شده است که در سال‌های اخیر وجود معنا در وب به‌ویژه در مورد تصویر مورد توجه محققان قرار گیرد. حاشیه‌نویسی تصویر به معنای انتصاب یک یا چند کلمه برای توصیف تصویر است. ورودی سیستم حاشیه‌نویسی، ویژگی‌های استخراج‌شده از تصویر است. این مقاله، الگوریتم جدیدی برای حاشیه‌نویسی تصویر ارائه می‌دهد که در آن برای استخراج ویژگی‌های تصاویر، از مفهوم کیف واژگان (BoW) استفاده شده است و از توصیفگر SIFT نیز برای این منظور کمک گرفته شده است. برای رسیدن به کارایی مناسب و با توجه به بالا بودن ابعاد ویژگی‌های SIFT از روش کاهش ابعاد ویژگی PCA-SIFT و الگوریتم K-Means بهره گرفته شده است. آزمایش‌های انجام‌شده بر روی سیستم پیشنهادی و با استفاده از مجموعه تصاویر Corel5k، گویای عملکرد بهتر سیستم ارائه‌شده در دو معیار مورد بررسی زمان و دقت نسبت به کارهای گذشته است.

**واژه‌های کلیدی:** حاشیه‌نویسی خودکار تصویر، کیف واژگان، توصیفگر PCA، الگوریتم K-Means، PCA-SIFT

## Automatic Image Annotation Using Bag of Words

Farzin Yaghmaee, Assistant Professor<sup>1</sup>, Vafa Maihami, Ph.D Candidate<sup>2</sup>, Alireza Noohi, M.S Degree<sup>3</sup>

1- Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran, Email: F\_yaghmaee@semnan.ac.ir

2- Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran, Email: Maihami@semnan.ac.ir

3- Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran, Email: Noohi@semnan.ac.ir

**Abstract:** Due to increase using images in different life application especially internet, recently many researchers interested in understanding in web and images. Automatic image annotation means attaching words, keywords or comments to an image. The inputs for image annotation system are features which are extracted from image. In this paper, a new algorithm for automatic image annotation using bag of words (BOW) and SIFT descriptor is presented. Considering the high dimensionality of SIFT features and to achieve satisfying efficiency, we apply dimension reduced technique PCA-SIFT and K-Means algorithm. Experimental results based on the images of Corel5k dataset show that the proposed method has better performance in precision and time measures.

**Keywords:** Automatic image annotation, bag of words, K-means, PCA, PCA-SIFT.

تاریخ ارسال مقاله: ۱۳۹۵/۰۶/۰۷

تاریخ اصلاح مقاله: ۱۳۹۵/۰۷/۲۴ و ۱۳۹۵/۰۸/۳۰

تاریخ پذیرش مقاله: ۱۳۹۵/۰۹/۰۶

نام نویسنده مسئول: فرزین یغمایی

نشانی نویسنده مسئول: ایران - سمنان - روبروی پارک سوکان - پردیس یک دانشگاه سمنان - دانشکده مهندسی برق و کامپیوتر

## ۱- مقدمه

با افزایش تعداد تصاویر در وب، لزوم دسته‌بندی و بازیابی مؤثر اطلاعات اهمیت بیشتری پیدا کرده است. همین امر سبب شده است که در سال‌های اخیر وجود معنا در وب به‌ویژه در مورد مالتی مدیا از جمله تصویر مورد توجه محققان قرار بگیرد. اکثر روش‌های قدیمی از ویژگی‌های تصویر شبیه رنگ، بافت و شکل استفاده می‌کردند که کارایی چندانی نداشت. به این دلیل که کاربران ترجیح می‌دهند جستجوی تصویر را بر اساس پرس‌وجوهای متنی و بر اساس کلمات کلیدی انجام دهند. از نظر کاربر بیان درخواست بازیابی با استفاده از کلمات به‌منظور توصیف محتوای تصویر به‌مراتب ساده‌تر و سریع‌تر از ارائه تصویر نمونه برای بازیابی اطلاعات است. این امر لزوم حاشیه‌نویسی تصویر را بیشتر کرده است.

