



وزارت علوم، تحقیقات و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران

پژوهشکده مدیریت فناوری اطلاعات

رساله دکتری در رشته مهندسی فناوری اطلاعات

# **طراحی سامانه‌ای برای بازیابی تصاویر با استفاده از روش‌های هوشمند برای کاربرد در جویشرهای تحت وب**

نکارنده:

**داور گیوکی**

**استاد راهنما: دکتر غلامعلی منتظر**

**دی ۱۳۹۵**

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



وزارت علوم، تحقیقات و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران

پژوهشکده مدیریت فناوری اطلاعات

رساله دکتری در رشته مهندسی فناوری اطلاعات

# **طراحی سامانه‌ای برای بازیابی تصاویر با استفاده از روش‌های هوشمند برای کاربرد در جویشرهای تحت وب**

نکارنده:

**داور گیوکی**

**استاد راهنما: دکتر غلامعلی منتظر**

**دی ۱۳۹۵**



وزارت علوم، تحقیقات و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران

## صور تجلسه دفاع از رساله دكترای تخصصی

با تأییدات خداوند متعال جلسه دفاع از رساله دكترای تخصصی آقای..... در رشته  
(گرایش).....تحت عنوان .....

که در تاریخ / / در جلسه شورای پژوهشی به تأیید رسیده است، با حضور اساتید راهنما، مشاور و هیأت  
داوران در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) در تاریخ / / تشکیل گردید. در این جلسه،  
رساله فوق با موفقیت مورد دفاع قرار گرفت و نامبرده موفق به دریافت نمره ..... (به  
حروف.....) و درجه .....گردید.

- |                                 |                    |       |       |
|---------------------------------|--------------------|-------|-------|
| • استاد راهنما:                 | نام و نام خانوادگی | درجه: | امضا: |
|                                 | مؤسسه:             |       |       |
| • استاد مشاور:                  | نام و نام خانوادگی | درجه: | امضا: |
|                                 | مؤسسه:             |       |       |
| • هیأت داوران:                  | نام و نام خانوادگی | درجه: | امضا: |
|                                 | مؤسسه:             |       |       |
|                                 | نام و نام خانوادگی | درجه: | امضا: |
|                                 | مؤسسه:             |       |       |
|                                 | نام و نام خانوادگی | درجه: | امضا: |
|                                 | مؤسسه:             |       |       |
| • نماینده معاون پژوهش پژوهشگاه: | نام و نام خانوادگی |       | امضا: |



## تعهدنامه اصالت اثر

اینجانب **داور گیوکی** متعهد می‌شوم که مطالب مندرج در این پایان نامه/ رساله حاصل کار پژوهشی اینجانب تحت نظارت و راهنمایی اساتید راهنما و مشاور مشخص شده بوده و به دستاوردهای دیگران که در این پژوهش از آنها استفاده شده است مطابق مقررات و روال متعارف ارجاع و در فهرست منابع و مآخذ ذکر گردیده است. همچنین این پایان نامه/ رساله قبلاً برای احراز هیچ مدرک هم‌سطح یا بالاتر ارائه نگردیده است.

کلیه حقوق مادی نشر و تکثیر پایان نامه/ رساله و درآمدهای حاصل از آنها متعلق به پژوهشگاه علوم و فناوری اطلاعات ایران می باشد، ولی حقوق معنوی پدید آورندگان محفوظ خواهد بود. هرگونه استفاده مستقیم از نتایج علمی و عملی، واگذاری اطلاعات به دیگران یا چاپ و تکثیر، نسخه‌برداری، ترجمه و اقتباس از این رساله بدون موافقت کتبی پژوهشگاه علوم و فناوری اطلاعات ایران ممنوع است. نقل مطالب با ذکر مآخذ بلامانع است.

**داور گیوکی**

امضا

## تقديم

سپاسگزاری

## چکیده

گسترش ارتباطات و فناوری اطلاعات تمامی جوانب زندگی انسان‌ها را تحت تأثیر قرار داده است. بدون تردید در دنیایی که به درستی «دنیای اطلاعات» نام گرفته است یافتن اطلاعات و داده‌های مشابه می‌تواند در بسیاری از زمینه‌ها راهگشا باشد. امروزه مجموعه‌های عظیمی از تصاویر مختلف بر روی وب در دسترس است که چالش‌های فنی بسیاری را برای انجام پردازش‌های گوناگون به وجود آورده است. جست‌وجوی تصاویر دیجیتال در پایگاه داده‌های بزرگ یکی از مهم‌ترین این چالش‌ها است که در دهه‌های اخیر توجه بسیاری از محققان علوم کامپیوتر را به خود معطوف کرده است.

در این رساله، با تمرکز بر روش بسته‌واژگان، روش جدیدی برای بازیابی معنایی تصاویر تکنیک جدیدی برای استخراج ویژگی‌های تصویر، معرفی می‌شود. مقایسه نتایج روش پیشنهادی با سایر روش‌ها، کارایی روش پیشنهادی را در بازیابی تصاویر نشان می‌دهد.

در ادامه با تعمیم رویکرد پیشنهادی، قابلیت آن برای حل مسئله بازیابی معنایی تصاویر ارزیابی شده است. همچنین برای افزایش بیشتر کارایی روش بسته‌واژگان، دسته‌بند جدیدی مبتنی بر شبکه‌های عصبی معرفی شده است. در این مرحله، روش‌های پیشنهادی با تعداد زیادی مجموعه داده معیار ارزیابی شده و با استفاده از معیارهای آماری با روش‌های مرسوم مقایسه شده است. تحلیل نتایج آماری، برتری روش‌های پیشنهادی را در مقایسه با روش‌های مرسوم نشان می‌دهد. در نهایت الگوریتم‌های پیشنهادی در قالب سامانه‌ای که قابلیت‌هایی چون کنش‌گرا بودن و تحت وب بودن را دارا است، به کار گرفته شده‌اند. نتایج آزمایش‌های انجام شده نشان می‌دهد که هر یک از رهیافت‌های پیشنهادی توانسته‌اند کارایی بازیابی محتوای محور تصاویر را در مقایسه با سایر روش‌های به روز، به میزان مناسبی افزایش دهند.

**کلید واژه‌ها:** بازیابی معنایی تصاویر، سیفت، بسته‌واژگان، الگوی دودویی محلی، دسته‌بند RBF.

## فهرست مطالب

| عنوان                                              | شماره صفحه |
|----------------------------------------------------|------------|
| فهرست مطالب.....                                   | أ          |
| فهرست جدول‌ها.....                                 | د          |
| فهرست شکل‌ها.....                                  | و          |
| <b>فصل اول کلیات پژوهش.....</b>                    | <b>۱</b>   |
| ۱-۱. مقدمه.....                                    | ۲          |
| ۱-۲. تعریف مسئله تحقیق.....                        | ۳          |
| ۱-۳. اهمیت موضوع تحقیق.....                        | ۴          |
| ۱-۴. نوآوری تحقیق.....                             | ۵          |
| ۱-۵. فرضیه‌های تحقیق.....                          | ۶          |
| ۱-۶. سؤال‌های اصلی تحقیق.....                      | ۶          |
| ۱-۷. روش اجرای پژوهش.....                          | ۶          |
| ۱-۸. واژه‌های کلیدی تحقیق.....                     | ۸          |
| ۱-۹. ساختار رساله.....                             | ۸          |
| <b>فصل دوم مبانی پژوهش.....</b>                    | <b>۱۱</b>  |
| ۲-۱. مقدمه.....                                    | ۱۱         |
| ۲-۲. بازنمایی تصویر.....                           | ۱۲         |
| ۲-۲-۱. مدل‌های بازنمایی سطح پایین تصویر.....       | ۱۲         |
| ۲-۲-۱-۱. توصیف‌گرهای رنگ.....                      | ۱۳         |
| ۲-۲-۱-۲. توصیف‌گرهای بافت.....                     | ۱۵         |
| ۲-۲-۱-۳. توصیف‌گرهای شکل.....                      | ۱۸         |
| ۲-۳. نتیجه‌گیری.....                               | ۲۱         |
| <b>فصل سوم بررسی روش‌های مؤثر توصیف تصویر.....</b> | <b>۲۴</b>  |

|           |                                                                   |     |
|-----------|-------------------------------------------------------------------|-----|
| ۳-۱       | مقدمه                                                             | ۲۵  |
| ۳-۲       | بازنمایی معنای تصاویر                                             | ۲۵  |
| ۳-۳       | استخراج ویژگی‌های تصویر                                           | ۲۸  |
| ۳-۳-۱     | تشخیص‌گر نقطه مورد علاقه                                          | ۲۹  |
| ۳-۳-۲     | توصیف‌گر ویژگی                                                    | ۳۰  |
| ۳-۳-۲-۱   | تشخیص‌گر نقطه کلیدی سیفت                                          | ۳۰  |
| ۳-۳-۲-۲   | بررسی عملکرد روش سیفت                                             | ۳۴  |
| ۳-۳-۲-۳   | توصیف‌گر LBP                                                      | ۴۶  |
| ۳-۳-۲-۴   | توصیف‌گر HOG                                                      | ۴۸  |
| ۳-۴       | روش بسته‌واژگان تصویری                                            | ۴۹  |
| ۳-۵       | مسائل موجود در بازیابی تصاویر                                     | ۵۳  |
| ۳-۶       | نتیجه‌گیری                                                        | ۵۸  |
| <b>۵۸</b> | <b>فصل چهارم طراحی سامانه بازیابی تصاویر</b>                      |     |
| ۴-۱       | مقدمه                                                             | ۶۰  |
| ۴-۲       | مدل پیشنهادی برای بازیابی تصاویر                                  | ۶۱  |
| ۴-۳       | روش پیشنهادی برای استخراج ویژگی‌های تصاویر                        | ۶۲  |
| ۴-۳-۱     | روش‌های پیشنهادی برای بازیابی تصاویر با استفاده از ویژگی‌های سیفت | ۶۳  |
| ۴-۴       | تعمیم روش پیشنهادی                                                | ۷۶  |
| ۴-۴-۱     | بهبود توصیف‌گر سیفت فشرده                                         | ۷۶  |
| ۴-۴-۲     | استخراج اطلاعات مکانی                                             | ۷۸  |
| ۴-۵       | دسته‌بند پیشنهادی                                                 | ۸۱  |
| ۴-۵-۱     | تعیین وزن‌های لایه نهانی                                          | ۸۲  |
| ۴-۵-۲     | الگوریتم بهینه‌سازی PSO                                           | ۸۲  |
| ۴-۶       | سامانه طراحی شده                                                  | ۸۸  |
| ۴-۷       | نتیجه‌گیری                                                        | ۹۵  |
| <b>۹۷</b> | <b>فصل پنجم جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده</b>     |     |
| ۵-۱       | مقدمه                                                             | ۹۷  |
| ۵-۲       | معیارهای ارزیابی در حوزه بازیابی تصاویر                           | ۹۷  |
| ۵-۳       | نتایج ارزیابی روش‌های پیشنهادی روی پایگاه داده‌های معیار          | ۹۸  |
| ۵-۳-۱     | ارزیابی روش سیفت فشرده                                            | ۹۸  |
| ۵-۳-۲     | ارزیابی توصیف‌گر سیفت فشرده بهبود یافته                           | ۱۰۱ |
| ۵-۳-۳     | ارزیابی روش پیشنهادی برای ضبط اطلاعات مکانی                       | ۱۰۴ |

- ۴-۳-۵. ارزیابی iPSO-OSD در روش بسته واژگان ..... ۱۰۸
- ۴-۵. محدودیت‌های پژوهش ..... ۱۱۴
- ۵-۵. پیشنهادهایی برای تحقیقات آتی ..... ۱۱۵
- ۶-۵. نتیجه‌گیری ..... ۱۱۷
- فهرست منابع ..... ۱۱۹

## فهرست جدول‌ها

| عنوان                                                                                             | شماره صفحه |
|---------------------------------------------------------------------------------------------------|------------|
| جدول ۱-۱ مراحل انجام پژوهش و روش‌های مورد استفاده.....                                            | ۸          |
| جدول ۱-۲ نمونه‌های مختلف روش‌های رنگ و دسته‌بندی آن‌ها.....                                       | ۱۴         |
| جدول ۲-۲ مزایا و معایب روش‌های استخراج ویژگی مبتنی بر رنگ.....                                    | ۱۵         |
| جدول ۳-۲ نمونه‌های مختلف روش‌های بافت و دسته‌بندی آن‌ها.....                                      | ۱۷         |
| جدول ۳-۲. نمونه‌های مختلف روش‌های بافت و دسته‌بندی آن‌ها.....                                     | ۱۷         |
| جدول ۴-۲ مزایا و معایب روش‌های استخراج ویژگی مبتنی بر بافت.....                                   | ۱۸         |
| جدول ۵-۲ روش‌های بازیابی تصاویر با استفاده از دسته‌بندی آن‌ها.....                                | ۱۹         |
| جدول ۵-۲ روش‌های بازیابی تصاویر با استفاده از دسته‌بندی آن‌ها <b>Error! Bookmark not defined.</b> |            |
| جدول ۵-۲ روش‌های بازیابی تصاویر با استفاده از دسته‌بندی آن‌ها <b>Error! Bookmark not defined.</b> |            |
| جدول ۱-۳ نتایج اجرای الگوریتم سیفت برای بررسی انطباق بین یک تصویر و ۵۸ تصویر دیگر.....            | ۴۰         |
| جدول ۲-۳ مسائل موجود در روش بسته‌واژگان.....                                                      | ۵۶         |
| جدول ۱-۴ نتایج محاسبه بردارهای گرادیان در جهت‌های مختلف.....                                      | ۷۷         |
| جدول ۱-۵ جدول ارزیابی.....                                                                        | ۹۷         |
| جدول ۲-۵ متوسط دقت حاصل از روش سیفت فشرده در مقایسه با سایر روش‌ها.....                           | ۱۰۰        |
| جدول ۳-۵ نتایج محاسبه بردارهای گرادیان در جهت‌های مختلف.....                                      | ۱۰۱        |
| جدول ۴-۵ نتایج محاسبه بردارهای گرادیان در جهت‌های مختلف.....                                      | ۱۰۳        |
| جدول ۵-۵ شرح پایگاه داده OT.....                                                                  | ۱۰۴        |
| جدول ۶-۵ نتایج حاصل از اجرای روش پیشنهادی با پارامترهای مختلف.....                                | ۱۰۶        |
| جدول ۷-۵ مقایسه نتایج حاصل از روش پیشنهادی با سایر روش‌ها.....                                    | ۱۰۷        |



- جدول ۵-۸ شرح پایگاه داده FP ..... ۱۰۹
- جدول ۵-۹ اطلاعات مربوط به پایگاه داده LSP ..... ۱۱۰
- جدول ۵-۱۰ مقایسه نتایج حاصل از روش پیشنهادی با سایر روش‌ها ..... ۱۱۴

## فهرست شکل‌ها

| عنوان                                                                          | شماره صفحه |
|--------------------------------------------------------------------------------|------------|
| شکل ۱-۱ نمونه‌ای از سامانه‌بازيابی تصوير.....                                  | ۴          |
| شکل ۲-۱ مراحل انجام پژوهش.....                                                 | ۷          |
| شکل ۱-۲ مثال‌هایی از بازنمایی‌های سراسری و محلی تصاویر.....                    | ۱۳         |
| شکل ۲-۲ نمونه‌هایی از بافت‌های مختلف.....                                      | ۱۶         |
| شکل ۳-۲ دسته‌بندی روش‌های مبتنی بر بافت.....                                   | ۱۶         |
| شکل ۴-۲ تقسیم‌بندی کلی توصیف‌گرهای شکل.....                                    | ۱۹         |
| شکل ۱-۳ سیستم پیشنهادی توسط وگل و شيله.....                                    | ۲۷         |
| شکل ۲-۳ توابع گوسی مختلف و تفاضل آن‌ها.....                                    | ۳۱         |
| شکل ۳-۳ بیشینه و کمینه تصوير تفاضل گوسی با مقایسه یک پیکسل.....                | ۳۲         |
| شکل ۴-۳ تصوير اصلی و نقاطا کسترمم انتخاب شده.....                              | ۳۲         |
| شکل ۵-۳ تصويری با ۷۲۹ نقطه کلیدی (سمت چپ) تصويری با ۵۳۶ نقطه کلیدی.....        | ۳۲         |
| شکل ۶-۳ تشکیل هیستوگرام جهت به منظور تعیین جهت چیره.....                       | ۳۳         |
| شکل ۷-۳ تشکیل توصیف‌گر نقطه کلیدی.....                                         | ۳۳         |
| شکل ۸-۳ نتایج ارزیابی تأثیر بهبود تصوير ۱ بر روش سیفت.....                     | ۳۷         |
| شکل ۹-۳ نتایج ارزیابی تأثیر بهبود تصوير ۲ بر روش سیفت.....                     | ۳۸         |
| شکل ۱۰-۳ مدت زمان تولید نقاط کلیدی برای تصاویر با تعداد نقاط کلیدی متفاوت..... | ۳۹         |
| شکل ۱۱-۳ رابطه بین سرعت اجرای الگوریتم سیفت و وضوح تصاویر.....                 | ۳۹         |
| شکل ۱۲-۳ تطابق دو تصوير و تصاویر دستکاری شده دو تصوير اصلی.....                | ۴۲         |
| شکل ۱۳-۳ تطابق دو تصوير و تصاویر دستکاری شده دو تصوير اصلی.....                | ۴۴         |
| شکل ۱۴-۳ تصوير سمت راست با ۱۰۹۹ و تصوير سمت چپ با ۱۰۸۳ نقطه کلیدی.....         | ۴۴         |
| شکل ۱۵-۳ یافتن تنها یک نقطه مشابه به وسیله روش سیفت.....                       | ۴۵         |
| شکل ۱۶-۳ دو تصوير مشابه به ترتیب از راست به چپ با ۳۹۵ و ۸۴۶ نقطه کلیدی.....    | ۴۵         |

- شکل ۳-۱۷ عدم یافتن هیچ دو نقطه مشابهی ..... ۴۵
- شکل ۳-۱۸ نمونه‌ای از تصویر چهره به همراه تصویر LBP متناظر با آن و بردار ویژگی استخراج شده ..... ۴۶
- شکل ۳-۱۹ تعیین وزن‌های همسایه‌ها با الگوی تولید شده ..... ۴۷
- شکل ۳-۲۰ الگوی تولید شده توسط توصیف‌گر LBP ..... ۴۷
- شکل ۳-۲۱ تصاویر بافتی مختلف و تکستون‌های مربوط به آن‌ها ..... ۵۰
- شکل ۳-۲۲ نمایش بسته واژگان تصویری برای سه تصویر بافتی ..... ۵۱
- شکل ۳-۲۳ تصاویر واژگان مربوط به گیتار، دوچرخه و صورت انسان ..... ۵۱
- شکل ۳-۲۴ دایره واژگان ساخته شده توسط تصاویر گیتار، دوچرخه و صورت ..... ۵۱
- شکل ۳-۲۵ مراحل تصویری پیاده‌سازی روش بسته‌واژگان تصویری ..... ۵۳
- شکل ۳-۲۶ تأثیر تغییرات روشنایی الف ..... ۵۴
- شکل ۳-۲۷ چالش تنوع درون دسته‌ای ..... ۵۵
- شکل ۳-۲۸ تنوع برون دسته‌ای ..... ۵۶
- شکل ۴-۱ نمای کلی روش پیشنهادی برای بازیابی تصاویر ..... ۶۲
- شکل ۴-۲ تشابه‌یابی با استفاده از روش اول ..... ۶۴
- شکل ۴-۳ تشابه‌یابی با استفاده از روش دوم ..... ۶۵
- شکل ۴-۴ تشابه‌یابی با استفاده از روش سوم ..... ۶۷
- شکل ۴-۵ تشابه‌یابی با استفاده از روش چهارم ..... ۶۸
- شکل ۴-۶ تشابه‌یابی با استفاده از روش پنجم ..... ۶۹
- شکل ۴-۷ تشابه‌یابی با استفاده از روش ششم ..... ۷۰
- شکل ۴-۸ تشابه‌یابی با استفاده از روش پنجم و تعداد خوشه‌ها برابر ۳۲ ..... ۷۱
- شکل ۴-۹ تشابه‌یابی با استفاده از روش پنجم و تعداد خوشه‌ها برابر ۶۴ ..... ۷۲
- شکل ۴-۱۰ تشابه‌یابی با استفاده از روش پنجم و تعداد خوشه‌ها برابر ۱۲۸ ..... ۷۳
- شکل ۴-۱۱ تشابه‌یابی با استفاده از روش پنجم و تقسیم تصویر به ۴ قسمت ..... ۷۵
- شکل ۴-۱۲ تشابه‌یابی با استفاده از روش پنجم و تقسیم تصویر به ۱۶ قسمت ..... ۷۶
- شکل ۴-۱۳ بردارهای گرادین محاسبه شده در هشت جهت در نقطه‌ای دلخواه از تصویر ..... ۷۷
- شکل ۴-۱۴ تصویری دلخواه و تقسیم آن به چهار بلوک ..... ۷۹
- شکل ۴-۱۵ بلوک‌بندی  $4 \times 4$  ..... ۷۹
- شکل ۴-۱۶ بلوک‌بندی  $8 \times 8$  ..... ۸۰
- شکل ۴-۱۷ تصویر بلوک‌بندی شده  $I$  و تداخل پنجره‌ها ..... ۸۰
- شکل ۴-۱۸ شبه کد ضبط اطلاعات مکانی ..... ۸۱
- شکل ۴-۱۹ شبکه عصبی RBF برای دسته‌بندی الگوها ..... ۸۱