حاشیه‌نویسی تصاویر به فرآیند تولید کلماتی که محتوای تصویر را توصیف کنند، اطلاق می‌شود. هدف حاشیه‌نویسی تصاویر، تولید کلماتی است که تصاویر را به شکل مناسبی توصیف می‌کنند [۱، ۲]. در واقع در حاشیه‌نویسی مجموعه‌ای از کلمه یا کلماتی که بیانگر معنا و مفهوم واقعی تصویر است، با تصاویر همراه می‌شود. منظور از معنا و مفهوم واقعی مفاهیمی نزدیک به برداشت انسان‌ها از تصویر است. یافتن تصویر مورد نیاز یک کاربر از بین تصاویر بسیار زیاد، یک کار چالش‌انگیز است. تحقیقات نشان داده است که یک فاصله معنایی بین بازیابی تصاویر بر اساس محتوا و معنای تصویر قابل فهم برای انسان، وجود دارد [۳]. بهترین راه برای پر کردن این فاصله معنایی، حاشیه‌نویسی خودکار تصویر است. حاشیه‌نویسی تصویر می‌تواند توسط انسان یا ماشین انجام شود اما از آنجاکه مجموعه‌های بزرگی از تصاویر نیازمند حاشیه‌نویسی است، انجام این عمل توسط انسان زمان‌بر، پرهزینه و همراه با اشتباه یا ابهام است. لذا عملاً منظور از حاشیه‌نویسی تصویر، انجام این عمل به‌صورت خودکار و توسط ماشین است که به آن حاشیه‌نویسی خودکار تصاویر گفته می‌شود. برای تولید سیستم حاشیه‌نویسی خودکار باید قوانینی را به ماشین یاد داد تا بتواند با استفاده از آن‌ها، پس از دریافت تصویر ورودی، حاشیه‌های تصویر را تولید کند.

تشخیص اشیاء به‌عنوان یکی از حوزه‌های پرکاربرد بینایی ماشین شناخته می‌شود. الگوریتم‌های متعددی بر پایه کیف واژگان بصری<sup>۱</sup> در شناسایی اشیاء استفاده شده‌اند [۱۶-۱۲، ۱۸، ۱۹]. در بینایی ماشین، واژه‌های بصری یک بردار از تعداد وقوع یک واژه از ویژگی‌های تصویر است. واژه‌های بصری به دلیل سادگی و عملکرد خوبی که بر روی شیء و صحنه دارند مورد توجه قرار گرفته‌اند. همچنین به دلیل موفقیت استفاده از کیف واژگان در بازیابی اطلاعات، استفاده از آن در پردازش تصویر و ویدئو نیز مورد توجه است. در مدل کیف واژگان یک سند به‌صورت هیستوگرامی از واژه‌ها نشان داده می‌شود.

در این مقاله یک روش جدید حاشیه‌نویسی خودکار تصاویر با استفاده از کیف واژگان ارائه شده است. رویکرد جدید از سه گام تشخیص ویژگی، توصیف ویژگی و ایجاد کتاب‌کد<sup>۲</sup> برای هر تصویر استفاده می‌کند.

در مرحله آموزش ابتدا نقاط کلیدی هر تصویر با استفاده از توصیف‌گر SIFT<sup>۳</sup> استخراج شده و در ماتریس بزرگی ذخیره و همچنین مشخص می‌گردد که هر ویژگی متعلق به کدام کلاس است. سپس با استفاده از الگوریتم خوشه‌بندی K-means ویژگی‌های موجود در این ماتریس خوشه‌بندی می‌شوند. خروجی این مرحله، ماتریسی خواهد بود که در بردارنده مرکز هر خوشه است. در مرحله آزمون، یک تصویر به سیستم وارد شده و ویژگی‌های SIFT آن استخراج می‌شود. سپس فاصله تک‌تک این ویژگی‌ها با مراکز خوشه‌های مرحله آموزش محاسبه می‌شود. اگر این فاصله از مقدار آستانه تعیین‌شده کمتر باشد، این نقطه کلیدی به خوشه نسبت داده می‌شود و در نهایت تعداد ویژگی نسبت داده‌شده به هر خوشه (هیستوگرام ویژگی‌ها) برای این تصویر به دست می‌آید. در مرحله آخر خوشه‌هایی که بیشترین فراوانی ویژگی را دارند به‌عنوان حاشیه (برچسب) های تصویر انتخاب می‌شوند.

هدف اصلی در این مقاله ارائه یک روش جدید برای حاشیه‌نویسی خودکار تصاویر است. در بخش دوم مروری بر کارهای گذشته خواهیم داشت، سپس در بخش سوم سیستم طراحی‌شده به کمک واژه‌های بصری همراه با جزئیات ارائه می‌شود، نمودارها و استنتاج‌های لازم آمده است، در بخش چهارم نتایج تجربی الگوریتم پیشنهادی ارائه می‌گردد و در نهایت نتیجه‌گیری تحقیق در بخش پنجم بیان می‌شود.

## ۲- پیشینه تحقیق

تاکنون روش‌ها و سامانه‌های گوناگونی برای حاشیه‌نویسی تصاویر توسط افراد مختلف ارائه شده است که هدف و کاربرد آن‌ها با همدیگر متفاوت است. از نظر نوع سیستم حاشیه‌نویسی استفاده‌شده می‌توان آن‌ها را به سه دسته تقسیم کرد:

- روش‌های بر اساس دسته‌بندی: در این روش از دسته‌بندی‌های مختلف برای تقسیم تصاویر به دسته‌های مختلف استفاده می‌شود و بر اساس این دسته‌ها کلمات کلیدی مناسب هر دسته تعیین می‌شود. مقالات [۴، ۵] نمونه‌هایی از این دسته‌اند.

- روش‌های بر اساس مدل احتمالاتی: در مدل احتمالی، اساس کار استخراج ویژگی‌های بصری و سطح پایین تصویر (رنگ، بافت، شکل و ...) از داده‌های آموزشی و سپس بررسی توزیع احتمال وابسته بین ویژگی‌های بصری تصویر و ویژگی‌های معنایی سطح بالا است و در نهایت از نتایج توزیع احتمال برای حاشیه‌نویسی تصویر جدید استفاده می‌شود. مقالات [۶، ۷] نمونه‌هایی از این دسته‌اند.

- روش‌های بر اساس بازیابی تصاویر: دسته سوم روش‌هایی هستند که با آن‌ها مانند بازیابی تصویر رفتار می‌شود. یعنی ابتدا بهترین تصاویر مرتبط با تصویر جدید بر اساس محتوای تصویر بازیابی شده و سپس کلمات کلیدی تصاویر بازیابی شده برای تصویر آزمون نیز استفاده می‌شود. مقالات [۸، ۹] نمونه‌هایی از این دسته‌اند.

در مقاله‌ای که اخیراً توسط نویسندگان مقاله حاضر، ارائه شده است، روشی خودکار برای حاشیه‌نویسی تصاویری با موضوع اماکن تاریخی و

خود کارایی بهتری را فراهم و ارائه کند. در مرحله انتخاب ویژگی‌ها از مفهوم کیف ویژگی‌ها (BoF<sup>T</sup>) استفاده شده است [۱۸]. در بینایی ماشین، مدل کیف واژگان، برای دسته‌بندی تصاویر به‌وسیله در نظر گرفتن ویژگی‌های تصویر به‌عنوان واژه استفاده می‌شود. در طبقه‌بندی اسناد، کیف واژگان، برداری از تعداد وقوع واژه‌ها است.

برای نمایش یک تصویر با استفاده از مدل کیف واژگان با تصویر می‌توان مانند یک سند رفتار کرد. یعنی باید واژه‌ها نیز در یک تصویر تعریف شوند. برای دستیابی به این منظور، دو گام اصلی زیر نیاز است:

۱. تشخیص ویژگی و توصیف ویژگی: یک تعریف برای BoF می‌تواند "نمایش هیستوگرام مبتنی بر ویژگی‌ها" باشد. بازیابی تصاویر بر اساس محتوا می‌تواند با این فن نمایش تصویر نیز انجام پذیرد. بعد از تشخیص ویژگی هر تصویر به‌وسیله چندین تکه تصویر خلاصه می‌شود. روش نمایش ویژگی با چگونگی نمایش تکه‌های تصویر به‌عنوان بردارهای عددی مرتبط می‌شود، این بردارها توصیف‌گرهای ویژگی نامیده می‌شوند. یک توصیف‌گر مناسب باید قادر به مدیریت شدت رنگ، چرخش، تغییر سایز و تبدیلات باشد. یکی از محبوب‌ترین توصیف‌گرها، توصیف‌گر SIFT است که در این مقاله نیز از آن استفاده شده است. SIFT هر تکه تصویر را به یک بردار، ۱۲۸ بعدی تبدیل می‌کند. بعد از این مرحله هر تصویر، مجموعه‌ای از بردارهای با ابعاد یکسان خواهد بود.

۲. ایجاد کتاب‌کد: مرحله بعدی برای مدل BoF تبدیل بردارهای به‌دست‌آمده به کلمه‌کد<sup>۲</sup> است. یک کلمه‌کد می‌تواند به‌عنوان یک نماینده برای چندین تکه تصویر مشابه در نظر گرفته شود. یکی از روش‌ها، استفاده از خوشه‌بندی K-Means روی تمام بردارها است. کلمه‌کدها به‌عنوان مراکز خوشه‌های یادگیری در نظر گرفته می‌شوند. تعداد خوشه‌ها برابر اندازه کلمه‌کد است. بنابراین هر تکه از یک تصویر به یک کلمه‌کد مناسب نگاشت می‌شود و تصویر به‌وسیله هیستوگرامی از کلمه‌کدها نمایش داده می‌شود.