- شکل ۴-۲۰ شبه‌کد روش محاسبه مراکز توابع گوسی با استفاده از الگوریتم iPSO ..... ۸۷
- شکل ۴-۲۱ اجزای اصلی تشکیل دهنده موتور جویشر طراحی شده ..... ۸۸
- شکل ۴-۲۲ نمای کلی سامانه بازیابی تصاویر مشابه ..... ۹۰
- شکل ۴-۲۳ تصاویر مشابه با تصویر دشت ..... ۹۱
- شکل ۴-۲۴ نتایج جست‌وجوی تصویر جنگل با شعاع دسته‌بندی برابر با ۱ ..... ۹۲
- شکل ۴-۲۵ بهبود نتایج جست‌وجوی تصویر جنگل با تغییر شعاع دسته‌بندی ۱ به ۳ ..... ۹۳
- شکل ۴-۲۶ قابلیت کنشگرا بودن سامانه طراحی شده ..... ۹۴
- شکل ۵-۱ تعدادی از تصاویر پایگاه داده ونگ ..... ۹۹
- شکل ۵-۲ ماتریس درهم‌ریختگی برای روش سیفت فشرده بهبود یافته با طول بردار ویژگی ..... ۱۴۰۰
- شکل ۵-۳ ماتریس درهم‌ریختگی سیفت فشرده بهبود یافته با طول بردار ویژگی ۳۲۰۰ ..... ۱۰۴
- شکل ۵-۴ برخی تصاویر پایگاه داده OT ..... ۱۰۵
- شکل ۵-۵ بازیابی تصاویر با استفاده از iPSO-OSD ..... ۱۰۸
- شکل ۵-۶ برخی تصاویر پایگاه داده FP ..... ۱۱۰
- شکل ۵-۷ ماتریس درهم‌ریختگی روش پیشنهادی با دسته‌بند iPSO-OSD روی پایگاه داده OT ..... ۱۱۱
- شکل ۵-۸ ماتریس درهم‌ریختگی برای روش پیشنهادی با دسته‌بند iPSO-OSD روی پایگاه داده FP ..... ۱۱۲
- شکل ۵-۹ ماتریس درهم‌ریختگی برای روش پیشنهادی با دسته‌بند iPSO-OSD روی پایگاه داده LSP ..... ۱۱۳
- شکل ۵-۱۰ تصاویری از حیواناتی که بسیار به هم شبیه هستند ..... ۱۱۵

## فصل اول

### کلیات پژوهش

## ۱-۱. مقدمه

گسترش ارتباطات و فناوری اطلاعات تمامی جوانب زندگی انسان‌ها را تحت تأثیر قرار داده است. بدون تردید در دنیایی که به درستی «دنیای اطلاعات» نام گرفته است یافتن اطلاعات و داده‌های مشابه می‌تواند در بسیاری از زمینه‌ها راهگشا باشد. همه روزه هزاران تصویر تولید می‌شوند که برای دسته‌بندی، سازماندهی و دستیابی به آن‌ها نیاز به روش‌های جدید و مؤثر پردازش تصویر است. با رشد نمایی تصاویر با کیفیت‌های بالا، دسته‌بندی معنایی تصاویر برای پردازش‌هایی نظیر جست‌وجوی تصاویر مشابه به شدت احساس می‌شود. امروزه دسته‌بندی و بازیابی تصاویر با در نظر گرفتن دسته آن‌ها امری مهم و پرکاربرد تلقی می‌شود.

برای یافتن تصاویر مشابه ابتدا ویژگی‌هایی از تصاویر استخراج می‌شود و سپس با تعیین میزان شباهت این ویژگی‌ها به یکدیگر، شبیه‌ترین تصاویر مشخص می‌شوند. از طرفی ویژگی‌هایی که برای تعیین میزان شباهت تصاویر استفاده می‌شوند، ویژگی‌های سطح پایین هستند. برای بازیابی معنایی تصاویر معمولاً از روش‌های دسته‌بندی استفاده می‌شود. به هر میزان که بردارهای ویژگی استخراج شده متمایزتر باشند، دسته‌بندی تصاویر از کارایی بالاتری برخوردار خواهد شد. همچنین هرچقدر دسته‌بندی کننده مورد استفاده قوی‌تر باشد بهتر می‌تواند به کاهش فاصله معنایی و یافتن دسته تصاویر مشابه کمک کند.

هدف این تحقیق ارائه روش جدیدی برای بهبود روش‌های بازیابی تصاویر است. در این فصل کلیات تحقیق و مواردی از قبیل تعریف مسئله تحقیق، اهمیت تحقیق، نوآوری تحقیق، فرضیه‌های اصلی تحقیق و تعریف واژگان اصلی تحقیق آورده شده است.

## ۲-۱. تعریف مسئله تحقیق

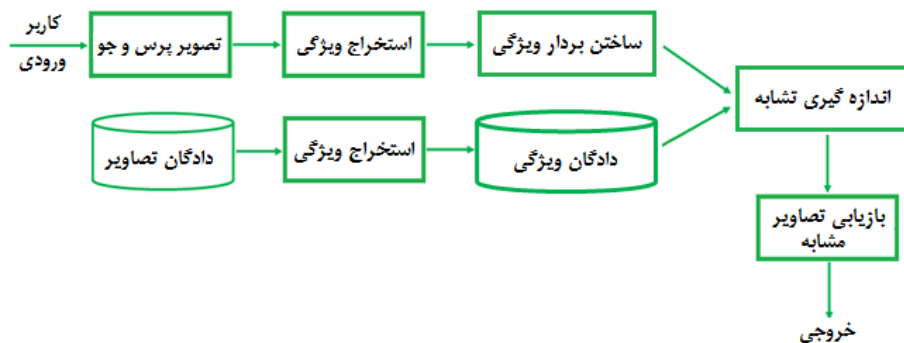
امروزه مجموعه‌های عظیمی از تصاویر مختلف در وب دسترس است که چالش‌های فنی بسیاری را برای انجام پردازش‌های گوناگون به وجود آورده است. جست‌وجوی تصاویر دیجیتال یکی از مهم‌ترین چالش‌ها است که در دهه‌های اخیر توجه بسیاری از محققان علوم کامپیوتر را به خود معطوف کرده است. برای جست‌وجوی تصاویر در یک پایگاه داده دو روش کلی وجود دارد:

۱- جست‌وجو براساس ابرداده‌های متنی<sup>۱</sup>، ۲- جست‌وجو براساس محتوا<sup>۲</sup>. در روش اول نیروی انسانی به هر تصویر برجستگی، که بیانگر محتوای آن است، نسبت می‌دهد و سپس جست‌وجو مبتنی بر این کلید واژه‌ها و با استفاده از روش‌های کلاسیک پایگاه داده و پرس‌وجو انجام می‌شود (Chan and Hsu 1992). در روش‌های جدید برای جست‌وجوی تصاویر مشابه در یک پایگاه داده، الگوریتم‌های مختلف پردازش تصویر به منظور استخراج بردارهای ویژگی (که حاوی اطلاعات گوناگونی از محتوای تصاویر است) به کار می‌روند (Deselaers, Keysers and Ney 2008).

شکل ۱-۱ سامانه‌ای را برای بازیابی تصویر مبتنی بر محتوا نشان می‌دهد. ابتدا با توجه به محتوای تصاویر پایگاه داده، الگوریتم‌های مناسب برای استخراج ویژگی‌های تصاویر موجود در پایگاه داده انتخاب می‌شوند و واحد «استخراج ویژگی»، با استفاده از این الگوریتم‌ها، ویژگی‌های مختلف هر تصویر را استخراج و بردارهای ویژگی تصاویر را در پایگاه داده جدیدی به نام «دادگان ویژگی» قرار می‌دهد. سپس کاربر تصویری را به عنوان ورودی (تصویر پرس‌وجو)، وارد می‌کند و مجدداً واحد استخراج ویژگی، بردار ویژگی را برای این تصویر استخراج می‌کند. بردار ویژگی تصویر پرس‌وجو به واحد «اندازه‌گیری تشابه» می‌رود. در واحد اندازه‌گیری تشابه میزان تشابه میان بردارهای دادگان ویژگی و بردار ویژگی تصویر پرس‌وجو اندازه‌گیری شده و تصاویر با توجه به میزان فاصله‌شان از تصویر پرس‌وجو رتبه‌بندی می‌شوند. سرانجام شبیه‌ترین تصاویر به عنوان خروجی توسط واحد «بازیابی تصاویر مشابه» به کاربر نشان داده می‌شوند.

<sup>۱</sup>Textual Meta Data

<sup>۲</sup>Content



شکل ۱-۱: نمونه‌ای از سامانه بازیابی تصویر (Traina et al 2002)

بنابراین در هر سامانه بازیابی تصویر دو واحد اصلی وجود دارد: «واحد استخراج ویژگی» و «واحد اندازه‌گیری تشابه». روش‌های مختلف بسیاری برای یافتن تصاویر مشابه معرفی شده‌اند که هر یک از ویژگی‌های مختلف تصویر مانند «رنگ»<sup>۱</sup>، «بافت»<sup>۲</sup> و «شکل»<sup>۳</sup> استفاده می‌کنند اما مهم‌ترین چالش‌های موجود در این راه عبارتند از شناسایی و استخراج ویژگی‌های مهم تصویر (که به طور کارایی بیانگر اطلاعات مهم تصویر است) و نیز دستیابی به معیاری کارا برای اندازه‌گیری میزان تشابه معنایی تصاویر با یکدیگر.

مسئله اصلی این رساله، بهبود روش‌های مؤثر استخراج ویژگی‌های تصویر، بازیابی تصاویر با توجه دسته‌بندی معنایی آن‌ها و طراحی و پیاده‌سازی سامانه‌ای برای بازیابی تصاویر است. این سامانه، تصویری را از کاربر دریافت و تصاویر مشابه با آن را به عنوان خروجی بازیابی می‌کند.

### ۳-۱. اهمیت موضوع تحقیق

با پیشرفت روز افزون فناوری دیجیتال و تولد دوربین‌های پیشرفته شاهد ظهور حجم وسیعی از تصاویر گوناگون هستیم. یکی از مهم‌ترین دغدغه‌های پژوهشگران، انجام پردازش‌های مؤثر و یافتن تصاویر مشابه بین تعداد زیادی تصویر است. بدون هیچ تردیدی در دنیای امروز تصاویر مهم‌ترین عناصر اطلاعاتی به شمار می‌روند (Kooper 2016; پیروزمند ۱۳۹۱) و بر اهمیت تصاویر در اینترنت هر روز افزوده می‌شود (Clark 2012). در حوزه بازیابی معنایی تصاویر با استفاده از دسته‌بندی آن‌ها، همچنان تحقیقات گسترده‌ای در حال انجام است (Patheja, Wao and Maurya 2012).

<sup>۱</sup>color

<sup>۲</sup>texture

<sup>۳</sup>shape



آمار نشان می‌دهد روزانه ۳۰۰ میلیون تصویر در فیس‌بوک بارگذاری می‌شود. در هر دقیقه ۵۱۰ کامنت ارسال می‌شود و ۲۹۳,۰۰۰ کاربر صفحه شخصی خود را روزآمد می‌کنند (Kooper 2016).

در این رساله، الگوریتمی کارا برای بهبود بازیابی معنایی تصاویر پیشنهاد می‌شود. الگوریتم نوآوری‌های مختلفی دارد که در ادامه بیان خواهند شد.

## ۴-۱. نوآوری تحقیق

نوآوری‌های این رساله را می‌توان به سه بخش متفاوت که در ادامه توضیح داده خواهند شد، تقسیم کرد. در هر بخش نتایج آزمایش‌ها نشان می‌دهد دستاوردهای حاصل شده در مقایسه با سایر روش‌های موجود موجب بهبود نتایج شده است. لازم به ذکر است تمامی آزمایش‌ها روی پایگاه داده‌های معیاری که در مقالات معتبر علمی از آن‌ها استفاده شده‌است، انجام شده‌اند.

### الف. دسته‌بند

یکی از مهم‌ترین عواملی که کارایی بازیابی معنایی تصاویر مشابه در دسته‌های مختلف را تحت تأثیر قرار می‌دهد، دسته‌بندی است که از آن استفاده می‌شود. در این رساله یک دسته‌بند جدید مبتنی بر شبکه‌های عصبی مصنوعی معرفی شده‌است. دسته‌بند پیشنهادی از خانواده شبکه‌های عصبی «پایه شعاعی»<sup>۱</sup> است. برای تعیین مراکز و پهنای توابع گوسی لایه نهانی این شبکه روش‌های جدیدی پیشنهاد شده است که موجب افزایش کارایی شبکه عصبی پیشنهادی در مقایسه با دیگر دسته‌بندها مانند «ماشین بردار پشتیبان»<sup>۲</sup> است.

### ب. توصیفگر سیفت<sup>۳</sup> فشرده

یکی از موفق‌ترین الگوریتم‌های بازشناسی اشیا و یافتن انطباق‌ها میان تصاویر مختلف، الگوریتم سیفت است. در این رساله نقاط ضعف و قوت الگوریتم سیفت و امکان استفاده از آن در بازیابی تصاویر مشابه به دقت بررسی شد. با توجه به انجام آزمایش‌ها مختلف روش جدیدی برای فشرده‌کردن ویژگی‌های استخراج شده توسط تشخیص‌گر نقاط کلیدی سیفت برای بازیابی تصاویر پیشنهاد شده‌است.

### ج. ضبط اطلاعات مکانی<sup>۴</sup>

---

<sup>۱</sup> Radial Basic Function

<sup>۲</sup> Support Vector Machine

<sup>۳</sup> Scale Invariant Feature Transform

<sup>۴</sup> Spatial information

بدون تردید روش بسته‌ واژگان تصویری<sup>۱</sup> یکی از مؤثرترین و موفق‌ترین روش‌های بازیابی تصاویر است. برخلاف کارایی بسیار مطلوب، این روش معایبی نیز دارد. بزرگ‌ترین آن از بین رفتن اطلاعات مکانی است. در این رساله با در نظر گرفتن تصاویر در وضوح مختلف و با انجام تقسیم‌بندی‌های خاص روش جدیدی برای ضبط اطلاعات مکانی پیشنهاد شده‌است. آزمایش‌ها انجام شده نشان می‌دهد که روش مذکور در مقایسه با همتهای خود از کارایی بالاتری برخوردار است.

## ۵-۱. فرضیه‌های تحقیق

فرضیه‌های اصلی تحقیق به شرح زیر است:

الف- کارایی روش‌های بازیابی تصاویر را می‌توان با استفاده از دسته‌بندی‌های قوی‌تر بهبود داد.

ب- بهبود روش‌های استخراج ویژگی از تصاویر می‌تواند باعث افزایش کارایی بازیابی تصاویر شود.

## ۶-۱. سؤال‌های اصلی تحقیق

۱- مؤثرترین ویژگی‌ها برای بازیابی تصاویر مشابه کدامند؟

۲- آیا استفاده از ترکیبی از ویژگی‌ها باعث افزایش کارایی یک سامانه بازیابی تصاویر می‌شود؟

۳- آیا استفاده از روش‌های هوشمند و دسته‌بندی باعث افزایش کارایی یک سامانه بازیابی تصاویر می‌شود؟

## ۷-۱. روش اجرای پژوهش

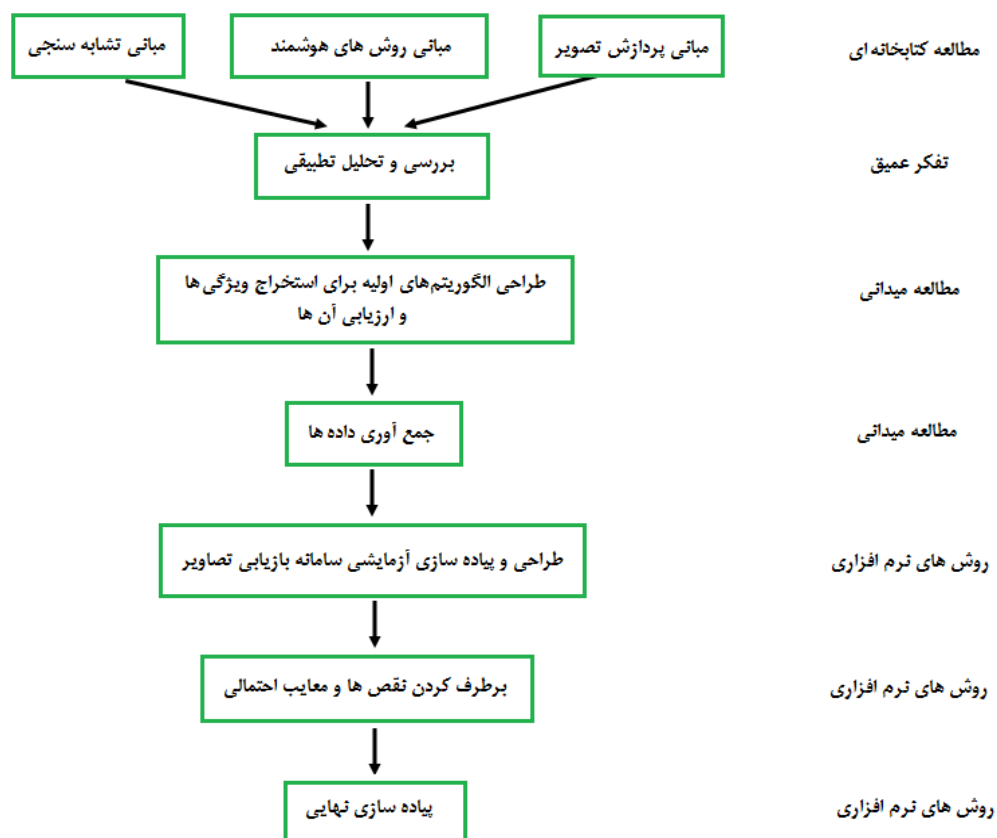
روش پژوهش شامل مجموعه‌ای از مراحل و گام‌هاست که پژوهشگر برای انجام پژوهش از طراحی و تحقیق تا ارائه گزارش و نتایج آن انجام می‌دهد (Kothari 2004). برخی از عوامل اثرگذار در انتخاب روش پژوهش عبارتند از: مسئله پژوهش، اهداف و چارچوب‌های آن، وضعیت پژوهش‌های گذشته، نوع پژوهش و پرسش‌های آن، ابزار پژوهش و کمی یا کیفی بودن پژوهش.

برای حل یک مسئله ابتدا می‌باید درک درستی از خود مسئله و زوایای آن داشت. لذا در گام اول مبانی تشابه‌یابی تصویری، انواع تصاویر، پردازش تصویر و نیز روش‌های هوشمند بررسی شده‌است. پس از فراگیری مبانی اولیه می‌باید راه حل‌های موجود در مورد آن مسئله بررسی شود، لذا در گام بعد به مطالعه و مرور روش‌های بازیابی تصاویر پرداخته شد. با تحلیل روش‌های کنونی و بررسی نقاط ضعف و

---

<sup>۱</sup> Bag of Visual Words (BoVW)

قوت این روش‌ها ضمن ایجاد زمینه فکری برای ارائه راه حل‌های جدید با زوایای پنهان صورت مسئله آشنایی بیشتری پیدا می‌شود. با جمع‌بندی اطلاعات به دست آمده در این مرحله و رسیدن به سطح مطلوبی در زمینه استخراج ویژگی‌ها از تصاویر، مشابهت‌سنجی و روش‌های مختلف دسته‌بندی و خوشه‌بندی می‌توان طراحی الگوریتم‌های اولیه را برای بازیابی تصاویر آغاز کرد. در این مرحله داده‌های مورد نیاز برای طراحی سامانه جمع‌آوری می‌شوند و سامانه مورد نظر به صورت آزمایشی پیاده‌سازی می‌شود و معایب احتمالی آن برطرف می‌شوند و سپس نسخه نهایی توسعه داده می‌شود. شکل ۱-۲ مراحل انجام پژوهش و ابزار مورد استفاده در آن را نشان می‌دهد.



شکل ۱-۲: مراحل انجام پژوهش

با توجه به موارد فوق، جدول ۱-۱ پرسش‌های پژوهش، هدف هر پرسش، شیوه پاسخ‌گویی به هر پرسش و نحوه ارزیابی خروجی آن را نشان می‌دهد.

## جدول ۱-۱: مراحل انجام پژوهش و روش‌های مورد استفاده

| پرسش                                                                                          | هدف                                                                                     | روش پاسخ‌گویی                  | خروجی                                                                             | نحوه ارزیابی                                                                   |
|-----------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|--------------------------------|-----------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| مؤثرترین ویژگی‌ها برای بازیابی تصاویر مشابه کدامند؟                                           | شناسایی، انتخاب و بهبود کاراترین توصیف‌گرهای تصویر برای بازیابی محتوا محور              | مطالعه کتابخانه‌ای و تفکر عمیق | انتخاب و بهبود توصیف‌گرهای مناسب برای بازیابی تصاویر                              | مقایسه نتایج حاصل از بهبود روش‌های استخراج ویژگی با دیگر روش‌ها در حوزه تخصصی  |
| آیا استفاده از ترکیبی از ویژگی‌ها باعث افزایش کارایی یک سامانه بازیابی تصاویر می‌شود؟         | بررسی امکان افزایش کارایی روش پیشنهادی با استفاده از ترکیبی از ویژگی‌ها                 | مطالعه کتابخانه‌ای و تفکر عمیق | تصمیم به ترکیب یا عدم ترکیب ویژگی‌های استخراج شده برای افزایش کارایی روش پیشنهادی | مقایسه نتایج حاصل از ترکیب ویژگی‌ها با دیگر روش‌ها در حوزه تخصصی               |
| آیا استفاده از روش‌های هوشمند و دسته‌بندی باعث افزایش کارایی یک سامانه بازیابی تصاویر می‌شود؟ | بررسی امکان افزایش کارایی روش پیشنهادی با استفاده از دسته‌بندی‌هایی مانند شبکه‌های عصبی | مطالعه کتابخانه‌ای و تفکر عمیق | معرفی یک دسته‌بند با کارایی بالا برای بازیابی محتوا محور تصاویر                   | مقایسه نتایج حاصل از استفاده از دسته‌بند پیشنهادی با دیگر روش‌ها در حوزه تخصصی |

## ۸-۱. واژه‌های کلیدی تحقیق

**الف. بازیابی محتوا محور تصاویر :** یافتن تصاویر مشابه با استفاده از محتوای آن‌ها را بازیابی محتوا محور تصاویر گویند. منظور از محتوای تصاویر ویژگی‌هایی از تصویر است که از آن‌ها استخراج می‌شود (Traina et al 2002).

**ب. دسته‌بندی :** منظور انتساب دسته جدید به یکی از دسته‌های از پیش تعریف شده در مسئله است (Vogel and Schiele 2007).

**ج. خوشه‌بندی :** خوشه‌بندی نوعی فرایند یادگیری بدون نظارت است که به دنبال یافتن رابطه یا شباهت فاصله‌ای بین نمونه‌های موجود و در نهایت گروه‌بندی آنهاست (Gutierrez-Osuna 2002).

**د. بسته واژگان/ویژگی‌های تصویری :** بسته واژگان/ویژگی‌های تصویری مدلی است که یک تصویر را به صورت مجموعه‌ای از واژگان خاص، که این واژگان روی یک مجموعه واژه تعریف شده‌اند، توصیف می‌کند (Eitz et al 2011).

## ۹-۱. ساختار رساله

ساختار رساله به شرح ذیل است : در فصل دوم، مروری اجمالی بر روش‌های بازیابی تصویر، مزایا و معایب هر یک از روش‌ها بیان می‌شود. در فصل سوم، روش‌های کاربردی در رساله توضیح داده می‌شوند.

در فصل چهارم، سامانه پیشنهادی برای حل مسئله رساله طراحی می‌شود. فصل پنجم به نتیجه‌گیری و ارائه پیشنهاد برای پژوهش‌های آتی اختصاص یافته است.

## فصل دوم

### مبانی پژوهش

## ۱-۲. مقدمه

حجم عظیم تصاویر که با توجه به پیشرفت سریع فناوری ایجاد شده است دشواری‌های بسیاری را برای کاربران مختلف فراهم کرده است. روش‌های قدیم بازیابی تصاویر دارای معایبی نظیر سلیقه‌ای بودن، وقت‌گیر بودن و عدم امکان توصیف بسیاری از تصاویر با واژگان است (Chan and Hsu 1992). بنابراین دانشمندان به فکر یافتن راه‌های مؤثرتر برای یافتن تصاویر مشابه افتادند. آن‌ها تلاش می‌کنند جست‌وجویی مؤثر و مبتنی بر محتوای تصویر (رنگ، بافت، شکل و...) انجام دهند. یکی از مزایای مهم بازیابی تصویر مبتنی بر محتوا، قابلیت اجرا به صورت خودکار است. از روش بازیابی مبتنی بر محتوا در موارد گوناگونی چون تشخیص چهره، تشخیص اثرانگشت، کتابخانه‌های دیجیتالی، پایگاه‌های اطلاعاتی نیروی پلیس، پایگاه داده‌های پزشکی، تحقیقات تاریخی و... استفاده می‌شود (Ramana, Rao and Prasad 2016).

در این فصل کارهایی که در حوزه بازیابی محتوا محور تصاویر انجام شده‌اند، مرور می‌شوند. این پژوهش‌ها مبتنی بر استخراج ویژگی‌های مختلف تصاویر نظیر رنگ، بافت و شکل هستند. در ادامه روش‌های نوین بازیابی تصاویر با استفاده از توصیف‌گرهایی نظیر سیفت<sup>۱</sup>، الگوی دودویی محلی<sup>۲</sup> و هیستوگرام گرادیان‌های جهت‌دار<sup>۳</sup> مرور می‌شوند. در ادامه بیان می‌شود که چگونه استفاده از روش‌های دسته‌بندی تصاویر باعث کاهش فاصله معنایی در بازیابی محتوا محور تصاویر می‌شود.

---

<sup>۱</sup> Scale Invariant Feature Transform (SIFT)

<sup>۲</sup> Local Binary Pattern (LBP)

<sup>۳</sup> Histogram of Oriented Gradients (HOG)

## ۲-۲. بازنمایی تصویر

روش‌های مختلفی برای بازنمایی تصاویر پیشنهاد شده است. در یک تقسیم بندی کلی می‌توان انواع روش‌های بازنمایی تصاویر را به سه دسته تقسیم کرد :

- الف) روش‌هایی که از ویژگی‌های سطح پایین تصاویر استفاده می‌کنند.
  - ب) روش‌هایی که از یک بازنمایی معنایی سراسری از تصاویر بهره می‌جویند.
  - ج) روش‌هایی که از قطعات محلی تصویر برای بازنمایی معنایی تصویر استفاده می‌کنند.
- در ادامه به اهداف هر یک از این روش‌ها و ویژگی‌های آن‌ها پرداخته می‌شود.

### ۱-۲-۲. مدل‌های بازنمایی سطح پایین تصویر

اولین روش‌هایی که برای بازیابی محتوای محور تصاویر معرفی شدند از ویژگی‌های سطح پایین تصاویر بهره می‌بردند (Stricker and Orengo 1995; Bimbo 1999). اگر تصاویر در دسته‌های معنایی قرار داشته باشند این ویژگی‌ها توسط یک دسته‌بند آموخته می‌شوند و پس از آموزش دسته‌بند می‌توان از آن برای یافتن دسته تصاویر شبیه به تصویر پرس‌وجو استفاده کرد. اگر تصاویر در دسته‌های معنایی قرار نداشته باشند تمامی بردارهای ویژگی استخراج شده از تصاویر موجود در پایگاه داده با بردار ویژگی استخراج شده از تصویر پرس‌وجو مقایسه می‌شوند و سپس شبیه‌ترین تصاویر به تصویر پرس‌وجو به عنوان خروجی بازگردانده می‌شوند.

روش‌های مبتنی بر ویژگی‌های سطح پایین بر این فرض استوار هستند که محتوای یک تصویر را می‌توان با ویژگی‌های سطح پایین آن توصیف کرد.

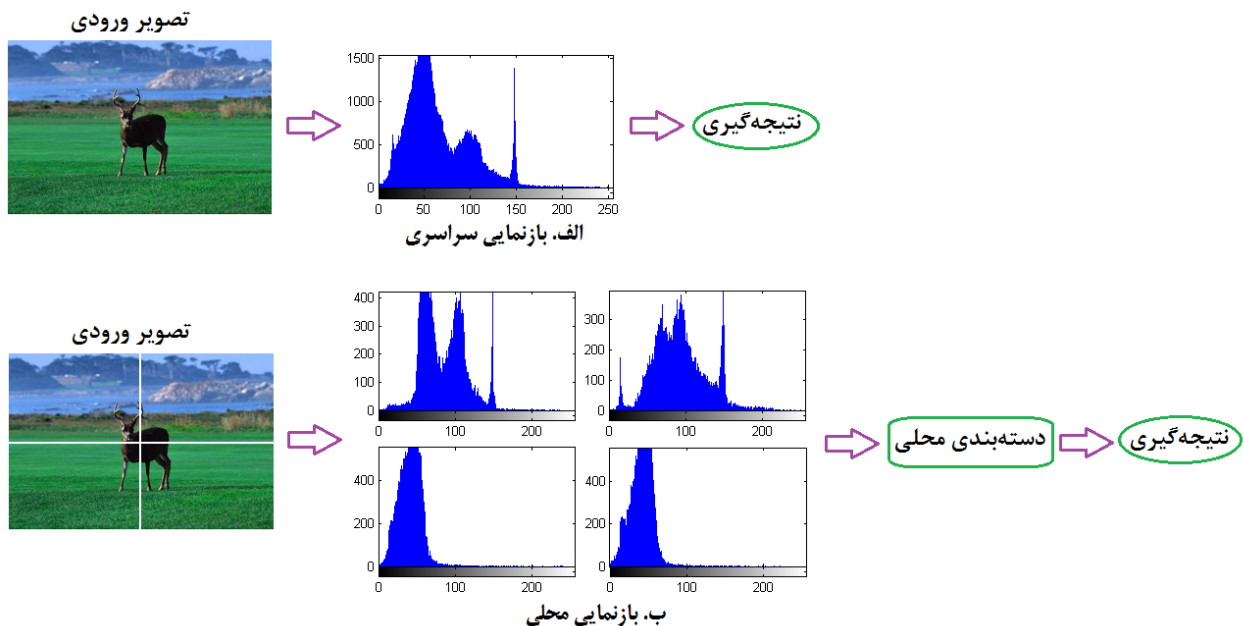
این رهیافت‌ها را می‌توان به دو دسته اصلی تقسیم کرد: بازنمایی سراسری<sup>۱</sup> و بازنمایی محلی<sup>۲</sup>. در بازنمایی سراسری ویژگی‌های سطح پایین تصاویر از کل تصویر استخراج می‌شوند اما در بازنمایی محلی ابتدا تصویر به قطعات یا بخش‌هایی تقسیم می‌شود و سپس ویژگی‌ها از هر یک از این بخش‌ها استخراج می‌شوند. در گام بعد ویژگی‌های استخراج شده با هم ترکیب می‌شوند و بردار ویژگی نهایی را می‌سازند. شکل ۱-۲ نمونه‌ای از ویژگی‌های سطح پایین تصاویر را که به صورت سراسری و محلی استخراج شده‌اند، نشان می‌دهد.

---

<sup>۱</sup> Global Representation

<sup>۲</sup> Local Representation





شکل ۲-۱: مثال‌هایی از بازنمایی‌های سراسری و محلی تصاویر

چنانچه در شکل (۲-۱-الف) ملاحظه می‌شود در بازنمایی سراسری ویژگی هیستوگرام رنگ تصویر از همه تصاویر استخراج می‌شود. در شکل (۲-۱-ب) برای استخراج ویژگی‌های محلی سطح پایین تصویر، ابتدا تصویر به بلوک‌هایی تقسیم می‌شود و سپس ویژگی هیستوگرام رنگ از این بلوک‌ها استخراج می‌شود.

روش‌هایی که از ویژگی‌های سطح پایین تصویر برای توصیف آن استفاده می‌کنند از توصیف‌گرهای رنگ، بافت و شکل استفاده می‌کنند.

#### ۲-۲-۱-۱. توصیف‌گرهای رنگ

رنگ از محبوب‌ترین ویژگی‌هایی است که برای توصیف محتوای تصویر استفاده می‌شود. می‌توان توصیف‌گرهای رنگ را به سه دسته زیر تقسیم کرد (Penatti, Valle and Torres 2012):

**الف. توصیف‌گرهای مبتنی بر روش‌های سراسری:** این توصیف‌گرها اطلاعات رنگ تصویر را به طور سراسری در نظر می‌گیرند و از روش‌های پیش‌پردازشی چون بخش‌بندی و بلاک‌بندی تصویر استفاده نمی‌کنند، این روش‌ها سریع هستند اما راجع به چیدمان پیکسل‌ها در تصویر اطلاعاتی به دست نمی‌دهند.

**ب. توصیف‌گرهای مبتنی بر ناحیه‌های با اندازه ثابت:** در این روش تصویر به سلول‌هایی با اندازه ثابت تقسیم می‌شود و در نتیجه اطلاعات بیشتری در بازه چیدمان پیکسل‌ها در تصویر در مقایسه با دسته

روش‌های اول مدنظر قرار می‌گیرد.

ج. توصیف‌گرهای مبتنی بر بخش‌بندی تصویر: در این روش هر تصویر با استفاده از الگوریتم‌های مختلف خوشه‌بندی یا بخش‌بندی به ناحیه‌هایی با اندازه‌های متفاوت تقسیم می‌شود و سپس ویژگی‌ها از این نواحی استخراج می‌شوند.

در جدول ۱-۲ برخی از روش‌های بازیابی تصاویر با استفاده از ویژگی رنگ آن‌ها ملاحظه می‌شوند.

جدول ۱-۲: نمونه‌های مختلف روش‌های رنگ و دسته‌بندی آن‌ها

| روش                  | نام روش      | توضیح مختصر                                                                                                        | مرجع                                          |
|----------------------|--------------|--------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|
| سراسری               | CW-LUV       | استخراج ضرایب موجک از هیستوگرام رنگ در فضای رنگی CIELuv                                                            | (Nallaperumal, Banu and Christiyana 2007)     |
| سراسری               | CW-HSV       | استخراج ضرایب موجک از هیستوگرام رنگ در فضای رنگی HSV                                                               | (Utenpattananant, Chitsobhuk and Khawne 2006) |
| سراسری               | Chaira Fuzzy | استفاده از هیستوگرام رنگ و تشابه‌سنج فازی مبتنی بر توزیع گاما                                                      | (Chaira and Ray 2005)                         |
| سراسری               | WC           | کرلوگرام رنگ مبتنی بر ضرایب موجک                                                                                   | (Moghaddam et al 2005)                        |
| سراسری               | SCD          | استفاده از تبدیل هار بر هیستوگرام رنگ                                                                              | (Manjunath et al 2001)                        |
| سراسری               | ACC          | اتوکرولوگرام رنگ                                                                                                   | (Huang et al 1997)                            |
| سراسری               | CGCH         | هیستوگرام تجمعی رنگ                                                                                                | (Stricker and Orengo 1995)                    |
| سراسری               | GCH          | هیستوگرام رنگ                                                                                                      | (Swain and Ballard 1991)                      |
| نواحی با اندازه ثابت | Color Bitmap | ضبط اطلاعات سراسری و محلی تصویر با مقایسه متوسط رنگ بلوک‌های تصویر و متوسط رنگ کل تصویر                            | (Lu and Chang 2007)                           |
| نواحی با اندازه ثابت | CDE          | ضبط اطلاعات مکانی رنگ با استفاده از انتروپی                                                                        | (Sun, Zhang and Zhou 2006)                    |
| مبتنی بر بخش‌بندی    | DCDI         | خوشه‌بندی رنگ‌ها                                                                                                   | (Yang et al 2008)                             |
| مبتنی بر بخش‌بندی    | CBC          | استفاده از خوشه‌بندی با کمک رنگ                                                                                    | (Stehling Nascimento, and Falcão 2001)        |
| مبتنی بر بخش‌بندی    | CBRD         | بخش‌بندی تصویر، استخراج ویژگی‌های رنگ، بافت و شکل از هر ناحیه و ترکیب این ویژگی‌ها برای ساختن بردار ویژگی نهایی    | (Shrivastava and Tyagi 2015)                  |
| سراسری               | Color Ton    | استخراج ویژگی‌های رنگ با استفاده از ماتریس هم‌رخدادی و ترکیب آن‌ها با ویژگی‌های بافت استخراج شده به کمک تبدیل موجک | (Rahimi and Moghaddam 2015)                   |
| سراسری               | CDH          | استخراج ویژگی‌های رنگ با در نظر گرفتن موقعیت مکانی پیکسل‌ها در فضای Lab                                            | (Liu and Yang 2013)                           |
| سراسری               | CDH-ART      | استخراج ویژگی‌های هیستوگرام رنگ و ترکیب آن‌ها با ویژگی‌های شکل                                                     | (Wilia and Pal 2014)                          |
| سراسری               | HSVHIST      | استخراج ویژگی رنگ در فضای HSV و ترکیب آن با ویژگی‌های لبه                                                          | (Patil and Kiumar 2016)                       |
| محلی                 | LED          | تشخیص نقاط کلیدی تصویر، تقسیم بندی تصویر و استخراج ویژگی‌های رنگ و بافت و ترکیب آن‌ها با یکدیگر                    | (Pham et al 2016)                             |

با مرور روش‌های پیشنهادی در جدول ۲-۱ می‌توان تحقیقات در زمینه بازیابی تصاویر با استفاده از ویژگی رنگ را می‌توان در یکی از سه گروه زیر جای داد :

۱- استفاده از فضای رنگی مناسب بر اساس هدفی که از قبل برای کاربردی خاص تعیین شده‌است.

۲- معرفی الگوریتم‌های مناسب برای استخراج ویژگی رنگ.

۳- مطالعه و ارزیابی تشابه‌سنج‌های مختلف مبتنی بر رنگ.

همچنین می‌توان مزایا و معایب هر یک از روش‌های بازیابی تصاویر با استفاده از ویژگی رنگ را به صورت جدول ۲-۲ خلاصه کرد.

**جدول ۲-۲: مزایا و معایب روش‌های استخراج ویژگی مبتنی بر رنگ**

| معایب                                                                               | مزایا                                                                          | روش                  |
|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|----------------------|
| عدم در نظر گرفتن موقعیت مکانی پیکسل‌ها                                              | سرعت بالا<br>بردار ویژگی نسبتاً کوچک<br>مقاوم در برابر نویز<br>نامتغیر با چرخش | سراسری               |
| حساس به نویز<br>کندتر در مقایسه با روش‌های سراسری<br>بردارهای ویژگی با ابعاد بزرگتر | در نظر گرفتن موقعیت مکانی پیکسل‌ها                                             | نواحی با اندازه ثابت |
| پیچیدگی بالا (به دلیل استفاده از الگوریتم‌های<br>بخش‌بندی یا خوشه‌بندی)             | کارایی بالا به دلیل در نظر گرفتن موقعیت مکانی<br>پیکسل‌ها                      | مبتنی بر بخش‌بندی    |

## ۲-۲-۱-۲. توصیف گرهای بافت<sup>۱</sup>

برای بافت نمی‌توان تعریف دقیقی ارائه کرد. بافت درکی است که بیانگر ساختار یک سطح است و با نگاه به آن می‌توان دریافت که الگوهایی به نحوی در حال تکرار شدن هستند. می‌توان بافت‌ها را از حالت کاملاً «اتفاقی»<sup>۲</sup> تا «ساخت‌یافته»<sup>۳</sup> مرتب کرد. بافت‌های اتفاقی شبیه به نویز هستند درحالی‌که بافت‌های ساخت‌یافته از الگوهایی منظم تشکیل شده‌اند (Zujovic, Pappas and Neuhoff 2013). شکل ۲-۲ نمونه‌ای از بافت‌های گوناگون را نشان می‌دهد.

<sup>۱</sup>Texture descriptors

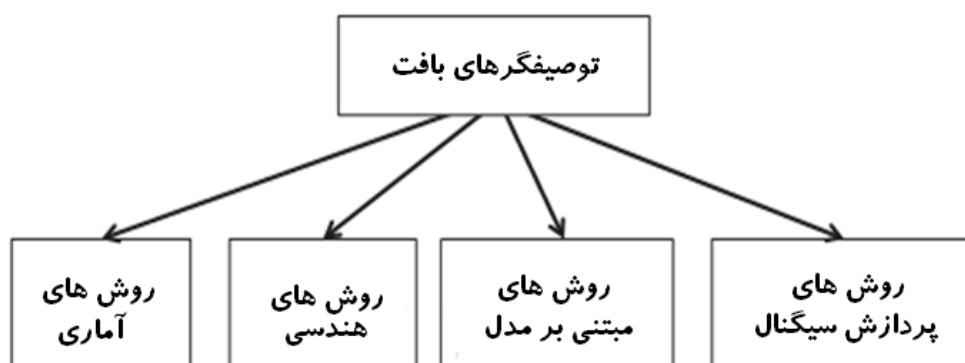
<sup>۲</sup>Stochastic

<sup>۳</sup>Structural



شکل ۲-۲: نمونه‌هایی از بافت: های مختلف (Wikipedia 2013)

روش‌های گوناگونی برای استخراج ویژگی‌های بافت ارائه شده است. همانطور که در شکل ۲-۳ ملاحظه می‌شود توصیف‌گرهای بافت به چهار دسته کلی تقسیم می‌شوند (Penatti, Valle and Torres 2012):



شکل ۲-۳: دسته بندی روش‌های مبتنی بر بافت (Penatti, Valle and Torres 2012)

**الف. روش‌های آماری:** این روش‌ها سنتی‌ترین روش‌ها هستند و خود به سه دسته روش‌های آماری مرتبه اول، دوم و سوم تقسیم می‌شوند که بهترین این روش‌ها ماتریس هم‌رخدادی<sup>۱</sup> است (Verdoolaege and Scheundres 2011).

**ب. روش‌های هندسی:** این روش‌ها ویژگی‌های هندسی «عناصر بافت» مانند اندازه، شکل، ناحیه و طول را تحلیل می‌کند (Nevatia 1982). این روش‌ها برای تحلیل بافت‌های تصاویر طبیعت مانند دریاها یا ابرها اصلاً مناسب نیستند، در عوض برای تصاویری مانند تصاویر دیوار یا تصاویر شبکه‌ای بسیار خوب عمل می‌کنند زیرا عناصر سازنده آن‌ها از شکل‌های یکسان و با اندازه‌های برابر ساخته شده‌اند (Torres et al. 2013).