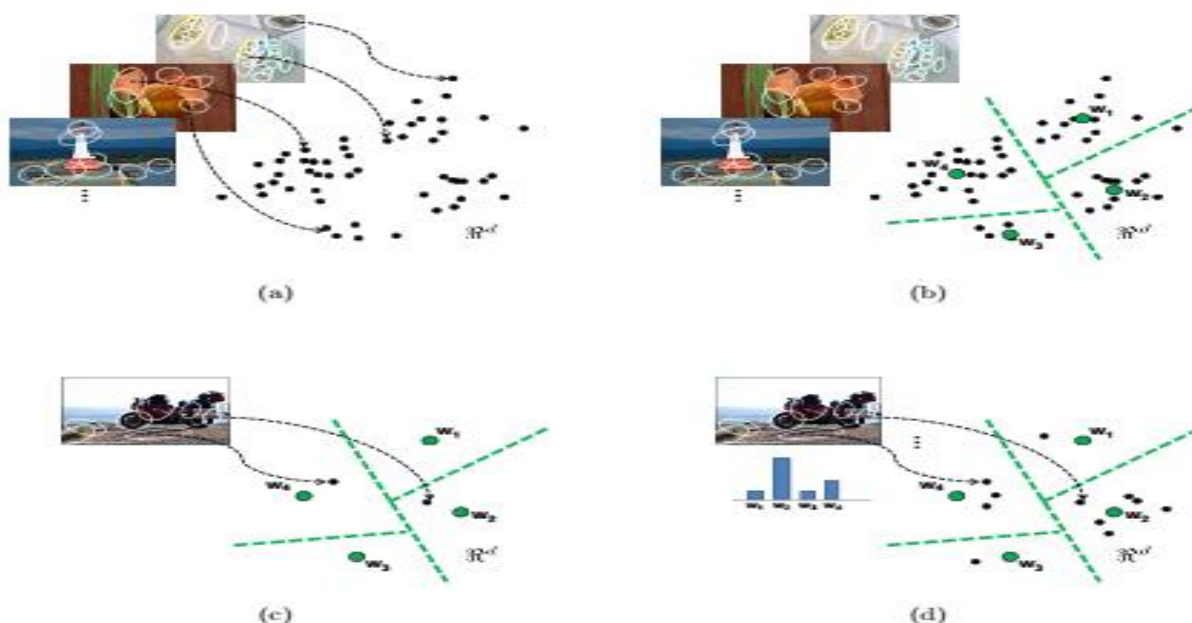
شکل ۱ مفهوم و شمای کلی پیاده‌سازی کیف واژگان و انتساب کلمات را نشان می‌دهد. برای هر تصویر مجموعه‌ای از توصیف‌گرها موجود است. با استفاده از الگوریتم خوشه‌بندی K-means توصیف‌گرها خوشه‌بندی می‌شود. برای این کار، تمام توصیف‌گرها، همه تصاویر را در یک ماتریس بزرگ قرار داده، که هر سطر ماتریس بیانگر یک توصیف‌گر است. سپس یک عدد دلخواه مثبت به K نسبت داده و ماتریس فوق به همراه مقدار K تعریف‌شده برای خوشه‌بندی توصیف‌گرها استفاده می‌شود. بر اساس آزمایش‌های صورت گرفته و بررسی خروجی‌ها، بهترین نتیجه به ازای  $K=200$  به دست آمده است (شکل ۲). بعد از خوشه‌بندی یک ماتریس C وجود دارد که هر سطر آن نشان‌دهنده مرکز هر خوشه در فضای ویژگی و تعداد سطرها برابر تعداد مراکز خوشه‌ها است. به‌عبارت‌دیگر، ماتریس C نگه‌دارنده مکان مراکز خوشه‌ها است. همچنین یک بردار IDX وجود دارد که هر عنصر آن دربردارنده برجستگی به ازای یک توصیف‌گر خاص است. با انجام این کار، خوشه‌بندی به پایان می‌رسد.

مذهبی ایران معرفی شده است [۱۰]. برای این منظور، ابتدا این تصاویر را به همراه متون اطراف آن‌ها ذخیره کرده و پس از انجام پیش‌پردازش‌های لازم، لغت‌نامه‌ای از کلمات موجود در متون ایجاد می‌شود. سپس با استفاده از روش‌های یادگیری ماشین و روش وزن دهی TF-IDF به هر کدام از کلمات وزنی نسبت داده می‌شود که نشان‌دهنده اهمیت هر یک از لغات است. برای تحلیل نتایج حاصل از الگوریتم، با کمک ۷ کاربر انسانی، ابتدا حاشیه‌نویسی دستی روی تصاویر انجام و به هر تصویر ۵ کلمه کلیدی نسبت داده می‌شود. نتیجه نشان‌دهنده شباهت حدود ۲/۸۶ کلمه (بیش از ۵۰ درصد)، بین تفسیر انسانی و الگوریتم پیاده‌سازی شده است. Sh.Wan [۱۱] از درخت تصمیم برای تولید سیستم حاشیه‌نویسی تصاویر استفاده کرده است. آن‌ها از یک چارچوب حاشیه‌نویسی مبتنی بر ناحیه بهره گرفته‌اند به طوری که در حاشیه‌نویسی مبتنی بر ناحیه، کلمات کلیدی با مناطق خاصی در مجموعه داده‌های آموزشی مرتبط هستند. در این مقاله از الگوریتم درخت تصمیم برای سیستم حاشیه‌نویسی استفاده شده است که این سیستم می‌تواند تعداد زیادی از تصاویر را سریع‌تر و بهتر طبقه‌بندی کند. الگوریتم ارائه‌شده در این مقاله از مجموعه تصاویر Corel5k برای آزمایش‌های خود استفاده کرده‌اند. نویسندگان مقاله تصویر را به تعدادی ناحیه تقسیم‌بندی کرده‌اند که هر ناحیه محتوای معنا و ویژگی‌های بصری متفاوتی دارد. ابتدا، تصویر به نواحی مختلفی تقسیم شده که ویژگی‌های بصری مشابهی دارند. سپس اطلاعات بصری هر ناحیه استخراج شده و مجموعه داده آموزشی ساخته می‌شود. سپس سیستم از الگوریتم ensemble برای یادگیری مجموعه آموزش استفاده می‌کند. هر کلاس یک برجسب دارد و برجسب کلاس می‌تواند کلمات کلیدی باشد. نتایج الگوریتم بر روی ۴۷۵ تصویر از مجموعه Corel و با استفاده از الگوریتم KNN دقت ۶۶/۹۵ را نشان می‌دهد. در مقاله [۱۲] به کاربرد بازنمایی تنک ساخت‌یافته در حاشیه‌نویسی تصاویر پرداخته شده است. در این مقاله ابتدا حاشیه‌نویسی خودکار تصاویر و بازنمایی تنک بررسی شده است و در ادامه مهم‌ترین پژوهش‌هایی که در حوزه حاشیه‌نویسی خودکار تصاویر با استفاده از بازنمایی تنک انجام شده است، مرور شده و در انتها از نظر ویژگی‌های مختلف با یکدیگر مقایسه شده است. همچنین در مقاله [۱۳] یک روش جدید برای حاشیه‌نویسی خودکار تصاویر به کمک فازی کردن روش رأی‌گیری همسایگی با استفاده از نزدیک‌ترین تصاویر به تصویر پرس‌وجو ارائه شده است.

در مقاله مروری [۱۴] به استفاده از کیف واژگان در حاشیه‌نویسی تصاویر پرداخته شده است و کارهای مرتبط با این حوزه از سال ۲۰۰۶ تا ۲۰۱۲ در آن مورد بررسی و مقایسه قرار گرفته است. در مقالات [۱۷-۱۵] به استفاده از الگوریتم مدل کیف کلمات در تشخیص اشیا و مفاهیم در تصاویر پرداخته شده است.

### ۳- الگوریتم پیشنهادی با استفاده از واژه‌های بصری

هدف اصلی این تحقیق طراحی و پیاده‌سازی سامانه‌ای خودکار برای حاشیه‌نویسی مجموعه‌ای از تصاویر است که در مقایسه با انواع مشابه



شکل ۱: شمای کلی پیاده‌سازی واژه‌های بصری و انتساب کلمات (a) مجموعه‌ای از تصاویر و ویژگی‌های استخراج شده از آن‌ها. (b) خوشه‌بندی ویژگی‌ها و تعیین مراکز خوشه‌ها (c) تصویر ورودی جدید و تعیین نزدیک‌ترین واژه بصری به هر ویژگی از تصویر جدید (d) هیستوگرام ایجاد شده برای تصویر ورودی که مشخص می‌کند هر واژه بصری چند بار در تصویر تکرار شده است.

#### ۴- آزمایش‌ها و نتایج

برای ارزیابی روش پیشنهادی از مجموعه داده Corel5k استفاده شده است. در این پایگاه داده ۵۰۰۰ تصویر وجود دارد که به صورت دستی توسط انسان حاشیه‌نویسی شده‌اند. در میان این تصاویر ۵۰۰ تصویر به‌عنوان آزمون در این مجموعه از گروه‌های متفاوت انتخاب شده است و نتایج به‌دست‌آمده حاصل میانگین نتایج بر روی تصاویر منتخب است. برای ارزیابی روش پیشنهادی از دو معیار میانگین دقت تصاویر آزمایش و زمان آموزش استفاده شده است. دقت هر تصویر آزمایش نیز تعداد واژه‌های درست تشخیص داده‌شده تقسیم بر کل واژه‌های تشخیص داده شده توسط الگوریتم برای هر تصویر آزمایش و میانگین دقت برابر با مجموعه دقت به‌دست‌آمده برای همه تصاویر آزمایش تقسیم بر تعداد تصاویر آزمون است.

در ابتدا و در آزمایش اول از زیرمجموعه‌ای از تصاویر مجموعه تصاویر Corel5k استفاده شده است. به صورتی که ۱۰۰۰ تصویر همراه با برچسب‌های آن انتخاب شده است که در آن ۸۰۰ تصویر برای آموزش و ۲۰۰ تصویر برای آزمایش در نظر گرفته شده است. جدول ۱ دقت به‌دست‌آمده برای الگوریتم پیشنهادی را در مقایسه با پژوهش وان و همکاران نشان می‌دهد. در پژوهش وان و همکاران از ۴۷۵ تصویر استفاده شده است در حالی که در الگوریتم پیشنهادی از ۱۰۰۰ تصویر استفاده شده است و با این وجود دقت بهتری را نشان داده است. همچنین شکل ۳ دو نمونه از خروجی‌های الگوریتم پیشنهادی را نشان می‌دهد. در مقایسه با کارهای انجام‌شده، روش پیشنهادی حاشیه‌نویسی تصاویر از دقت بالایی برخوردار بوده و با توجه به افزایش تعداد تصاویر و کلاس‌های مختلف تصویری نتایج قابل‌تأمل و بیانگر بهبود است.