<sup>۱</sup> Cooccurrence matrix

**ج. روش‌های مبتنی بر مدل:** این روش‌ها بر مدل‌هایی بنا شده‌اند که ساختمان یک بافت را توصیف می‌کنند. پارامترهای استخراج شده توسط این مدل‌ها اطلاعات لازم برای توصیف یک بافت را ضبط می‌کنند. برای مثال می‌توان عناصر یک بافت را به صورت نقاط روشن یا تاریک، الگویی عمودی یا افقی، گوشه‌ها، خطوط و ... مدل کرد. این توصیف‌گرها برای بافت‌های منظم بسیار خوب عمل می‌کنند اما برای بازیابی تصاویر زیاد مورد استفاده قرار نمی‌گیرند (Oja, Pietikainen and Maenpaa 2002).

**د. روش‌های پردازش سیگنال:** این روش‌ها مبتنی بر اعمال فیلترهایی (در حوزه‌های زمان و مکان) روی تصویر هستند. توصیف‌گرهای مبتنی بر تبدیل موجک و فیلترهای گابور از این دست روش‌ها هستند. این روش‌ها از کارایی بسیار بالایی برخوردار هستند (Randen and Husoy 1999).

جدول ۲-۳ برخی روش‌های مبتنی بر بافت و تقسیم‌بندی آن‌ها را مشاهده کرد.

**جدول ۲-۳: نمونه‌های مختلف روش‌های بافت و دسته‌بندی آن‌ها**

| دسته                 | نام روش  | توضیح مختصر                                                                                       | مرجع                                        |
|----------------------|----------|---------------------------------------------------------------------------------------------------|---------------------------------------------|
| آماري                | 2DWAH    | روشی مبتنی بر هیستوگرام لبه در فضای مقیاس                                                         | (Kiranyaz, Ferreira and Gabbouj 2008)       |
| آماري                | MRH      | استفاده از مجموعه‌ای از هیستوگرام‌ها که در فضای مقیاس ساخته می‌شوند.                              | (Hadjidemetriou, Grossberg, and Nayar 2004) |
| آماري                | MOGG     | استفاده از ضرایب موجک و مخلوط‌های متناهی از چگالی‌های گوسی توسعه یافته (MOGG) <sup>۱</sup>        | (Allili 2012)                               |
| آماري                | EHD      | تقسیم تصویر به بلوک‌هایی و استفاده از هیستوگرام لبه در جهت‌های مختلف به کمک تشخیص‌دهنده‌های لبه   | (Manjunath 2001)                            |
| پردازش سیگنال        | IFLT     | استفاده از تبدیل موجک و استخراج ویژگی‌های نامتغیر نسبت به چرخش که از شدت نور تصویر به دست می‌آیند | (Janney and Yu 2007)                        |
| پردازش سیگنال        | HanGabor | استفاده از فیلترهای گابور و استخراج ویژگی‌هایی که نسبت به چرخش و تغییر مقیاس نامتغیر هستند        | (Han and Ma 2007)                           |
| پردازش سیگنال        | SimDCT   | استفاده از تبدیل کسینوسی گسسته                                                                    | (Sim, Kim and Park 2001)                    |
| پردازش سیگنال        | HTD      | استفاده از فیلترهای گابور در جهت‌ها و مقیاس‌های مختلف                                             | (Wu et al 2000)                             |
| پردازش سیگنال        | FCM-RCCR | استفاده از ویژگی‌های استخراج شده توسط تبدیل موجک و شبکه‌های عصبی برای بازیابی تصاویر              | (Dash, Mukhopadhyay and Gupta 2015)         |
| آماري- پردازش سیگنال | LBP-CLCM | ترکیب ویژگی‌های آماری و پردازش سیگنال برای بازیابی تصاویر چهره                                    | (Kumar and Stephena 2016)                   |
| پردازش سیگنال        | LBP-CCH  | ترکیب ویژگی‌های رنگ و پردازش سیگنال                                                               | (Jain, Zaveri and Patel 2016)               |
| پردازش سیگنال        | LBP-BEM  | ترکیب ویژگی‌های بافت، شکل و رنگ                                                                   | (Shrivastava and Tyag 2015)                 |
| پردازش سیگنال        | GC-Gabor | استخراج ویژگی‌های موجک گابور در فضاهای رنگی مختلف                                                 | (Li, Huang and Zhu 2017)                    |
| پردازش سیگنال        | DLTP     | استفاده از ویژگی‌های بافت که به وسیله الگوی دودویی محلی استخراج شده‌اند برای بازیابی تصاویر       | (Vipparthi and Nagar 2015)                  |

<sup>۱</sup>Mixtures of Generalized Gaussian (MOGG)

مزایا و معایب هر یک از روش‌های مذکور در جدول ۴-۲ خلاصه شده است.

**جدول ۴-۲: مزایا و معایب روش‌های استخراج ویژگی مبتنی بر بافت**

| روش           | مزایا                                                                                                  | معایب                                                              |
|---------------|--------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|
| هندسی         | مناسب برای تصاویر بافتی با عناصر سازنده قابل تمایز                                                     | پیچیدگی بالا<br>نامناسب برای تصاویر بافتی یکنواخت (طبیعت-دریا-ابر) |
| مبتنی بر مدل  | ضبط اطلاعات مربوط به چیدمان پیکسل‌ها                                                                   | حساس بودن به نویز عدم استفاده در بازیابی تصاویر                    |
| آماري         | سادگی                                                                                                  | سختی در انتخاب پارامترهای مناسب برای مدل                           |
| پردازش سیگنال | کارایی بسیار بالا قابلیت تحلیل چند وضوحی نمایش موثر<br>سیگنال در هر دو فرکانس و مکان ضبط اطلاعات مکانی | محاسبات ریاضی پیچیده‌تر                                            |

### ۳-۱-۲-۲. توصیف‌گرهای شکل

مطالعات نشان می‌دهد که بازنمایی شکل توسط لبه‌ها یکی از مهم‌ترین ویژگی‌ها برای بازیابی، دسته‌بندی و بازشناسی اشیاء است (Yang, Kpalma and Ronsin 2010). در همین راستا برخی از روش‌های استخراج ویژگی‌ها از شکل بیان شده و یک دسته‌بندی کلی از این روش‌ها ارائه می‌شود.

برای استخراج ویژگی‌های شکل از تصویر ابتدا توسط الگوریتم‌های بخش‌بندی، تصویر به ناحیه‌ها یا اشیایی تقسیم می‌شود. سپس با استفاده از توصیف‌گرهای شکل ویژگی‌های هر ناحیه/شیء استخراج می‌شود. توصیف‌گرهای شکل به دو دسته کلی «مبتنی بر مرز»<sup>۱</sup> و «مبتنی بر ناحیه»<sup>۲</sup> تقسیم می‌شوند. این روش‌ها بر استخراج ویژگی‌ها از کانتور و یا کل ناحیه مربوط به کانتور متمرکز هستند. پس با توجه به اینکه آیا کل شیء یا بخش‌هایی از آن نمایش داده می‌شود این دو روش به دو دسته محلی و سراسری تقسیم می‌شوند. تقسیم‌بندی دیگر نیز این است که توصیف‌گرهای شکل به دو دسته مبتنی بر «حوزه مکان»<sup>۳</sup> و «حوزه تبدیل»<sup>۴</sup> تقسیم می‌شوند. این دو روش بر این اصل استوارند که آیا اندازه‌گیری‌ها مستقیماً روی خود شکل انجام می‌گیرند یا تبدیلی بر شکل‌ها اعمال می‌شود (Safar, Shahabi and Sun 2000). شکل ۴-۲ تقسیم‌بندی کلی دیگری از توصیف‌گرهای شکل را نشان می‌دهد.

لازم به ذکر است که هر یک از توصیف‌گرهای شکل مزایا و معایب خاص خود را دارند و با توجه به کاربردی خاص و تصاویری که در دست است برای مسایل گوناگون از آن‌ها استفاده می‌شود.

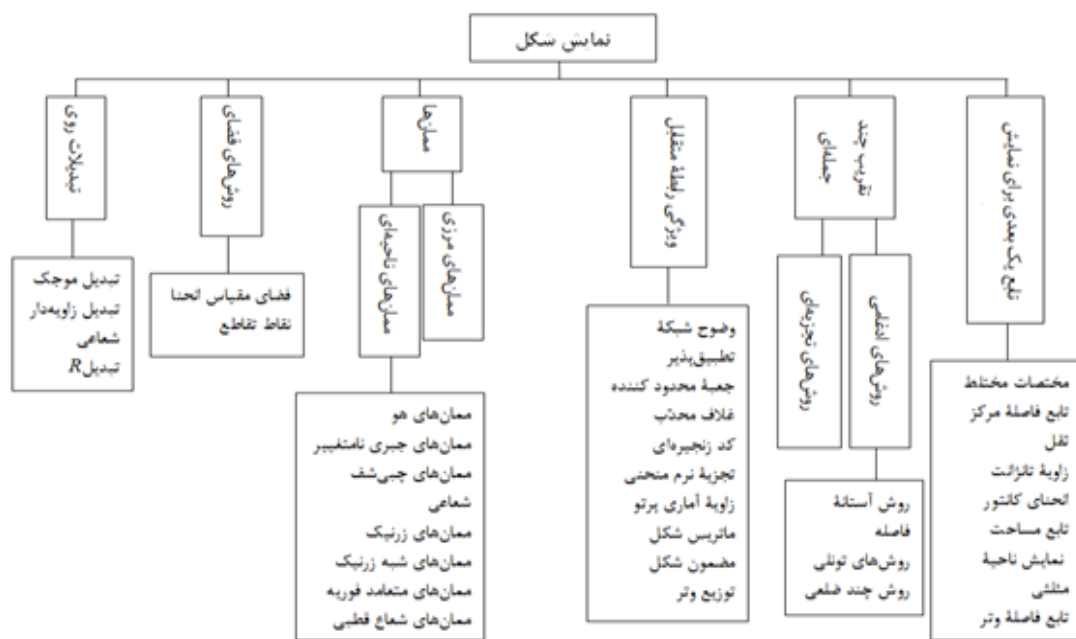
<sup>۱</sup>Boundary based

<sup>۲</sup>Region Based

<sup>۳</sup>Special domain

<sup>۴</sup>Transform Domain

از مهم‌ترین روش‌های استخراج ویژگی از شکل می‌توان به ممان‌های هو<sup>۱</sup>، ممان‌های زرنیک<sup>۲</sup>، فضای حالت انحنا<sup>۳</sup>، پرتو زاویه آماری<sup>۴</sup>، توصیف‌گرهای فوریه<sup>۵</sup>، تبدیل زاویه‌دار شعاعی<sup>۶</sup> و تبدیل  $R$  اشاره کرد.



شکل ۲-۴: تقسیم‌بندی کلی توصیف‌گرهای شکل (Kpalma and Ronsin 2010)

در روش‌های اشاره شده برای بازیابی تصاویر از ویژگی‌های رنگ، بافت و شکل یا ترکیبی از این ویژگی‌ها استفاده شده‌است. بسیاری از محققان برای افزایش کارایی بازیابی تصاویر از دسته‌بندی و خوشه‌بندی تصاویر استفاده کرده‌اند. در این روش‌ها ابتدا ویژگی‌ها از تصاویر استخراج می‌شود و سپس از یک دسته‌بند برای تشخیص دسته تصاویر پرس‌وجو استفاده می‌شود. جدول ۲-۵ تعدادی از این روش‌ها را نشان می‌دهد.

جدول ۲-۵: روش‌های بازیابی تصاویر با استفاده از دسته‌بندی آن‌ها

| دسته‌بند استفاده شده | توضیح مختصر                                                                       | مزایا                 | معایب                                     | مرجع                 |
|----------------------|-----------------------------------------------------------------------------------|-----------------------|-------------------------------------------|----------------------|
| MLP                  | استفاده از ترکیب ویژگی‌های رنگ، بافت و شکل و شبکه عصبی MLP برای تشخیص دسته تصاویر | دقت و سرعت بسیار زیاد | پیچیدگی الگوریتم پیشنهادی برای آموزش شبکه | (Kang and Park 2009) |

<sup>۱</sup>Hu's Moments

<sup>۲</sup>Zernike moments

<sup>۳</sup>Curvature Scale Space (CSS)

<sup>۴</sup>Beam angle statistics (BAS)

<sup>۵</sup>Fourier descriptors

<sup>۶</sup>Angular radial transform (ART)

|     |                                                                                           |                                                                                 |   |                              |
|-----|-------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---|------------------------------|
| MLP | استفاده از ترکیب ویژگی‌های رنگ، بافت و شکل و یک شبکه عصبی MLP جدید برای تشخیص دسته تصاویر | پایداری بالای شبکه طراحی شده در برابر نویز به دلیل طراحی خاص شبکه عصبی پیشنهادی | - | (Olkiewicz and Kaczmar 2010) |
|-----|-------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---|------------------------------|

### ادامه جدول ۲-۵: روش‌های بازیابی تصاویر با استفاده از دسته‌بندی آن‌ها

| دسته‌بند استفاده شده | توضیح مختصر                                                                                                                                                                                                                                                                                                                        | مزایا                                                                                           | معایب                                           | مرجع                               |
|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|-------------------------------------------------|------------------------------------|
| ART                  | استخراج ویژگی‌ها از تصاویر و خوشه‌بندی ویژگی‌های استخراج شده با استفاده از شبکه عصبی ART                                                                                                                                                                                                                                           | دقت مناسب و سرعت بالای روش پیشنهادی به دلیل بازیابی تصاویر تنها از یک خوشه                      | پیچیدگی بالای پیاده‌سازی شبکه ART               | (Park, Seo and Jang 2005)          |
| SOM                  | استخراج ویژگی‌ها از تصاویر و استفاده از شبکه عصبی نگاشت خود سامانده برای خوشه‌بندی تصاویر                                                                                                                                                                                                                                          | دقت خوب و سرعت بالای بازیابی تصاویر به دلیل استفاده از خوشه‌بندی                                | -                                               | (Almeida et al 2008)               |
| MLP                  | استخراج ویژگی‌ها بافت از تصاویر و استفاده از شبکه MLP برای تشخیص دسته تصاویر خوشه‌بندی تصاویر                                                                                                                                                                                                                                      | کوتاه بودن طول بردار ویژگی                                                                      | عدم دقت بالای روش پیشنهادی در تعیین دسته تصاویر | (Garcia, Azuelan and Riveron 2007) |
| LVQ <sup>۱</sup>     | استخراج ویژگی‌ها بافت از تصاویر بافتی و استفاده از شبکه عصبی LVQ برای تعیین دسته تصویر پرس‌وجو. در این روش اگر نتایج بازیابی مورد پسند کاربر واقع نمی‌شد تصاویر مشابه از دومین دسته‌ای که شبیه‌ترین دسته به تصویر پرس‌وجو هستند، بازیابی می‌شوند. این روند تا زمانی که کاربر از خروجی تولید شده توسط شبکه راضی شود، ادامه می‌یابد. | سرعت بالا و تعامل کاربر با سیستم بازیابی تصاویر                                                 | -                                               | (Bashar et al 2005)                |
| SVM                  | استخراج ویژگی‌های شکل و استفاده از دسته‌بند SVM برای تعیین دسته تصویر پرس‌وجو                                                                                                                                                                                                                                                      | پایدار بودن ویژگی پیشنهادی در برابر چرخش و تغییر مقیاس تصویر و تولید بردارهای ویژگی نسبتاً کوچک | -                                               | (Wong and Hsu 2006)                |
| SOM و SVM            | استخراج ویژگی‌های شکل از تصاویر و ترکیب دسته‌بند SVM و خوشه‌بند SOM برای بازیابی تصاویر                                                                                                                                                                                                                                            | دقت و سرعت بالای روش پیشنهادی                                                                   | پیچیدگی زیاد در پیاده‌سازی                      | (Wong, Shih and Liu 2007)          |
| MLP                  | قطعه‌بندی تصویر و تشخیص شیء اصلی تصویر با کمک حذف پس‌زمینه تصویر و سپس استخراج ویژگی‌های بافت از شیء اصلی تصویر و آموزش شبکه عصبی MLP با استفاده از ویژگی‌های استخراج شده                                                                                                                                                          | سرعت و دقت بسیار بالای روش به دلیل استخراج ویژگی‌ها از شیء اصلی درون تصویر                      | نیاز به پیش‌پردازش‌های مختلف                    | (Park, Lee and Kim 2004)           |

<sup>۱</sup> Learning Vector Quantization



## ادامه جدول ۲-۵: روش‌های بازیابی تصاویر با استفاده از دسته‌بندی آن‌ها

|     |                                                                                                                                                                                                                                                          |                                                                                            |                                                                            |                                            |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|--------------------------------------------|
| MLP | استخراج ویژگی‌های بافت و تعیین دسته تصاویر پرس‌وجو با استفاده از دسته‌بند MLP در حالی که کاربر با سیستم بازیابی تصاویر در تعامل است و در صورتی که کاربر از نتایج حاصل شده راضی نباشد وزن‌های شبکه طوری تغییر می‌کنند که نتایج مورد قبول کاربر حاصل شوند. | تعامل با کاربر به منظور افزایش کارایی سیستم بازیابی تصاویر                                 | کند بودن روش پیشنهادی به دلیل آموزش شبکه مجدد شبکه در صورت عدم رضایت کاربر | (Han et al 2005)                           |
| MLP | استخراج ویژگی‌های رنگ و تعیین دسته تصاویر پرس‌وجو با استفاده از شبکه عصبی MLP که از تابع فعال‌کننده بی‌اسپلین استفاده می‌کند.                                                                                                                            | استفاده از تابع فعال‌کننده بی‌اسپلین برای بهتر مشخص کردن رابطه غیر خطی میان بردارهای ویژگی | -                                                                          | (Sadek et al 2009)                         |
| MLP | استخراج ویژگی‌های رنگ و بافت تصاویر و تعیین دسته تصاویر پرس‌وجو با استفاده از شبکه عصبی MLP                                                                                                                                                              | سرعت و دقت بالا                                                                            | -                                                                          | (Hussain et al 2012)                       |
| MLP | استخراج ویژگی‌های بافت و شکل تصاویر و تعیین دسته تصاویر پرس‌وجو با استفاده از شبکه عصبی MLP                                                                                                                                                              | سرعت و دقت بالا                                                                            | -                                                                          | (Li and Liu 2011)                          |
| -   | استخراج ویژگی‌های لبه از تصاویر پزشکی                                                                                                                                                                                                                    | مناسب برای بازیابی اشیای نویزی و اشیای همپوشانی‌کننده                                      | -                                                                          | (Janan and Brady 2015)                     |
| SVM | استخراج ویژگی‌های شکل تصاویر با استفاده از ART                                                                                                                                                                                                           | سرعت و دقت بالا                                                                            | -                                                                          | (Khemchandani and Saigal 2015)             |
| -   | استفاده از ویژگی‌های هندسی برای بازیابی تصاویر                                                                                                                                                                                                           | سادگی و دقت نسبتاً خوب                                                                     | -                                                                          | (Shanmugavadivu, Sumathy and Vadivel 2016) |
| -   | معرفی یک توصیف‌گر جدید مبتنی بر شکل با استفاده از تبدیل رادون                                                                                                                                                                                            | پایدار در برابر نویز و چرخش                                                                | -                                                                          | (Hasegawa and Tabbone 2014)                |

با مطالعه و مرور روش‌های اخیر پیشنهادی برای بازیابی تصاویر چنین به نظر می‌رسد که استفاده از دسته‌بندهای مختلف مانند MLP و SVM باعث افزایش کارایی بازیابی تصاویر می‌شود. همچنین جای خالی شبکه‌های عصبی RBF برای بازیابی تصاویر به چشم می‌خورد.

## ۳-۲. نتیجه‌گیری

یکی از مهم‌ترین بحث‌های مربوط به بازیابی تصویر طراحی سیستمی تطبیق‌پذیر است. همانطور که در

بخش‌های قبل ملاحظه شد هنوز روشی جامع که بتوان از آن برای بازیابی تصاویری از دسته‌های مختلف با سطوح روشنایی متفاوت، در اندازه‌های گوناگون، استفاده کرد وجود ندارد. روش‌هایی که مبتنی بر ویژگی‌های مرحله پایین تصویر هستند، توسط عوامل گوناگونی چون سطوح روشنایی، اشیای همپوشانی شده، وضوح تصویر و نویز به مشکل برمی‌خورند. چالش بسیار مهم دیگر فاصله معنایی (فاصله بین ویژگی‌های استخراج شده در سطح پیکسل و برداشت و دریافت انسان از تصاویر) است. بعد فضای ویژگی، توجه به طراحی سیستمی خاص برای مسئله‌ای خاص، انتخاب روش‌ها، پارامترها و مقادیر آستانه نیز از چالش‌های موجود هستند. از دیگر نکات مهمی که برای طراحی یک سیستم بازیابی تصویر باید مد نظر قرار گیرد این است که برای رسیدن به کارایی مطلوب می‌بایست از همه ویژگی‌های سطح پایین (رنگ-بافت-شکل) استفاده شود. در این بین باید ویژگی‌ها به گونه‌ای انتخاب شوند که مکمل یکدیگر باشند و بتوانند به طور کارایی جوانب گوناگون تصاویر را بیان کنند. روش‌های بسیاری مبتنی بر استخراج و توصیف ویژگی‌های محلی و سراسری تصویر معرفی شده‌اند. هر یک از این روش‌ها بر ویژگی خاصی از تصویر مانند شدت پیکسل‌ها، رنگ، بافت، لبه و... تأکید دارند. ویژگی‌های محلی نسبت به تغییرات روشنایی، اعوجاج پرسپکتیو، تبدیلات تصویر و همپوشانی بسیار حساس هستند اما قدرت تمایز بیشتری به دست می‌دهند. از طرف دیگر ویژگی‌های سراسری نسبت به نویز، تغییر مقیاس، انتقال و چرخش نامتغیر هستند.

یکی دیگر از مهم‌ترین نکات در بازیابی تصاویر به توصیف‌گر تصویر مربوط می‌شود. توصیف‌گر تلاش می‌کند ویژگی‌های مختلف تصویر را به صورت عددی استخراج کند. یک توصیف‌گر باید ویژگی‌های زیر را داشته باشد:

- ۱- تا حد ممکن «کامل»<sup>۱</sup> باشد یعنی باید بتواند تا جایی که ممکن است اطلاعات محتوای شیء را استخراج کند.
- ۲- تا حد ممکن «متراکم»<sup>۲</sup> باشد یعنی اندازه بردار ویژگی که توصیف‌گر استخراج می‌کند نباید زیاد بزرگ باشد.

در ادامه به مرور روش‌های بازیابی تصاویر با استفاده از دسته‌بندهای مختلفی چون SVM و شبکه‌های عصبی پرداخته شد. با استفاده از این روش‌ها در مقایسه با روش‌های سنتی می‌توان به نتایج بسیار بهتری در بازیابی تصاویر رسید زیرا روش‌های مبتنی بر دسته‌بندی تصویر با توجه به در نظر گرفتن دسته‌ای که

---

<sup>۱</sup>complete

<sup>۲</sup>compact

تصویر در آن قرار دارد باعث کاهش فاصله معنایی می‌شوند. مطالعات نشان می‌دهد که از شبکه‌های عصبی RBF چندان برای بازیابی تصاویر استفاده نشده است. استفاده از شبکه‌های عصبی به دلایل بسیاری نظیر «قابلیت عمومیت بخشی بالا»، «مقاوم بودن در برابر داده‌های نویزی» و «داده‌های گم‌شده»<sup>۱</sup> موجب افزایش کارایی، سرعت و دقت سیستم بازیابی می‌شود. بنابراین برای طراحی یک سیستم کارای بازیابی تصاویر استفاده از شبکه‌های عصبی برای پیش‌بینی دسته‌ای که تصویر پرس‌وجو به آن تعلق دارد بسیار مؤثر است. علاوه بر آن استفاده از شبکه‌های عصبی موجب افزایش سرعت بازیابی تصاویر می‌شود زیرا پس از آن که دسته مربوط به تصویر توسط شبکه پیش‌بینی شد فضای جست‌وجو تنها به تصاویر مربوط به آن دسته خاص محدود می‌شود و نیاز به بررسی کل تصاویر پایگاه داده نیست. از نکات مهم استفاده از شبکه‌های عصبی انتخاب معماری و پارامترهای مختلف شبکه مدل مانند نرخ یادگیری است. همچنین استفاده از الگوریتم‌های مؤثرتر آموزش می‌تواند باعث بالا رفتن سرعت یادگیری مدل شود. در این زمینه تحقیقات بسیاری در حال انجام شدن است و هر ساله شبکه‌های عصبی بهبود یافته‌ای معرفی می‌شوند که از آن‌ها در بازیابی تصاویر استفاده می‌شود. این بهبودها به طور عمده شامل یافتن الگوریتم‌های آموزش بهینه‌تر و در عین حال سریع‌تر است که موجب بالا رفتن کارایی شبکه در دسته‌بندی الگوها می‌شود.

---

<sup>۱</sup> Missed data

## فصل سوم

بررسی روش‌های مؤثر توصیف تصویر

### ۳-۱. مقدمه

در این فصل روش‌های مختلفی که در این رساله برای بازیابی تصاویر از آن‌ها استفاده شده است، توضیح داده می‌شوند. چنان‌که در فصل ۲ ذکر شد روش‌های جدیدتر بازیابی تصاویر از دسته‌بندی‌کننده‌های مختلف برای کم کردن فاصله معنایی استفاده می‌کنند. همچنین استفاده از توصیف‌گرهای جدید یکی دیگر از مواردی است که به افزایش کارایی سیستم‌های بازیابی تصاویر کمک شایانی کرده است. از دیگر عوامل مهم تأثیرگذار بر حوزه بازیابی تصاویر چگونگی ترکیب و استفاده از ویژگی‌های استخراج شده از تصاویر به کمک روش‌های جدید مبتنی بر «بسته واژگان تصویری»<sup>۱</sup> بسته ویژگی‌ها<sup>۲</sup> است.

در این فصل ابتدا روش‌های «بازنمایی معنایی تصویر»<sup>۳</sup> بیان می‌گردد. سپس روش بسته واژگان تصویری توضیح داده می‌شود و در نهایت چالش‌هایی که در این مسیر وجود دارند، بیان می‌شوند.

### ۳-۲. بازنمایی معنایی تصاویر

«معنای» یک تصویر مفهومی منحصر به فرد نیست و بازنمایی‌های معنایی مختلفی برای تصاویر بیان شده است. در حالت کلی می‌توان بازنمایی‌های معنایی تصاویر را به دو دسته تقسیم کرد:

الف) روش‌های سراسری : این روش‌ها ویژگی‌های معنایی تصویر را به صورت سراسری استخراج می‌کنند (Oliva and Torralba 2001).

ب) روش‌های محلی : این روش‌ها معنای تصویر را با مشخص کردن اشیای محلی تصویر مانند دریا، آفتاب، کوه و ... مشخص می‌کنند (Kanimozhi and Latha 2015).

---

<sup>۱</sup> Bag of Visual Words (BoVW)

<sup>۲</sup> Bag of Features (BoF)

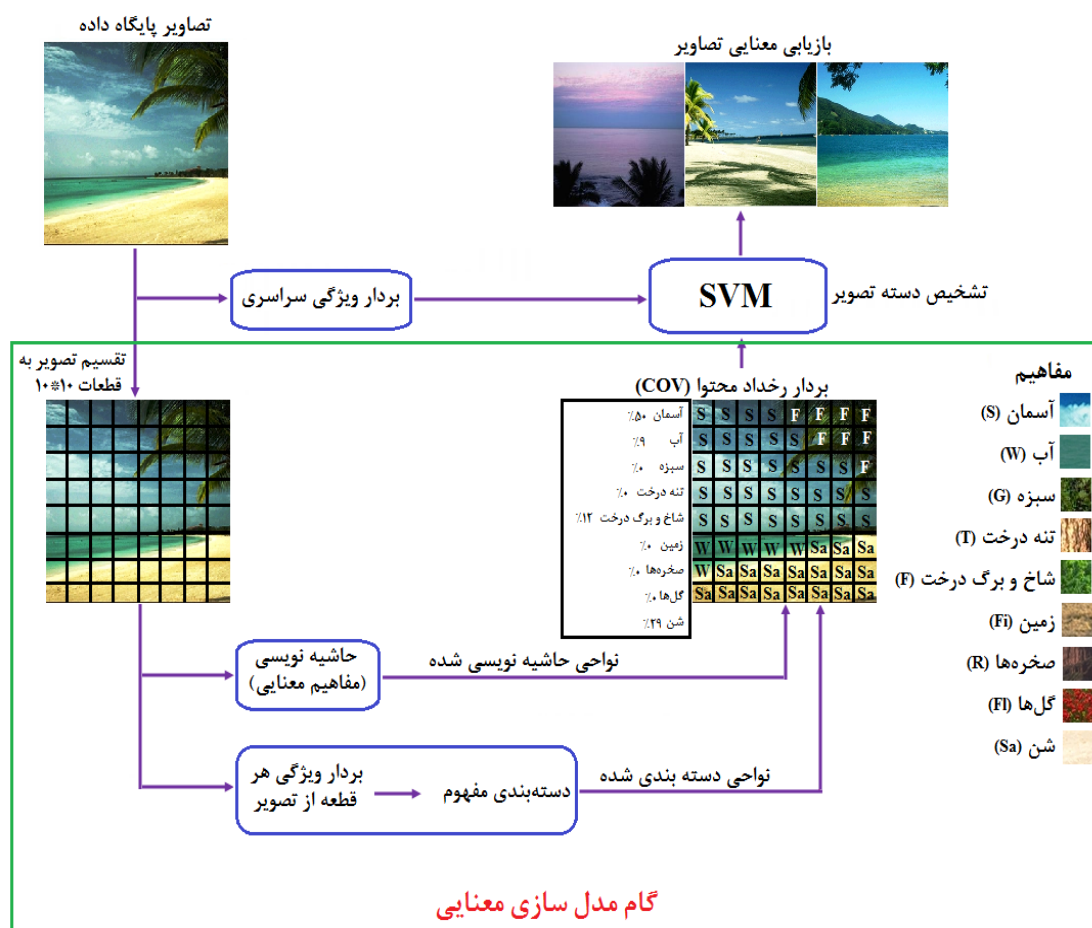
<sup>۳</sup> Semantic image representation

در روش‌های دسته اول معمولاً از روش‌های آماری مختلف برای توصیف تصویر استفاده می‌شود. این روش‌ها در مقایسه با روش‌های محلی از کارایی چندانی برخوردار نیستند (Alzu'bi, Amira and Ramzan2015).

محتوای معنایی محلی تصویر می‌تواند به عنوان روش بازنمایی مؤثرتری برای کاهش فاصله معنایی بین ویژگی‌های سطح پایین تصویر و ویژگی‌های سطح بالا، استفاده شود. به عنوان نمونه وجود بینی، چشم و ابرو در یک تصویر به معنای تصویر چهره و حضور درخت، کوه و برف نشانگر تصویری کوهستانی است (Alzu'bi, Amira and Ramzan2015). در روش‌های مبتنی بر بازنمایی محلی ابتدا تصویر به ناحیه‌های مختلفی تقسیم می‌شود و سپس دسته‌بندی کننده‌های محلی برای برچسب زدن این نواحی و تصمیم‌گیری درباره اینکه این نواحی به چه اشیایی تعلق دارند، استفاده می‌شوند. در نهایت با استفاده از این اطلاعات محلی دسته معنایی تصویر مشخص و شبیه‌ترین تصاویر به آن بازیابی می‌شوند (Yang et al 2014).

موجسیلوویچ و همکاران (Mojsilovic, Gomes and Rogowitz 2002) سعی کردند با استفاده از ویژگی‌های رنگ و بافت تصویر را قطعه‌بندی و اشیای مختلف تصویر نظیر آسمان، آب و... را شناسایی کنند. در نهایت از این اشیا برای یافتن دسته‌های معنایی تصاویر نظیر انسان‌ها، طبیعت، شیرها و... استفاده شد. رهیافت‌های مشابه دیگری نیز مبتنی بر قطعه‌بندی تصویر در تحقیقات دیگر پیشنهاد شده‌اند (Aksoy et al 2005; Barnard et al 2003; Duygulu et al 2002; Singhal, Luo and Zhu 2003; Fredembach, Schröder and Süssstrunk 2004).

از آنجا که قطعه‌بندی تصویر پیش پردازشی سنگین و گران محسوب می‌شود برای اولین بار وگل و شیله استفاده ساختاری شبکه‌ای را برای پیش‌پردازش پیشنهاد دادند (Vogel and Schiele 2007). ساختار شبکه‌ای، ساختاری است که در آن تصویر به زیرناحیه‌های منظمی تقسیم می‌شود. در این تحقیق نویسندگان از ویژگی‌های رنگ و بافت برای بازیابی و دسته‌بندی تصاویر استفاده کردند. روش پیشنهادی آن‌ها از دو مرحله تشکیل می‌شود: در مرحله اول تصویر به زیر ناحیه‌هایی به اندازه  $10 \times 10$  افراز می‌شود (شکل ۳-۱).



شکل ۳-۱: سیستم پیشنهادی توسط وگل و شیله (Vogel and Schiele 2007)

هر یک از این زیرناحیه‌ها به وسیله دسته‌بندی‌هایی چون K-NN یا SVM دسته‌بندی می‌شوند. سپس مفهومی به نام «بردار رخداد محتوا»<sup>۱</sup> تعریف می‌شود. این مفهوم تعداد دفعات وقوع اشیای مختلف در یک تصویر خاص را مشخص می‌سازد. به این ترتیب برای هر تصویر یک الگو ساخته می‌شود. از این الگوها می‌توان مستقیماً برای بازیابی تصاویر یا دسته‌بندی آن‌ها استفاده کرد.

سودرث و همکاران (Sudderth et al 2005) از یک روش آماری سلسله مراتبی برای بازشناسی اشیاء در تصاویر طبیعت با پس زمینه‌های شلوغ و پیچیده استفاده کردند. روش مذکور بر پایه ویژگی‌هایی که از بخش‌های تصویر استخراج می‌شود، استوار است. هر تصویر با اندازه‌گیری توزیع دسته‌های معنایی از تصاویر روی این بخش‌ها مدل می‌شود.

این روش‌ها سعی می‌کنند با توجه به برداشت انسان از تصاویر آن معنا را ضبط کنند. به این ترتیب با

<sup>۱</sup> Concept occurrence vector(COV)

استفاده از این روش‌ها امکان بازیابی تصاویری که در دسته‌های معنایی زیادی قرار گرفته‌اند با دقتی بالا فراهم می‌شود. هم‌چنان که روش‌های فوق مزیت‌هایی دارند، معایب مهمی نیز می‌توان برای آن‌ها در نظر گرفت. اکثر این روش‌ها بر بخش‌بندی تصویر استوار شده‌اند. چنانچه ملاحظه می‌شود روش‌های بخش‌بندی تصویر هنوز هم از قدرت و توانایی کافی برای بخش‌بندی کارایی انواع تصویر برخوردار نیستند. بنابراین بسته به قدرت روش بخش‌بندی استفاده شده، ممکن است کارایی این روش‌ها تحت تأثیر قرار گیرد. به علاوه برخی از این روش‌ها بر استخراج و کشف اشیای موجود در تصویر استوار هستند. اگر نتوان اشیای موجود در تصویر را به درستی استخراج و کشف کرد، این روش‌ها دچار شکست می‌شوند.

چنانچه اشاره شد دسته دیگر از روش‌های بازیابی تصاویر بر بازنمایی تصویر با استفاده از نواحی/ بلوک‌ها/ قسمت‌های محلی آن استوار هستند. در این روش‌ها با استفاده از تشخیص گر ناحیه<sup>۱</sup>، نواحی مهم تصویر شناسایی می‌شود و سپس با استفاده از توصیف‌گرهای تصویر، ناحیه‌های تشخیص داده شده توصیف و ویژگی‌هایشان استخراج می‌شوند. استفاده از توصیف‌گرهای محلی در تشخیص و بازشناسی اشیا به روشی بسیار کارآمد تبدیل شده است.

فرگوس و همکاران (Fergus, Perona and Zisserman 2003) دسته‌های تصاویر را به صورت مجموعه‌های احتمالاتی از ناحیه‌ها مدل می‌کنند. در این مقاله ظاهر هر ناحیه و نیز رابطه این ناحیه با سایر نواحی تصویر با استفاده از توزیع‌های گوسی مدل می‌شود. چنین مدل‌هایی بعدها به عنوان بسته واژگان تصویری معروف شدند.

روش‌های بسیاری بر پایه مدل بسته واژگان تصویری بنا نهاده شده‌اند. مروری کامل بر چنین روش‌هایی رامی‌توان در (Alzu'bi, Amira and Ramzan 2015) مشاهده کرد.

بنابراین در مدل‌های معنایی بازیابی تصاویر چند نکته قابل ملاحظه است : ۱. ویژگی‌هایی که قطعات تصویر استخراج می‌شود، ۲. چگونگی قطعه‌قطعه کردن تصویر، ۳. دسته‌بندی‌کننده استفاده شده برای آموختن دسته‌های معنایی تصاویر و ۴. چگونگی ترکیب ویژگی‌های استخراج شده از قطعات تصویر و تشکیل بردارهای ویژگی نهایی. در ادامه توضیحاتی درباره هر یک از موارد مذکور بیان می‌شود.

### ۳-۳. استخراج ویژگی‌های تصویر

به طور کلی استخراج ویژگی‌های تصویر به دو روش صورت می‌گیرد:

<sup>۱</sup> Region detector



در روش اول با استفاده از تشخیص گره‌های نقطه مورد علاقه<sup>۱</sup> نقاط مهم تصویر مانند گوشه‌ها، کیمنه‌های سراسری و محلی و... تشخیص داده می‌شوند. نقاط تشخیص داده شده، با استفاده از توصیف گره‌های<sup>۲</sup> مختلف، توصیف می‌شوند یعنی ویژگی‌هایشان استخراج می‌شود.

در روش دوم برخلاف روش اول، نقاط مهم تصویر استخراج نمی‌شوند و از همان ابتدا از توصیف گره‌های تصویر برای استخراج ویژگی‌ها استفاده می‌شود.

نقاط مورد علاقه یا همان نقاط کلیدی<sup>۳</sup> تصویر نقاطی هستند که حاوی اطلاعات غنی از تصویر می‌باشند. استخراج ویژگی‌ها از این نقاط باعث افزایش کارایی و متمایز بودن ویژگی‌های استخراجی می‌شود.

### ۱-۳-۳. تشخیص گره نقطه مورد علاقه

هر تشخیص گره مهمی با توجه به چگونگی طراحی ویژگی‌هایی از تصویر را مشخص می‌کند که نسبت به مواردی چون چرخش، مقیاس، پرسپکتیو، هم‌پوشانی و تغییرات روشنایی تا حدود متفاوت مقاوم است (Mikolajczyk and Schmid 2005). اساس کار بسیاری از تشخیص گره‌ها بر پایه آنالیز ریاضی و جبرخطی استوار است. در ادامه فهرستی از مهم‌ترین تشخیص گره‌های نقطه مورد علاقه ملاحظه می‌شود (Tuytelaars and Mikolajczyk 2008):

- (۱) Laplacian of Gaussian (LoG)
- (۲) Moravec corner detector
- (۳) Harris and Stephens corner detection
- (۴) Shi-Tomasi corner detector
- (۵) Difference of Gaussian (DoG ; an approximation of LoG)
- (۶) Harris-Laplace and Hessian-Laplace detectors
- (۷) Determinant of Hessian (DoH)
- (۸) Salient regions
- (۹) SUAN
- (۱۰) MSER

در این رساله به دلیل مزایای توصیف گره DoG از آن استفاده شده است (Tuytelaars and Mikolajczyk 2008).

---

<sup>۱</sup> Interest point detectors

<sup>۲</sup> Descriptor

<sup>۳</sup> Keypoints

## ۲-۳-۳. توصیف‌گر ویژگی

توصیف‌گرهای تصویر برای استخراج ویژگی‌های تصویر و بیان عددی ویژگی‌ها به کار می‌روند. برای توصیف یک تصویر روش‌های خاصی وجود دارد. برخی از این روش‌ها تصویر را مجموعه‌ای متشکل از تعدادی نواحی مهم، برخی تصویر را به صورت مجموعه‌ای متشکل از اشیای مهم و برخی نیز تصویر را به صورت مجموعه‌ای از نقاط مهم (نقاط کلیدی) توصیف می‌کنند. انواع مختلفی از توصیف‌گرهای تصویر وجود دارند که در این رساله از توصیف‌گرهای سیفت<sup>۱</sup>، LBP<sup>۲</sup> و HOG<sup>۳</sup> در انجام آزمایش‌های مختلف استفاده شده است؛ به همین دلیل تنها به توضیح آن‌ها اکتفا می‌شود.

### ۱-۲-۳-۳. تشخیص‌گر نقطه کلیدی سیفت

سیفت روشی برای توصیف تصویر با استفاده از نقاط کلیدی و ویژگی‌های استخراج شده از این نقاط است (Lowe 2004). با استفاده از این روش می‌توان اشیای مشابه را بازشناسی کرد. از مزایای این روش می‌توان به نامتغیر بودن نسبت به چرخش، تغییر مقیاس، تغییرات روشنایی تصویر و اعوجاج اشاره کرد. به این ترتیب که اگر شیء در تصویری وجود داشته باشد با کمک روش سیفت می‌توان همان شیء را اگر چرخیده، تغییر مقیاس داده شده، جابجا شده و یا کج و معوج شده باشد، بازشناسی کرد. ویژگی‌های استخراج شده با استفاده از سیفت بسیار از یکدیگر متمایز هستند که این مهم در یافتن تطابق‌ها در تصاویر بسیار حائز اهمیت است. برای استخراج ویژگی‌ها با استفاده از این روش چهار گام اصلی زیر اعمال می‌شود:

الف. تشخیص نقاط اکسترمم فضای مقیاس

ب. تعیین نقاط کلیدی

ج. تخصیص جهت

د- توصیف نقاط کلیدی

### الف. تشخیص نقاط اکسترمم فضای مقیاس

در این مرحله تصویر با توابع گوسی مختلف فیلتر می‌شود و در نتیجه ویژگی‌های کلیدی تصویر که نسبت به تغییر مقیاس نامتغیر هستند، استخراج می‌شوند. روش کار به این صورت است که ابتدا تصویر

---

<sup>۱</sup> Scale Invariant Feature Transform (SIFT)

<sup>۲</sup> Local Binary Patterns (LBP)

<sup>۳</sup> Histogram of Oriented Gradients (HOG)

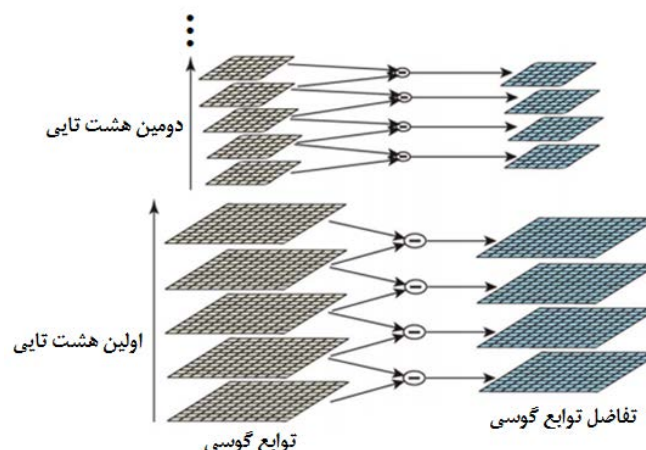
اصلی  $I(x, y)$  با توابع گوسی مختلف  $G(x, y, \sigma)$  کانولوشن می‌شود و تصاویر مختلفی را در فضای مقیاس به صورت زیر تولید می‌کند:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (۳-۱)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (۳-۲)$$

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (۳-۳)$$

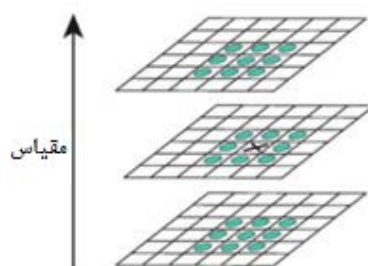
در این صورت نقاطی که نسبت به تغییر مقیاس ثابت هستند، نامزد می‌شوند. شکل ۲-۳ چگونگی انجام این روند را برای تصویر اصلی و تغییر اندازه یافته آن نشان می‌دهد.