برای هر تصویر هیستوگرامی از نقاط کلیدی ایجاد می‌شود، که تعداد وقوع هر ویژگی (مرکز خوشه) را در یک تصویر خاص نشان می‌دهد. از آنجایی که برچسب تمام توصیف‌گرهای یک گروه (مرکز خوشه) مشخص است در نتیجه می‌توان به‌سادگی تعداد وقوع هر ویژگی را با توجه به برچسب توصیف‌گرها محاسبه کرد. بعد از قرار دادن همه هیستوگرام‌ها (بردارهای سطری) در یک ماتریس، ماتریس بزرگی ایجاد می‌شود که هر سطر آن هیستوگرامی از یک تصویر است به‌طوری‌که هر ستون مربوط به یک مرکز خاص و نشان‌دهنده تعداد وقوع در تصویر است. در مرحله بعد، برای هر تصویر در مجموعه آموزش و آزمایش هیستوگرام آن محاسبه و تشکیل می‌شود.

در مرحله آموزش ابتدا نقاط کلیدی هر تصویر با استفاده از توصیفگر SIFT استخراج شده و در ماتریس بزرگی ذخیره و همچنین مشخص می‌گردد که هر ویژگی متعلق به کدام کلاس است. سپس با استفاده از الگوریتم خوشه‌بندی K-means ویژگی‌های مورد نظر به دست می‌آید. خروجی این مرحله، ماتریسی خواهد بود که در بردارنده مرکز هر خوشه است. در مرحله آزمون، یک تصویر به سیستم وارد شده و ویژگی‌های SIFT آن استخراج می‌شود. سپس فاصله تک‌تک این ویژگی‌ها با مراکز خوشه‌های مرحله آموزش محاسبه می‌شود. اگر این فاصله از مقدار آستانه تعیین‌شده کمتر باشد، این نقطه کلیدی به خوشه نسبت داده شده و در نهایت تعداد ویژگی نسبت داده شده به هر خوشه (هیستوگرام ویژگی‌ها) برای این تصویر به دست می‌آید. در مرحله آخر خوشه‌هایی که بیشترین فراوانی ویژگی را دارند به‌عنوان حاشیه (برچسب) های تصویر انتخاب می‌شوند.



کوه، سنگ، ساختمان، چمن، ابر، پنجره



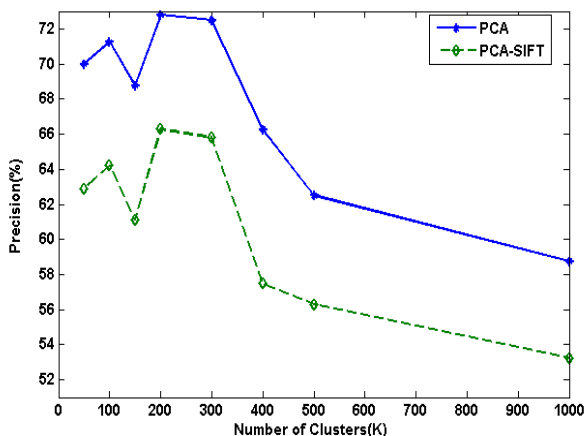
ساختمان، سنگ، ماشین، کوه، جاده، ابر

شکل ۲: مثال‌های از حاشیه‌نویسی تصاویر با استفاده از مدل ارائه شده

الگوریتم پیشنهادی با SIFT و PCA-SIFT برای تعداد خوشه‌های مختلف در شکل ۳ نشان داده شده است.

جدول ۲: مقایسه عملکرد SIFT و PCA-SIFT در الگوریتم پیشنهادی

الگوریتم	زمان آموزش (دقیقه)	میانگین دقت (درصد)
SIFT	۵۱۷	۷۲/۸
PCA-SIFT	۱۳۳	۶۶/۹۵



شکل ۳: میزان دقت به دست آمده به ازای k های مختلف با دو روش

#### PCA-SIFT و SIFT

در ادامه برای مقایسه روش پیشنهادی با روش‌های دیگر از همه تصاویر مجموعه تصاویر Corel5k استفاده شده است. تعداد تصاویر آزمایش و آموزش مانند مجموعه واقعی ۵۰۰ و ۴۵۰۰ در نظر گرفته شده است. جدول ۳ عملکرد الگوریتم پیشنهادی را در مقابل تعدادی از روش‌های دیگر حاشیه‌نویسی نشان می‌دهد. همان‌طور که در جدول مشخص است دقت روش پیشنهادی بسیار بهتر از گذشته بوده است. دلیل به دست آمدن این نتیجه استفاده از ویژگی‌های مناسب‌تر برای توصیف محتوای تصویر است.