شکل ۲-۳: توابع گوسی مختلف و تفاضل آن‌ها در هر مرحله هشت تابع گوسی (Lowe 2004)

### ب. تعیین نقاط کلیدی

پس از انتخاب تعدادی نقطه به عنوان نقاط نامزد که نقاط کلیدی و با ارزش تصویر محسوب می‌شوند، نقطه‌ای مانند  $X$  از این مجموعه نقاط نامزد شده، در صورتی انتخاب می‌شود که این نقطه بین ۲۶ همسایه‌اش بیشینه یا کمینه باشد. شکل ۳-۳ یک نقطه نامزد و ۲۶ همسایه‌اش را نشان می‌دهد. شکل ۳-۴ نقاط نامزد انتخاب شده با این روش را نشان می‌دهند.



شکل ۳-۳: بیشینه و کمینه تصویر تفاضل گوسی با مقایسه یک پیکسل (مانند X)



تصویر اصلی ۱۸۹×۲۲۲

۸۳۲ نقطه اکسترم انتخاب شده اولیه

شکل ۳-۴: تصویر اصلی و نقاط اکسترم انتخاب شده (Lowe 2004)

در مرحله بعدی مجدداً از بین نقاط انتخاب شده تعدادی نقطه با توجه به معیارهای خاصی مانند کیفیت نقاط (شکل ۳-۵ سمت چپ) و قرار داشتن روی گوشه‌ها و یا لبه‌ها (شکل ۳-۵ سمت راست) انتخاب می‌شوند.



شکل ۳-۵: تصویری با ۷۲۹ نقطه کلیدی (سمت چپ) تصویری با ۵۳۶ نقطه کلیدی (سمت راست) (Lowe 2004)

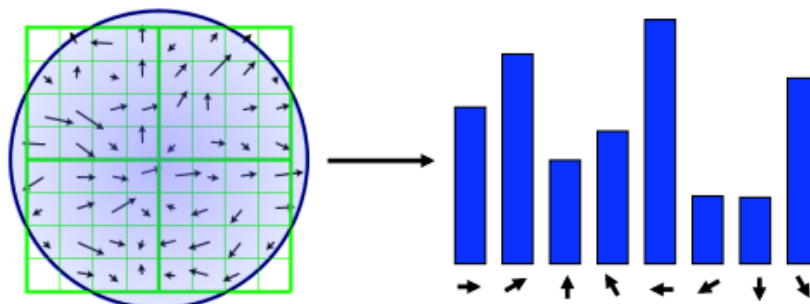
### ج. تخصیص جهت

اکنون تعدادی نقاط مناسب در اختیار است. در این مرحله هدف اختصاص جهت یا جهت‌هایی به نقاط انتخاب شده با استفاده از بردار گرادیان محلی تصویر است. از نتایج این مرحله توصیف نقاط به قسمی است که تحت تغییر مقیاس و چرخش، بدون تغییر باشند. به همین منظور با استفاده از مقیاس نقطه کلیدی (مقیاسی که نقطه کلیدی در آن انتخاب شده است، یعنی  $(\sigma)$ )، تصویر مورد نظر یعنی  $L(x, y, \sigma) = L(x, y)$  انتخاب می‌شود و سپس برای هر  $L(x, y)$  بزرگی بردار گرادیان  $(m(x, y))$  و جهت آن  $(\theta(x, y))$  به صورت زیر محاسبه می‌شوند:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (3-4)$$

$$\theta(x, y) = \tan^{-1}(L(x, y+1) - L(x, y-1) / L(x+1, y) - L(x-1, y)) \quad (3-5)$$

اکنون جهت چیره بردار گرادیان با محاسبه هیستوگرام روی ناحیه کوچکی در اطراف نقطه کلیدی تعیین می‌شود. شکل ۳-۶ نمونه‌ای از یک ناحیه انتخاب شده و هیستوگرام جهت‌های مختلف در آن ناحیه را نشان می‌دهد.

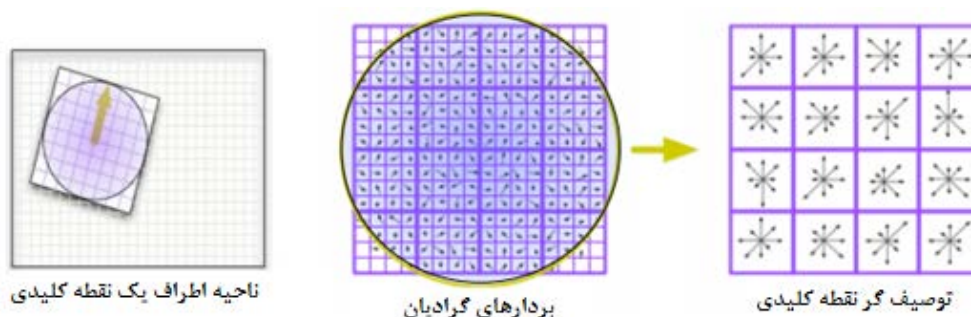


شکل ۳-۶: تشکیل هیستوگرام جهت به منظور تعیین جهت چیره (Schiele 2011)

اکنون مشخص کردن قله هیستوگرام شکل ۳-۷ امری بسیار مهم است زیرا این قله بیانگر جهت چیره لبه در آن ناحیه است. برای مشخص کردن جهت چیره، از قله‌هایی که ۸۰٪ مقدار بزرگ‌ترین قله را دارند، استفاده می‌شود.

#### د. توصیف نقاط کلیدی

در این مرحله به یک توصیف‌گر محلی در اطراف نقاط کلیدی برای توصیف آن نقطه نیاز است. به همین منظور هیستوگرام گرادیان روی پنجره‌ای به ابعاد  $4 \times 4$  و در ۸ جهت محاسبه می‌شود. بنابراین بردار ویژگی با  $4 \times 4 \times 8 = 128$  بُعد برای هر نقطه ساخته می‌شود (شکل ۳-۷).



شکل ۳-۷: تشکیل توصیف گر نقطه کلیدی (Schiele 2011)

پس از ساخته شدن بردار ویژگی به منظور تعبیه کردن ویژگی پایدار بودن نسبت به تغییرات روشنایی، بردار ویژگی طوری هنجار می‌شود که طول آن برابر با واحد می‌شود. پس از تشکیل بردارهای

ویژگی برای نقاط می‌توان نقاط منطبق بر هم در تصاویر را یافت. البته در این میان برای بهبود کارایی روش ممکن است نیاز به تشخیص و پاک‌سازی نقاط پرت<sup>۱</sup> باشد. ویژگی استخراج شده برای هر نقطه کلیدی با استفاده از روش سیفت، نسبت به انتقال، چرخش، تغییر مقیاس، اعوجاج‌های محلی هندسی و تغییرات روشنایی پایدار هستند. آزمایش‌های انجام شده در رساله نشان می‌دهد که در استفاده از روش سیفت برای بازیابی تصاویر مشکلاتی وجود دارد.

## ۲-۳-۳. بررسی عملکرد روش سیفت

روش سیفت برای هر یک از نقاط کلیدی تشخیص داده شده بردار ویژگی ۱۲۸ بعدی به دست می‌دهد. ویژگی‌های استخراج شده ویژگی‌هایی محلی با قدرت تمایزدهی بسیار زیاد هستند. کاربردهای بسیار مهم این الگوریتم در زمینه‌هایی چون «ردیابی ویدئویی»<sup>۲</sup>، «بازشناسی ژست»<sup>۳</sup>، «بازشناسی شیء»<sup>۴</sup> و «مدل‌سازی سه بعدی»<sup>۵</sup> (Lowe 2004) از انگیزه‌های بسیار مهم رساله در استفاده از این روش برای بازیابی تصاویر مشابه بوده‌است.

ویژگی استخراج شده برای هر نقطه کلیدی با استفاده از روش سیفت، نسبت به انتقال، چرخش، تغییر مقیاس، اعوجاج‌های محلی هندسی و تغییرات روشنایی پایدار هستند. اما از معایب این روش می‌توان به زمان‌بر بودن آن و نیاز به فضای بالا برای ذخیره‌سازی بردارهای ویژگی اشاره کرد.

ویژگی‌های استخراج شده با استفاده از روش سیفت مبتنی بر هیستوگرامی از لبه‌های تصویر هستند. بنابراین بررسی تحلیلی نقاط کلیدی و بردار ویژگی این نقاط اولین گام در شناخت روش سیفت است. در این بررسی پاسخ به سوالات زیر ضروری به نظر می‌رسد.

۱. تعداد نقاط تولید شده در روش سیفت به چه عواملی بستگی دارد؟
۲. آیا این روش همواره تعداد نقاط کلیدی زیادی به دست می‌آورد؟
۳. اگر تعداد نقاط کلیدی کم باشد قدرت روش سیفت تغییر می‌کند؟
۴. اگر تعداد نقاط کلیدی زیاد باشد قدرت روش سیفت افزایش می‌یابد؟
۵. از میان نقاط کلیدی چه نقاطی بیشتر اهمیت دارند؟

---

<sup>۱</sup> Outlier

<sup>۲</sup> Video tracking

<sup>۳</sup> Gesture recognition

<sup>۴</sup> Object recognition

<sup>۵</sup> 3D modeling

۶. آیا می‌توان از روش سیفت مستقیماً برای بازیابی تصاویر مشابه استفاده کرد؟  
با پاسخ به این سوالات می‌توان درک مناسبی از نقاط قوت و ضعف روش سیفت به دست آورد. یافتن پاسخ پرسش‌های بالا نیازمند اجرای روش روی تصاویر متفاوت است. بدین منظور در ادامه با اعمال روش سیفت روی تصاویر مختلف سعی در یافتن پرسش‌های بالا خواهد شد.  
در این مرحله آزمایشات مختلفی روی انواع مختلف تصاویر اعمال شد. در این آزمایشات سعی شده‌است که تأثیر عوامل مختلف نظیر قدرت روش سیفت در تولید نقاط کلیدی و زمان اجرای الگوریتم تطابق سیفت بررسی شود.

### ۱- روش سیفت و کیفیت تصویر

در تصاویر دیجیتال معمولاً مواردی از قبیل نویز، بلور شدن و تغییر شفافیت و کیفیت اتفاق می‌افتد. در آزمایشات صورت گرفته چنین مواردی مد نظر قرار داده شده‌اند.

در ادامه مثال‌هایی از تأثیر این عوامل روی نتیجه نهایی روش سیفت بررسی شده‌است. برای مثال شکل‌های ۳-۸ و ۳-۹ نتایج حاصل از اجرای روش سیفت روی پنج تصویر اصلی و تصاویر تغییر یافته آن‌ها نشان داده شده است. در این شکل‌ها برای هر تصویر نقاط کلیدی تصاویر با رنگ آبی نمایش داده شده‌اند. با کمی دقت می‌توان دریافت که در تصاویری که یکنواختی زیادی دارند و یا تغییرات آن‌ها کم است، مانند شن‌زارها، دریا، آسمان، تصاویر یک رنگ و ... تعداد نقاط کلیدی به شدت کم و در تصاویری با تغییرات زیاد و جزئیات فراوان، مانند تصاویر جنگل، تصاویر گیاهان، چمنزار و .. تعداد این نقاط بسیار زیاد است.

شکل‌های ۳-۸ و ۳-۹ نشان می‌دهند که روش سیفت همواره تعداد نقاط کلیدی زیادی تولید نمی‌کند. با کمی دقت در تصاویر مشخص است که این نقاط نقاطی هستند که با تغییر اندازه تصویر هنوز در تصویر حضور دارند. برای نمونه، لبه‌هایی که با تغییر اندازه تصویر هنوز از اهمیت بالایی برخوردارند هنوز هم به عنوان نقاط کلیدی انتخاب می‌شوند.

در تصاویری با جزئیات زیاد نقاط کلیدی در تمامی تصویر پراکنده هستند. اما در تصاویری که یک شیء خاص در آن از اهمیت بیشتری برخوردار است این نقاط در مرزهای این شیء به هم نزدیک‌تر هستند و محدوده آن شیء را مشخص می‌کنند.

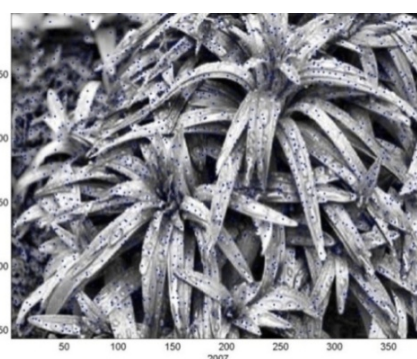
با انجام آزمایش روی پایگاه داده لی مشاهده می‌شود که بلور و نویز بازده روش سیفت را کاهش می‌دهند. به عبارت بهتر نویز باعث می‌شود که نقاط در مقیاس‌های مختلف از پایایی کمتری برخوردار

باشند. همچنین بلور کردن تصویر موجب کاهش کیفیت تصویر و از بین رفتن لبه‌ها و جزئیات آن می‌شود و در نتیجه تعداد نقاط کلیدی کاهش می‌یابد. از سوی دیگر با کاهش نویز می‌توان روش سیفت را بهبود بخشید.

دسته دیگر از آزمایش‌ها برای بررسی تأثیر روش سیفت و کیفیت تصاویر انجام شد. در تمامی این آزمایش‌ها ملاحظه شد با بهبود کیفیت تصاویر کارایی روش سیفت افزایش می‌یابد. دلیل این امر آن است که با افزایش کیفیت تصویر جزئیات تصویر نمایان‌تر و واضح‌تر می‌شوند و به همین دلیل روش سیفت می‌تواند تعداد نقاط کلیدی بیشتری از تصاویر استخراج کند.



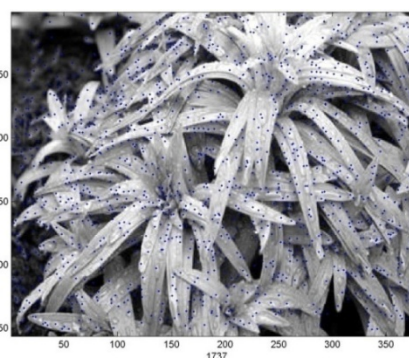
نقاط کلیدی بدست آمده از اجرای روش سیفت روی تصویر اصلی  
تعداد نقاط به دست آمده = ۱۶۰۳



نقاط کلیدی بدست آمده از اجرای روش سیفت روی تصویر اصلی  
با هیستوگرام سازگار شده<sup>۱</sup> - تعداد نقاط به دست آمده = ۲۰۰۷



نقاط کلیدی بدست آمده از اجرای روش سیفت روی بلور شده  
تعداد نقاط به دست آمده = ۳۶۵



نقاط کلیدی بدست آمده از اجرای روش سیفت روی تصویر اصلی  
با هیستوگرام تنظیم شده<sup>۲</sup> - تعداد نقاط به دست آمده =

<sup>۱</sup>Adaptive Histogram

<sup>۲</sup>Adjusted Histogram

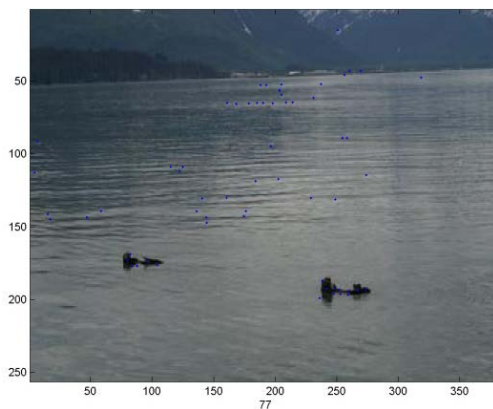


### ۳۷ □ بررسی روش‌های مؤثر توصیف تصاویر

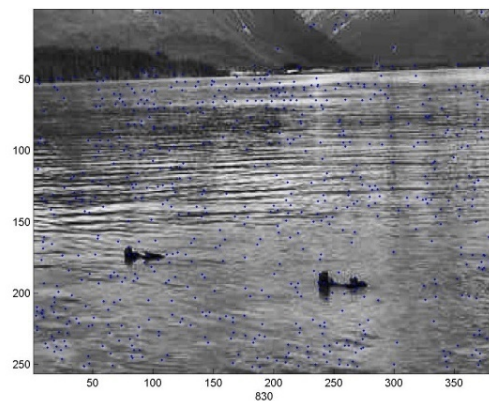


نقاط کلیدی بدست آمده از اجرای روش سیفت روی تصویر اصلی همراه با مقداری نویز تعداد نقاط به دست آمده = ۱۲۶۲

### شکل ۳-۸: نتایج ارزیابی تأثیر بهبود تصویر ۱ بر روش سیفت



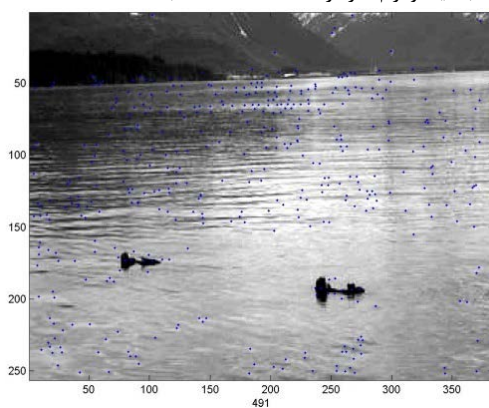
نقاط کلیدی بدست آمده از اجرای روش سیفت روی تصویر اصلی  
تعداد نقاط به دست آمده = ۷۷



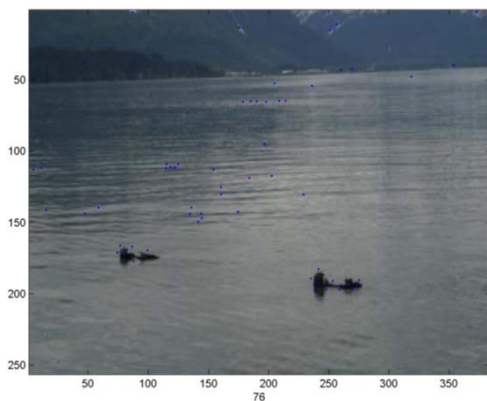
نقاط کلیدی بدست آمده از اجرای روش سیفت روی تصویر اصلی  
با هیستوگرام سازگار شده - تعداد نقاط به دست آمده = ۶۳۰



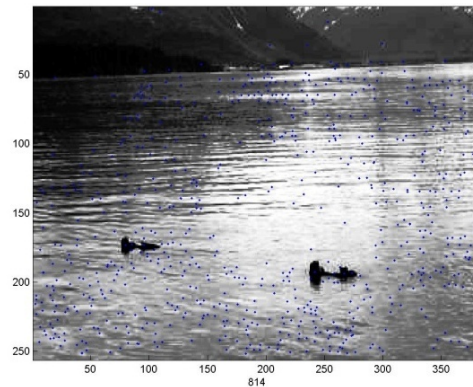
نقاط کلیدی بدست آمده از اجرای روش سیفت روی بلور شده  
تصویر اصلی تعداد نقاط به دست آمده = ۱۳



نقاط کلیدی بدست آمده از اجرای روش سیفت روی تصویر اصلی  
با هیستوگرام تنظیم شده - تعداد نقاط به دست آمده = ۴۹۱



نقاط کلیدی بدست آمده از اجرای روش سیفت روی تصویر اصلی  
با حذف نویز تعداد نقاط به دست آمده = ۷۶



نقاط کلیدی بدست آمده از اجرای روش سیفت روی تصویر اصلی با هیستوگراف  
یکسان شده - تعداد نقاط به دست آمده = ۸۱۴

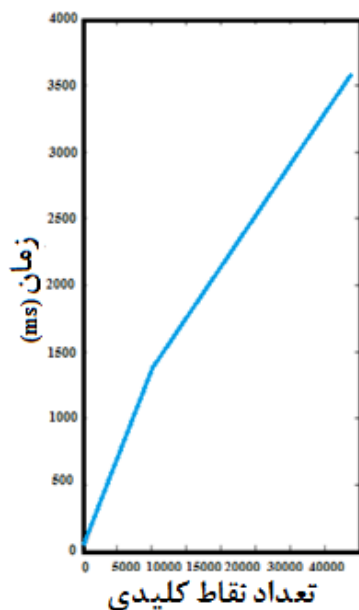


نقاط کلیدی بدست آمده از اجرای روش سیفت روی تصویر اصلی همراه با مقداری نویز  
تعداد نقاط به دست آمده = ۲۰۱

شکل ۳-۹: نتایج ارزیابی تأثیر بهبود تصویر ۲ بر روش سیفت

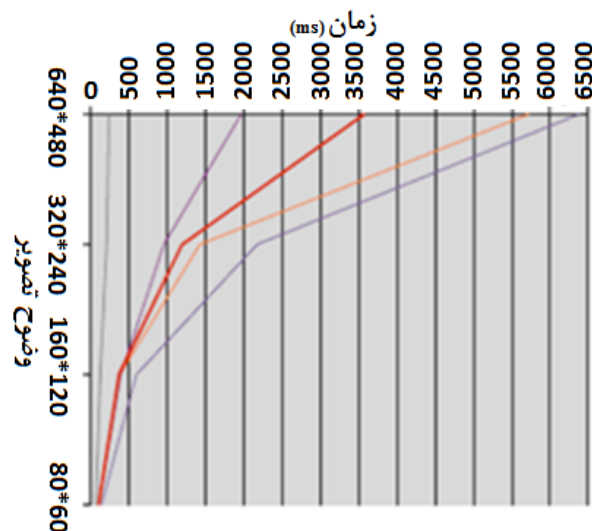
## ۲- رابطه میان نقاط کلیدی و زمان اجرای روش سیفت

در این آزمایش تعدادی تصویر با وضوح و اندازه‌های متفاوت انتخاب و روش سیفت روی آن‌ها اجرا شد. چنانچه از شکل ۳-۱۰ بر می‌آید زمان اجرای روش سیفت برای تولید و یافتن نقاط کلیدی با تعداد نقاط کلیدی رابطه‌ای تقریباً خطی دارد.



شکل ۳-۱۰: مدت زمان تولید نقاط کلیدی برای تصاویر با تعداد نقاط کلیدی متفاوت

همچنین در آزمایش‌های صورت گرفته رابطه بین زمان اجرای روش‌سیفت و وضوح تصویر بررسی شد. همان‌طور که در شکل ۳-۱۱ ملاحظه می‌شود با کاهش وضوح تصویر، زمان اجرای روش‌سیفت کاهش می‌یابد که دلیل آن کاهش تعداد نقاط کلیدی تصویر در اثر کاهش کیفیت آن است.



شکل ۳-۱۱: رابطه بین سرعت اجرای الگوریتم سیفت و وضوح تصاویر

در آزمایش دیگر رابطه بین زمان اجرا برای انطباق یک تصویر با ۵۸ تصویر دیگر بررسی شد. در این آزمایش تصاویر مختلف و ۵۸ تصویر دیگر انتخاب و سرعت روش سیفت برای انطباق بین آن‌ها آزمایش

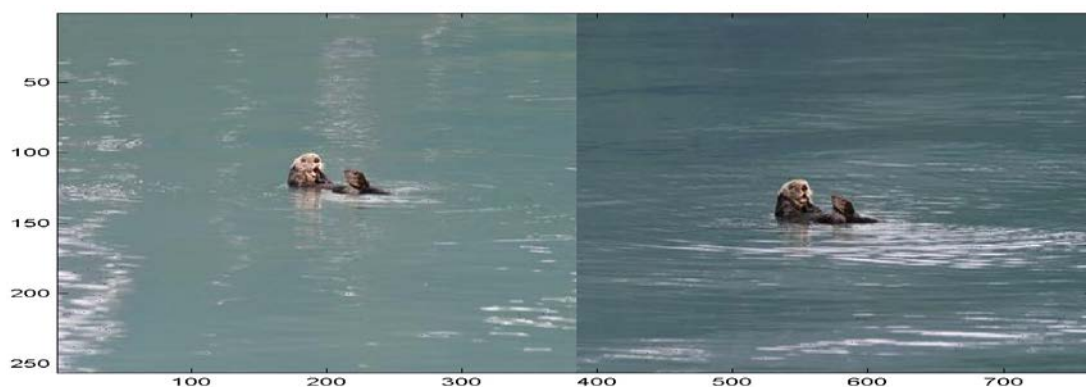
شد. که نتایج این آزمایش در جدول زیر ملاحظه می‌شود:

**جدول ۳-۱:** نتایج اجرای الگوریتم سیفت برای بررسی انطباق بین یک تصویر و ۵۸ تصویر دیگر

| ابعاد تصویر | زمان<br>(ثانیه:دقیقه) |
|-------------|-----------------------|
| ۳۲۰×۲۴۰     | ۷:۵۴                  |
| ۳۸۴×۲۸۸     | ۹:۵۸                  |
| ۹۶×۷۲       | ۱:۵۸                  |
| ۶۴۰×۴۸۰     | ۱۸:۴۲                 |

### ۳- حالت‌های خاص انطباق تصاویر<sup>۱</sup>

پس از بررسی ویژگی‌های و شرایط خاص در روش سیفت، بررسی الگوریتم انطباق نیز بسیار حائز اهمیت است. در زیر نتایج اجرای انطباق تصاویر روی دو دسته از تصاویر مشابه ملاحظه می‌شود. در شکل‌های ۳-۱۲ و ۳-۱۳ تصاویر بسیار شبیه هستند. این تصاویر منطبق بر هم به نظر می‌رسند و تفاوت کمی میان آن‌ها دیده می‌شود. با این حال الگوریتم تطابق سنجی روش سیفت هیچ مشابهتی میان این تصاویر پیدا نمی‌کند. نکته مهم اینجاست با انجام عملیات کاهش نویز در تصویر بردارهای ویژگی به یکدیگر نزدیک‌تر می‌شوند.



شکل الف). نتیجه اجرای الگوریتم تطابق روی تصویر اصلی، تعداد نقاط مشابه = ۰

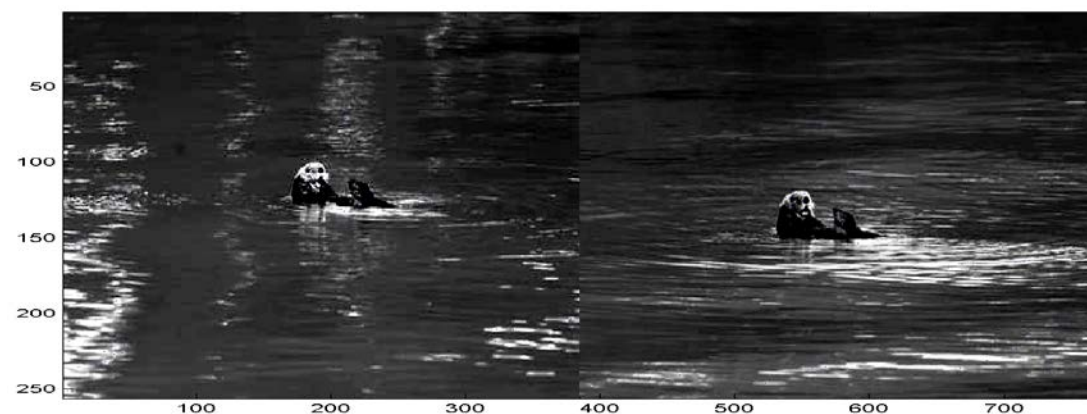




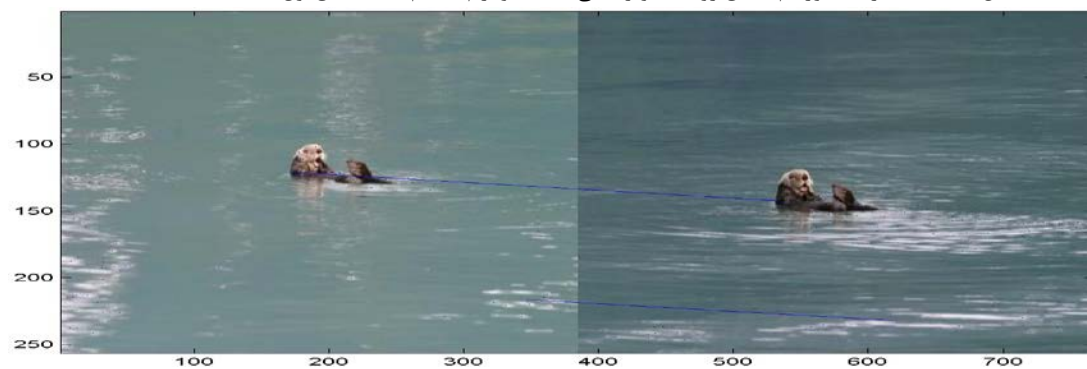
شکل ب) نتیجه اجرای الگوریتم تطابق روی تصاویر اصلی با هیستوگرام سازگار شده، تعداد نقاط مشابه = ۰



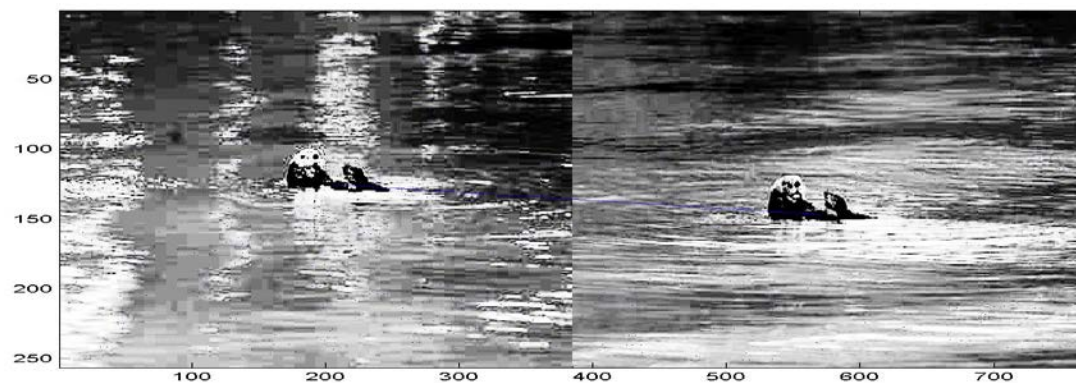
شکل ج) نتیجه اجرای الگوریتم تطابق روی بلور شده تصاویر اصلی (تطابق اول)، تعداد نقاط مشابه = ۱



شکل د) نتیجه اجرای الگوریتم تطابق روی تصاویر اصلی با هیستوگرام تنظیم شده (تطابق اول)، تعداد نقاط مشابه = ۰



شکل ه) نتیجه اجرای الگوریتم تطابق بر روی تصاویر اصلی با حذف نویزها (تطابق اول)، تعداد نقاط مشابه = ۲

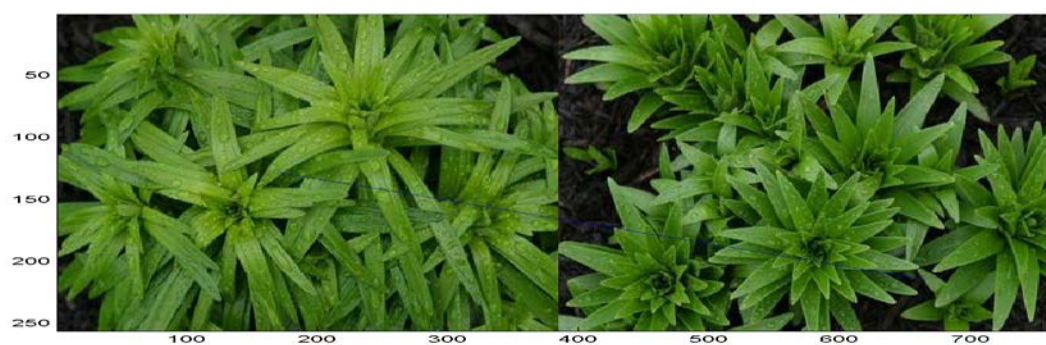


شکل و) نتیجه اجرای الگوریتم تطابق روی تصاویر اصلی با هیستوگرام یکسان شده (تطابق اول)، تعداد نقاط مشابه = ۱

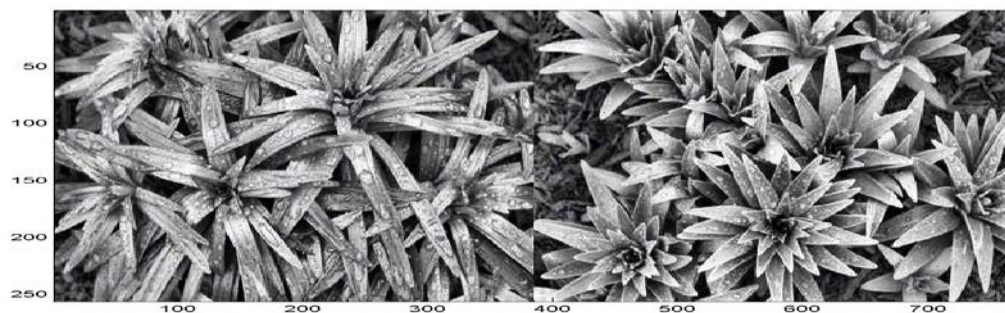


شکل ز) نتیجه اجرای الگوریتم تطابق تصاویر اصلی همراه با نویز (تطابق اول)، تعداد نقاط مشابه = ۰

شکل ۳-۱۲: (تطابق اول)، تطابق دو تصویر و تصاویر دستکاری شده دوتصویر اصلی

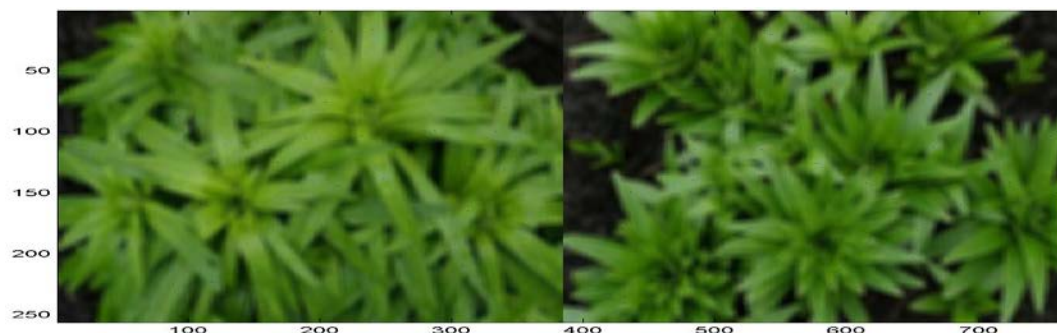


شکل الف) نتیجه اجرای الگوریتم تطابق روی تصاویر اصلی (تطابق دوم)، تعداد نقاط مشابه = ۱

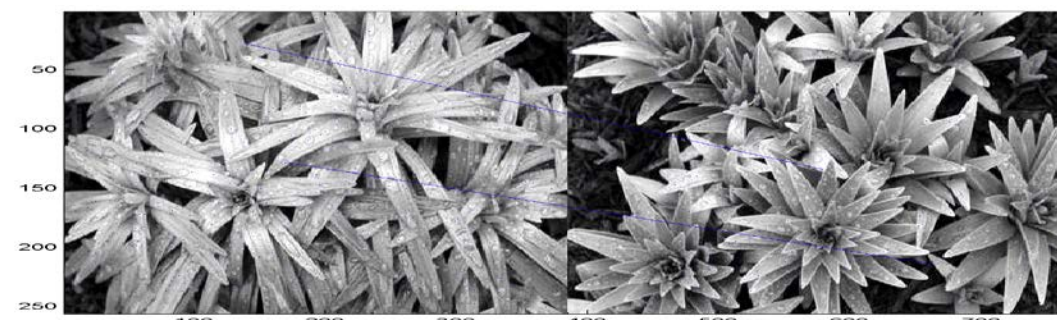


شکل ب) نتیجه اجرای الگوریتم تطابق روی تصاویر اصلی با هیستوگرام سازگار شده (تطابق دوم)، تعداد نقاط مشابه = ۰





شکل ج) نتیجه اجرای الگوریتم تطابق روی بلور شده تصاویر اصلی (تطابق دوم)، تعداد نقاط مشابه = ۰



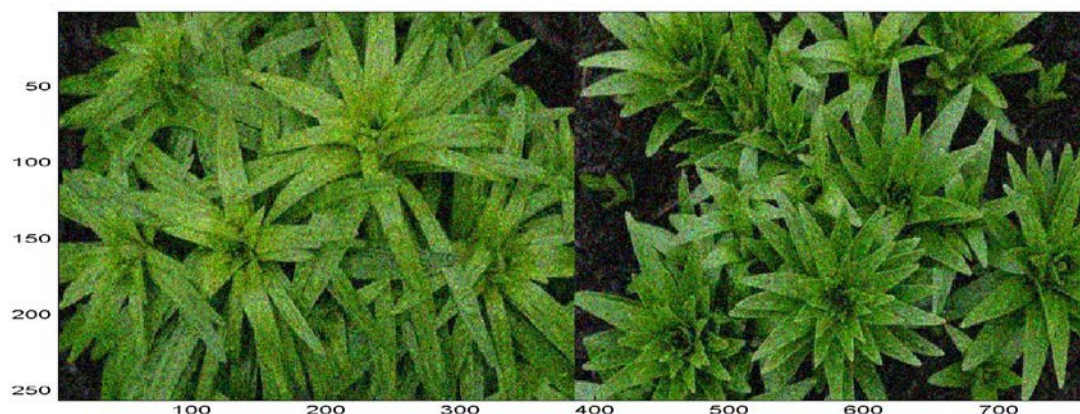
شکل د) نتیجه اجرای الگوریتم تطابق بر روی تصاویر اصلی با هیستوگرام تنظیم شده (تطابق دوم)، تعداد نقاط مشابه = ۲



شکل ه) نتیجه اجرای الگوریتم تطابق روی تصاویر اصلی با حذف نویزها (تطابق دوم)، تعداد نقاط مشابه = ۱



شکل و) نتیجه اجرای الگوریتم تطابق روی تصاویر اصلی با هیستوگرام یکسان شده ، تعداد نقاط مشابه = ۰



شکل ز) نتیجه اجرای الگوریتم تطابق تصاویر اصلی همراه با نویز، تعداد نقاط مشابه = \*

### شکل ۳-۱۳: تطابق دو تصویر و تصاویر دستکاری شده دو تصویر اصلی

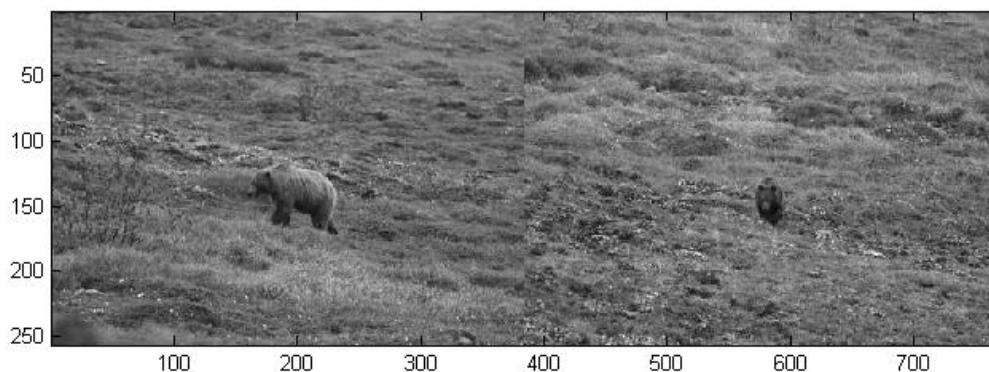
برای بازیابی تصاویر با استفاده از روش سیفت اولین نکته‌ای که به ذهن می‌رسد توجه به ماهیت این روش است. سیفت روشی است که نقاط کلیدی تصویر همراه با ویژگی‌های آن‌ها را استخراج می‌کند. پس می‌توان حدس زد که اگر دو تصویر بخواهند با هم مشابه باشند باید تعداد نقاط کلیدی آن‌ها و قاعده‌تاً تعداد نقاط منطبق آن‌ها که الگوریتم تطابق سیفت به دست می‌دهد، به هم نزدیک باشند. حال اگر به تصاویر شکل‌های ۳-۱۲، ۳-۱۳ و نیز تصاویر زیر و تعداد نقاط کلیدی آن‌ها توجه شود ملاحظه می‌شود چنین فرضی اصلاً برای روش سیفت برقرار نیست.



شکل ۳-۱۴: تصویر سمت راست با ۱۰۹۹ و تصویر سمت چپ با ۱۰۸۳ نقطه کلیدی

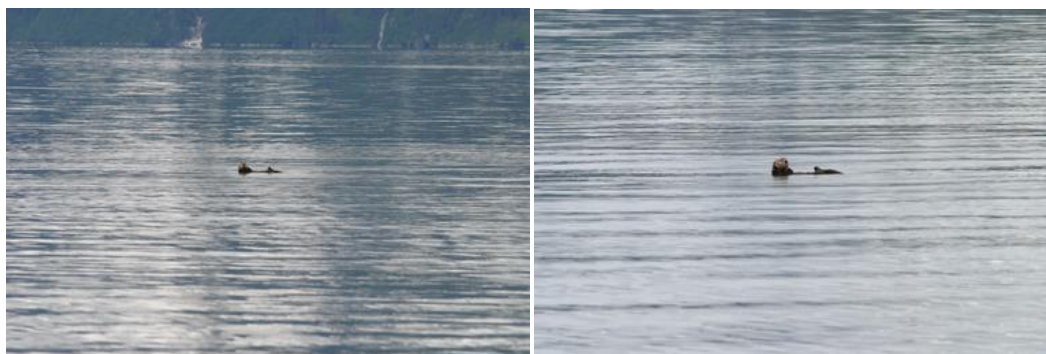
در شکل ۳-۱۴ دو تصویر مشابه با تعداد نقاط کلیدی زیاد و نسبتاً برابر دیده می‌شوند. در شکل ۳-۱۵ نتیجه اجرای روش سیفت برای این دو شکل دیده می‌شود.