#### ۵- نتیجه‌گیری

هدف نهایی سامانه‌های حاشیه‌نویسی خودکار، بازیابی تصاویر حاشیه‌دار بر اساس درخواست‌های متنی است. به همین دلیل ضرورت

جدول ۱: مقایسه میانگین دقت الگوریتم پیشنهادی

الگوریتم	میانگین دقت (درصد)
الگوریتم پیشنهادی	۷۲/۸
الگوریتم WAN [۱۱]	۶۶/۹۵

زمان آموزش برای الگوریتم پیشنهادی برابر ۵۱۷ دقیقه است. طبق نتایج به دست آمده از این الگوریتم، همچنین با توجه به تعداد و ابعاد زیاد ویژگی‌های هر تصویر، اجرا و یادگیری الگوریتم زمان‌بر است. برای مثال، تعداد ویژگی‌ها و نقاط کلیدی به دست آمده برای تصویری در حدود ۳۵۰۰ عدد بوده است. به همین دلیل تصمیم گرفتیم از PCA<sup>۷</sup> برای کاهش ابعاد ویژگی‌ها استفاده کنیم و الگوریتم استخراج ویژگی را به PCA-SIFT تغییر دادیم [۱۷]. PCA-SIFT سعی در بهبود توصیف گره‌های محلی تصویر با استفاده از SIFT دارد. PCA-SIFT بردار ویژگی SIFT را کاهش نمی‌دهد، بلکه ابعاد نقاط مورد علاقه تشخیص داده شده را کاهش می‌دهد. PCA-SIFT فقط از گام تشخیص SIFT استفاده می‌کند، سپس گام توصیف خود را اعمال می‌کند. برای هر نقطه مورد علاقه، PCA-SIFT با تمرکز بر روی نقطه مورد علاقه واقعی، یک تصویر ۴۱×۴۱ به دست می‌آورد؛ گرادیان افقی و عمودی را محاسبه می‌کند و آن‌ها را در یک بردار ویژگی ۳۹×۲۹×۲=۳۰۴۲ ذخیره می‌کند. بنابراین هر بردار ویژگی ۳۰۴۲ تایی بر روی یک فضای با ابعاد کمتر نگاهت می‌شود. این کاهش ابعاد ویژگی سرعت اجرا را افزایش داده، اما ممکن است دقت نتایج را در مقایسه با حالت استفاده از توصیف گره‌های SIFT کاهش دهد. PCA-SIFT وقتی که ابعاد بردار ویژگی ۳۶ بعدی باشد، نتایج بهتری را نشان می‌دهد.

جدول ۲ میزان دقت الگوریتم به ازای ۵ کلمه کلیدی نسبت داده شده به تصاویر آزمایشی را با دو روش SIFT و PCA-SIFT نشان می‌دهد. آزمایش‌های این قسمت بر روی یک سیستم با مشخصات Intel Core 2.3 GHz و حافظه 2.0 GB و ویندوز ۷ نسخه ۳۲ بیتی و نرم‌افزار MATLAB ۷.۱ نسخه R2010a انجام شده است.

همان‌طور که واضح است در روش PCA-SIFT با حذف تعدادی از ویژگی‌ها (نقاط کلیدی) دقت الگوریتم کاهش یافته است. اما زمان یادگیری کاهش چشم‌گیری داشته است. همچنین آنالیز حساسیت

[۱۰] علیرضا نوحی، فرزین یغمایی، «حاشیه نویسی تصاویر با استفاده از متن‌کاوی»، کنفرانس داده‌کاوی ایران، دانشگاه تهران، پاییز ۱۳۹۲.

[11] Wan S. "Image annotation using the simple decision tree," *International Conference on Management of e-Commerce and e-Government*, pp. 141-146, 2011.

[12] V. Maihami, and F.Yaghmaee, "A review on the application of structured sparse representation at image annotation," *Artificial Intelligence Review*, pp. 1-18, 2016, In Press.

[13] V.Maihami, and F. Yaghmaee, "Fuzzy Neighbor Voting for Automatic Image Annotation," *Journal of Electrical and Computer Engineering Innovations*, pp. 9-16, 2016.

[14] C.F.Tsai, "Bag-of-words representation in image annotation: A review," *ISRN Artificial Intelligence*, 2012.

[15] M.U. Kim, and K.Yoon, "Performance evaluation of large-scale object recognition system using bag-of-visual words model" *Multimedia Tools and Applications*, vol. 74, no. 7, pp. 2499-2517, 2015.

[16] M.Oszust, "BDSB: Binary descriptor with shared pixel blocks," *Journal of Visual Communication and Image Representation*, 2016, In Press.

[17] G. H. Liu, J. Y. Yang, and Z. Li, "Content-based image retrieval using computational visual attention model," *Pattern Recognition*, vol. 48, no. 8, pp. 2554-2566, 2015.

[18] A.G. Faheema, S. Rakshit, "Feature selection using bag of visual words representation", *IEEE 2<sup>nd</sup> International Advance Computing Conference (IACC)*, pp. 151-156, 2010.

[19] R. Valenzuela, W. Schwartz, H. Pedrini, "Dimensionality Reduction Through PCA over SIFT and SURF Descriptors," *University of Campinas, CIS*, 2012.

[20] D.Nister and H.Stewnius, "Scalable recognition with a vocabulary tree." *In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1261-1268, Washington DC, USA, 2006.

[21] J. Yang, Y. Jiang, A. G. Hauptmann and C. Ngo, "Evaluating Bag-Of-Visual-Words Representation in Scene Classification," *In Proc. of International Workshop on Multimedia Information Retrieval (MIR)*, pp. 197-206, 2007.

[22] J. Jeon, V. Lavrenko, R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models," *In 26<sup>th</sup> annual international ACM SIGIR conference on research and development in information retrieval*, ACM, Toronto, pp 119-126, 2003.

[23] V. Lavrenko, R. Manmatha, J. Jeon, "model for learning the semantics of pictures," *In 16<sup>th</sup> conference on advances in neural information processing systems (NIPS)*, Vancouver, MIT Press, Canada, pp. 8-13, 2003.

[24] A. Yavlinsky, E. Schofield, and S. Ruger, "Automated image annotation using global features and robust nonparametric density estimation," *In Proc. ACM Int. Conf. Image Video Retrieval*, pp. 507-517, 2005.

[25] S. Zhu, X. Tan, "A novel automatic image annotation method based on multi-instance learning," *Procedia Eng.*, vol. 15, pp. 3439-3444, 2011.

[26] N. El-Bendary, T. Kim, A. Hassanien, M. Sami, "Automatic image annotation approach based on optimization of classes scores," *Computing*, 2014.

حاشیه‌نویسی تصاویر در بحث بازیابی تصاویر وجود دارد. این مقاله یک الگوریتم حاشیه‌نویسی خودکار تصاویر با استفاده از واژه‌های بصری ارائه کرده است. با توجه به این‌که تولید سیستم حاشیه‌نویسی خودکار شامل دو هسته اصلی استخراج ویژگی و استفاده از الگوریتم‌های یادگیری ماشین است، لذا به منظور تأمین کارایی سیستم در هر دو بخش روش‌های مختلف بررسی و حالت‌های بهینه انتخاب شده است. ویژگی اصلی سیستم تولیدشده در مقایسه با پژوهش‌های انجام‌شده استفاده از تعداد تصاویر و کلاس تصویری بیشتر و بردارهای ویژگی با طول بسیار کمتر و در نتیجه کارایی بالای سیستم است.

### جدول ۳: مقایسه الگوریتم پیشنهادی در مقایسه با روش‌های دیگر

Models	Precision
CMRM [22]	۰/۱۰
CRM [23]	۰/۱۶
NPDE [24]	۰/۱۸
MIL [25]	۰/۲۰
RFC-PSO [26]	۰/۲۶
Proposed method	۰/۲۹

### مراجع

[1] D. Zhang, Md. Monirul Islam, G. J. Lu "A review on automatic image annotation techniques," *Pattern Recognition*, vol 45, pp. 346-362, 2012.

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

[3] Y. Liu, D. Zhang, G. Lu, and W. Ma, "survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, no. 1, pp. 262-282, 2007.

[4] Y. Han, F. Wu, Q. Tian, Y. Zhuang "Image annotation by input-output structural grouping sparsity," *IEEE Transactions on Image Processing*, vol. 21, no. 6, pp. 3066-3079, 2012.

[5] H. Fu, Z. Chi, D. Feng, "Recognition of attentive objects with a concept association network for image annotation," *Pattern Recognition*, vol 43, no. 10, pp. 3539-3547, 2010.

[6] J. Liu, M. Li, Q. Liu, H. Lu, and S. Ma, "Image annotation via graph learning," *Pattern Recognition*, vol. 42, no. 2, pp. 218-228, Feb. 2009.

[7] Su, Ja-Hwung, et al. "Effective semantic annotation by image-to-concept distribution model," *IEEE Transactions on Multimedia*, vol. 13, no. 3, pp. 530-538, 2011.

[8] S.Zhang, J. Huang, H. Li, and D. N. Metaxas, "Automatic image annotation and retrieval using group sparsity," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 3, pp. 838-849, 2012.

[9] Y. Yang, Z. Huang, Y. Yang, J. Liu, H. T. Shen, and J. Luo, "Local image tagging via graph regularized joint group sparsity," *Pattern Recognition*, vol.46, no.5, pp.1358-1368, 2013.

### زیرنویس‌ها

<sup>3</sup> Scale-invariant feature transform

<sup>4</sup> Bag of Features

<sup>5</sup> Words

<sup>6</sup> Codeword

<sup>7</sup> Principal Components Analysis

<sup>1</sup> Bag-of-visual-words

<sup>2</sup> Codebook