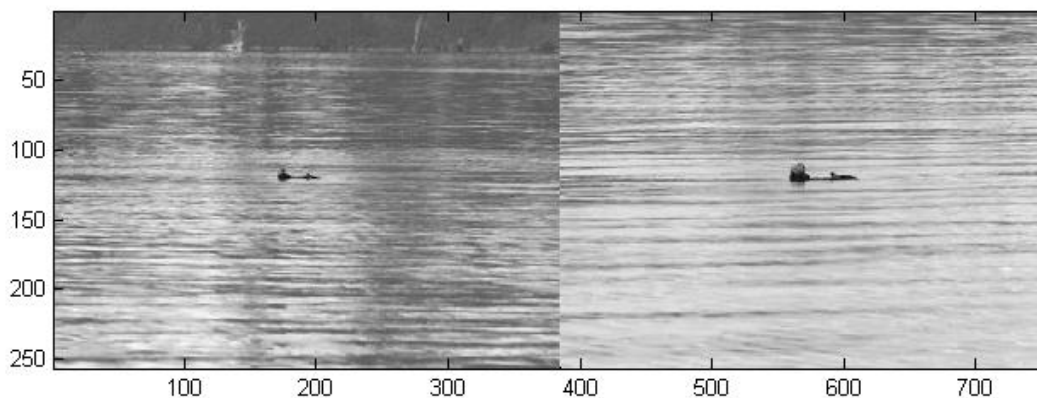


شکل ۳-۱۵: یافتن تنها یک نقطه مشابه به وسیله روش سیفت

همانطور که در شکل ۳-۱۵ ملاحظه می‌شود روش سیفت برای دو تصویر کاملاً مشابه هیچ نقطه انطباقی نیافته است. در شکل‌های ۳-۱۶ و ۳-۱۷ نمونه‌های دیگری از تصاویر دیده می‌شود که این تصاویر به هم شبیه هستند اما روش سیفت تشابهی بین آن‌ها نمی‌یابد.



شکل ۳-۱۶: دو تصویر مشابه به ترتیب از راست به چپ با ۳۹۵ و ۸۴۶ نقطه کلیدی



شکل ۳-۱۷: عدم یافتن هیچ دو نقطه مشابهی

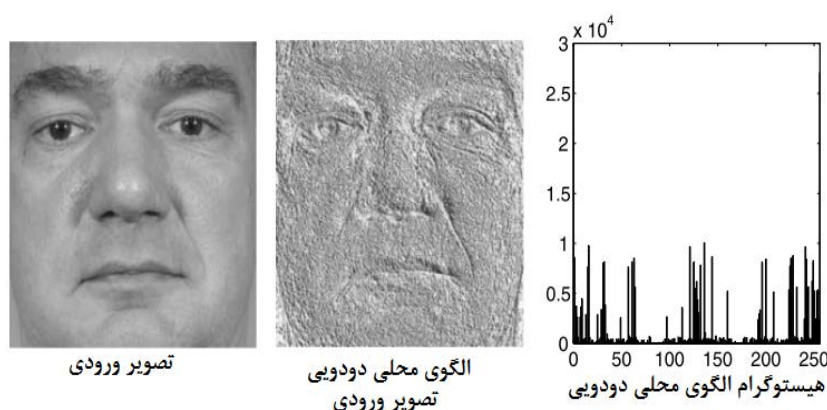
پس همان‌طور که ملاحظه می‌شود فرضیه استفاده مستقیم از روش سیفت برای بازیابی تصاویر مشابه پاسخ مطلوبی را تولید نمی‌کند. در این رساله نسخه بهبود یافته‌ای از توصیف‌گر سیفت معرفی می‌شود که

می‌توان از آن در بازیابی و دسته‌بندی تصاویر استفاده کرد.

### ۳-۳-۲-۳. توصیف‌گر LBP

LBP به عنوان روشی جدید برای استخراج ویژگی‌های الگو و تباین<sup>۱</sup> تصویر به منظور توصیف بافت معرفی شد (Ojala, Pietikäinen and Mäenpää 2002; Zhang et al 2010; Guo, Zhang and Zhang 2010). توصیف‌گر محلی LBP با انجام عملیاتی روی یک پیکسل و همسایگی آن ویژگی‌هایی به صورت هیستوگرام به دست می‌دهد. این توصیف‌گر برای توصیف پیکسلی دلخواه و همسایگی آن یک رشته دودویی تولید می‌کند. در نسخه اصلی LBP از یک همسایگی  $3 \times 3$  در اطراف پیکسل موردنظر استفاده می‌شود.

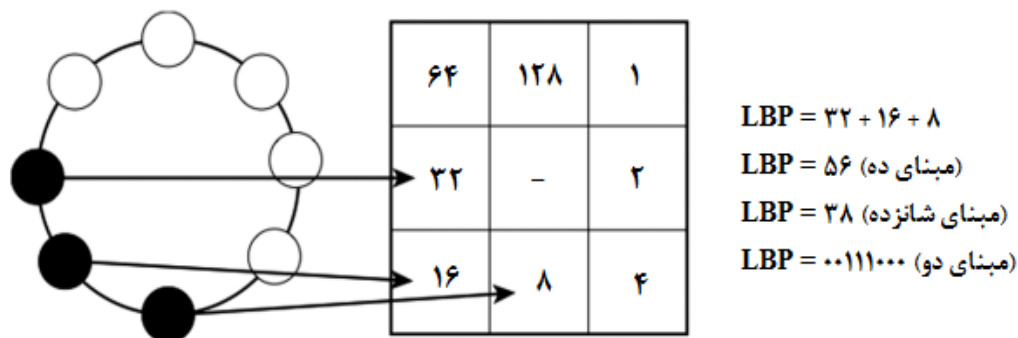
روش کار به این صورت است که پیکسل مرکزی با همسایه‌هایش مقایسه می‌شود اگر از همسایه‌ای بزرگتر بود، به جای آن عدد ۱ و در غیر این صورت به جای همسایه عدد ۰ قرار داده می‌شود. به این ترتیب برای نمایش یک همسایگی، یک الگوی دودویی تولید می‌شود. از آنجا که در یک همسایگی  $3 \times 3$ ، پیکسل مرکزی با هشت پیکسل دیگر همسایه است،  $2^8 = 256$  الگوی دودویی مختلف با توجه به مقادیر پیکسل‌ها متصور است شکل ۳-۱۸ تصویر چهره و الگوی دودویی شده آن به همراه بردار ویژگی استخراج شده را نشان می‌دهد.



شکل ۳-۱۸: نمونه‌ای از تصویر چهره به همراه تصویر LBP متناظر با آن و بردار ویژگی استخراج شده

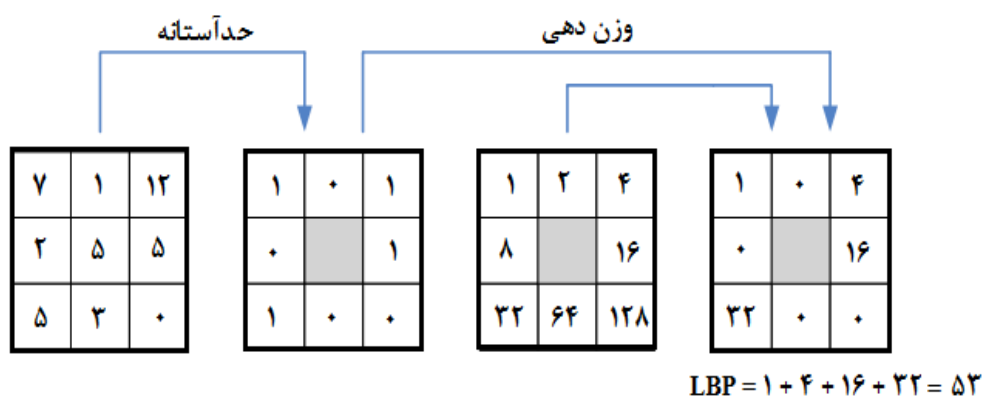
برای محاسبه عدد دودویی متناظر با الگوی دودویی تولید شده باید وزن هر یک از همسایه‌ها در مبنای دو مشخص گردد. برای تعیین وزن‌های دودویی همانند شکل ۳-۱۹ عمل می‌شود.

<sup>۱</sup> Contrast



شکل ۳-۱۹: تعیین وزن‌های همسایه‌ها در محاسبه عدد دسیمال متناظر با الگوی تولید شده

شکل ۳-۲۰ نمونه‌ای از الگوی تولید شده به وسیله توصیف‌گر LBP را نشان می‌دهد.



شکل ۳-۲۰: الگوی تولید شده توسط توصیف‌گر LBP

می‌توان برای تصویر خاکستری  $I(x, y)$ ، پیکسل مرکزی  $g_c$  (که  $g_c = I(x_c, y_c)$ ) و پیکسل‌هایی که در همسایگی  $R$  از  $g_c$  قرار دارند (و با  $g_p$  نشان داده می‌شوند)، توصیف‌گر LBP را به صورت زیر تعریف کرد:

$$LBP_{R,R}(x_c, y_c) = \sum_{p=0}^{P-1} S(g_p - g_c) 2^p \quad (3-6)$$

بنابراین توزیع سطح خاکستری تصویر یعنی بافت را می‌توان با هیستوگرامی با  $2^P$  بین تخمین زد. چنانچه ذکر شد در یک همسایگی  $3 \times 3$ ، بردارهای ویژگی در فضای  $256$  بُعدی تولید می‌شوند. می‌توان از نوع خاصی از LBP که از کارایی LBP اصلی برخوردار است و بردارهای ویژگی با ابعاد کوچکتر تولید می‌کند نیز استفاده کرد. این عملگر از توسعه LBP اصلی حاصل می‌شود و LBP یکنواخت<sup>۱</sup> نامیده می‌شود (Zhang et al 2010; Guo, Zhang and Zhang 2010).

<sup>۱</sup> Uniform

برای توضیح LBP یکنواخت ابتدا باید معیاری به نام «سنجه یکنواختی»<sup>۱</sup> را تعریف کرد. سنجه یکنواختی تعداد گذرهای بیتی از ۰ به ۱ و برعکس است. باید دقت شود در محاسبه سنجه یکنواختی الگوی تولید شده به صورت دایره‌ای در نظر گرفته می‌شود. یعنی اولین و آخرین بیت با هم همسایه محسوب می‌شوند. یک الگوی دودویی محلی، یکنواخت نامیده می‌شود اگر سنجه یکنواختی آن حداکثر برابر با دو باشد. برای مثال سنجه یکنواختی الگوهای ۰۰۰۰۰۰۰۰، ۰۱۱۱۰۰۰۰ و ۱۱۰۰۱۱۱۱ به ترتیب برابر است با صفر، دو و دو.

یکی از موارد مهمی که کارایی توصیف‌گر LBP را کاهش می‌دهد این است که این توصیف‌گر تنها روی نسخه خاکستری تصاویر عمل می‌کند. در این رساله ایده اعمال LBP در فضاهای رنگی مختلف و ترکیب ویژگی‌های استخراج شده از هر یک از این کانال‌ها مطرح می‌شود.

#### ۴-۲-۳. توصیف‌گر HOG

توصیف‌گر HOG یکی از محبوب‌ترین توصیف‌گرهایی است که در پردازش تصویر و بینایی ماشین استفاده می‌شود. این توصیف‌گر تعداد وقوع جهت بردار گرادیان را در یک تصویر قسمت‌بندی شده مشخص می‌کند (Dalal and Triggs 2005).

HOG با استفاده از توزیع جهت‌های لبه<sup>۲</sup> (گرادیان‌های شدت<sup>۳</sup>) تصویر، ویژگی‌هایی محلی ظاهری شیء و شکل آن را توصیف می‌کند. در این روش ابتدا تصویر به ناحیه‌های کوچک متصل به هم (که سلول نامیده می‌شوند)، تقسیم می‌شود. سپس برای هر سلول هیستوگرام جهت‌های لبه محاسبه می‌شود. ترکیب این هیستوگرام‌ها توصیف‌گر نهایی را ایجاد می‌کند. برای بهبود کارایی می‌توان هیستوگرام‌های محلی را هنجارسازی کرد. این کار با اندازه‌گیری تابینحول سلول و در یک ناحیه بزرگ‌تر از تصویر که بلوک نامیده می‌شود، صورت می‌گیرد. این هنجارسازی باعث مقاوم‌تر شدن ویژگی‌های استخراج شده در برابر تغییرات روشنایی و سایه می‌شود. به بیان دقیق‌تر استخراج ویژگی‌های تصویر به وسیله توصیف‌گر HOG در چهار مرحله زیر صورت می‌گیرد:

الف) محاسبه گرادیان

ب) «بین‌سازی جهت»<sup>۴</sup>

---

<sup>۱</sup> Uniformity measure

<sup>۲</sup> Edge directions

<sup>۳</sup> Intensity gradients

<sup>۴</sup> Orientation binning

ج) ساختن توصیف‌گر بلوکی

د) هنجارسازی بلوک

### الف. محاسبه گرادیان

اولین مرحله الگوریتم HOG محاسبه بردار گرادیان است. در اکثر موارد برای محاسبه گرادیان از فیلتر کردن تصویر خاکستری با کرنل‌های  $D_x = [-1 \ 0 \ 1]$  و  $D_y = [0 \ 1 \ -1]^T$  استفاده می‌شود.

### ب. «بین» سازی جهت

در دومین مرحله از الگوریتم HOG، اگر جهت بردارهای گرادیان علامت‌دار در نظر گرفته شوند دامنه تغییرات از  $0^\circ$  تا  $360^\circ$  خواهد بود، اما اگر جهت‌های محاسبه شده بدون علامت در نظر گرفته شوند، دامنه تغییرات از  $0^\circ$  تا  $180^\circ$  خواهد بود. در اکثر موارد برای کوانتیزه کردن جهت‌های محاسبه شده نه «بین» در نظر گرفته می‌شود.  $0^\circ$  تا  $20^\circ$ ،  $20^\circ$  تا  $40^\circ$ ، ... و در نهایت  $160^\circ$  تا  $180^\circ$ . تخصیص جهت‌ها به «بین»‌ها با استفاده از درون‌یابی انجام می‌شود.

### ج. توصیف‌گر بلوک‌ها

برای مقاوم کردن ویژگی‌های استخراج شده در برابر تباین و تغییرات روشنایی، سلول‌ها را در گروه‌هایی قرار داده (بلوک‌ها) و بردارهای ویژگی هر سلول را به بردارهای ویژگی سلول‌های دیگر الحاق می‌کنند و به این ترتیب بردار ویژگی برای هر بلوک حاصل می‌شود. با هنجارسازی بردارهای ویژگی بلوک‌ها، توصیف‌گر حاصل در برابر تباین و تغییرات روشنایی مقاوم‌تر می‌شود.

### د. هنجارسازی بلوک‌ها

برای هنجارسازی ویژگی‌های هر بلوک روش‌های مختلفی وجود دارد. برای نمونه می‌توان از نرم  $L_1$  استفاده کرد.

توصیف‌گرهای مذکور در قالب روش «بسته واژگان تصویری»<sup>۱</sup> برای بازیابی معنایی تصویر به کار می‌روند. در ادامه روش بسته واژگان تصویری به طور مختصر توضیح داده می‌شود.

## ۴-۳. روش بسته واژگان تصویری

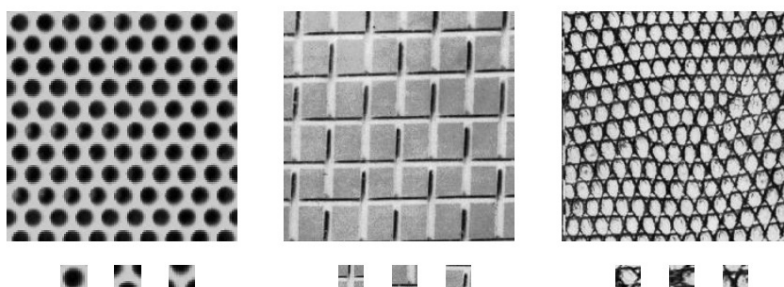
بدون تردید امروزه روش بسته واژگان تصویری یکی از کاراترین روش‌ها برای بازیابی معنایی تصاویر

---

<sup>۱</sup> Bag of Visual Words (BoVW)

است (Alzu'bi, Amira and Ramzan 2015). در این روش ابتدا تصویر به قطعاتی تقسیم و سپس ویژگی‌های مختلف (نظیر سیفت ، LBP و...) از هر یک از این قطعات استخراج می‌شود. هر یک از این ویژگی‌ها یک «واژه» (لغت) نام دارد. در گام بعدی برای کاهش تعداد واژه‌ها یک روش خوشه‌بندی روی همهٔ واژگان استخراج شده از همهٔ تصاویر اعمال می‌شود. مراکز خوشه‌ها جداسازی شده و «دایرهٔ لغت/لغت‌نامه/دایرهٔ واژگان» را تشکیل می‌دهد. سپس برای توصیف هر تصویر «هیستوگرامی» روی دایرهٔ واژگان برای هر تصویر محاسبه و به عنوان بردار ویژگی تصویر معرفی می‌شود.

ایدهٔ استفاده از این روش ابتدا به نمایش تصاویر بافتی و سپس به دسته‌بندی اسناد باز می‌گردد. در نمایش تصاویر بافتی، ویژگی بافت براساس عناصر سازندهٔ آن‌ها یعنی «تکستن<sup>۱</sup>» (که به آن‌ها «کدبوک<sup>۲</sup>» نیز گفته می‌شود)، استخراج می‌شوند. این ویژگی‌ها به عنوان ویژگی‌های سطح میانی از یک تصویر محسوب می‌شوند. شکل ۳-۲۱ سه تصویر بافتی و برخی عناصر سازندهٔ آن‌ها (تکستن‌ها) را نشان می‌دهد.



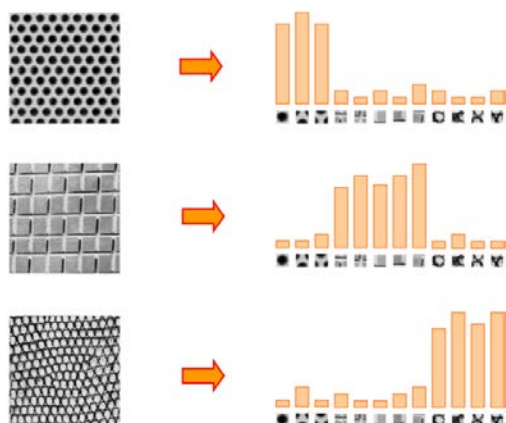
شکل ۳-۲۱: تصاویر بافتی مختلف و تکستون‌های مربوط به آن‌ها

در دسته‌بندی اسناد «بستهٔ واژگان» بردار تنکی<sup>۳</sup> از تعداد دفعات وقوع هر واژه در آن سند است. به همین ترتیب در یک تصویر بستهٔ واژگان برابر با هیستوگرامی است که هر یک از واژگان تصویر (تکستون‌ها در تصویر بافتی) را شمارش می‌کند. شکل ۳-۲۲ چنین هیستوگرام‌هایی را برای سه تصویر بافتی نشان می‌دهد.

<sup>۱</sup> Texton

<sup>۲</sup> codebook

<sup>۳</sup> Sparse



شکل ۳-۲۲: نمایش بسته واژگان تصویری برای سه تصویر بافتی



شکل ۳-۲۳: تصاویر واژگان مربوط به گیتار، دوچرخه و صورت انسان



شکل ۳-۲۴: دایره واژگان ساخته شده توسط تصاویر گیتار، دوچرخه و صورت

برای نمایش یک تصویر با استفاده از روش بسته واژگان به طور معمول سه مرحله زیر در نظر گرفته می‌شود (Fei-fei and Perona 2005).

۱- تشخیص ویژگی

۲- توصیف ویژگی

۳- تولید دایره واژگان

برای ساختن دایره واژگان ابتدا تصاویر به قطعاتی (معمولاً به اندازه  $10 \times 10$ ) تقسیم و سپس ویژگی‌هایی از این قطعات استخراج می‌شود به این ترتیب می‌توان هر تصویر را توسط ویژگی‌های

استخراج شده از قطعات آن توصیف کرد. می‌توان از هر یک از روش‌های توصیف ویژگی که در قسمت قبل بیان شد برای بیان ویژگی‌های استخراج شده از تصاویر، استفاده کرد. به عنوان نمونه فرض می‌کنیم از توصیف‌گر سیفت برای بیان ویژگی‌ها استفاده شود. چنان‌که گفته شد الگوریتم سیفت برای هر پیکسل تصویر یک بردار ۱۲۸ بُعدی به دست می‌دهد. آخرین مرحله از روش بسته‌ی واژگان تبدیل ویژگی‌های استخراج شده از قطعات به «کدواژه‌ها» و تولید دایره‌ی واژگان است.

یکی از مؤثرترین روش‌ها برای تولید دایره‌ی واژگان، خوشه‌بندی واژگان به روش K-means است. پس از انجام خوشه‌بندی می‌توان کدواژه‌ها را به عنوان مراکز خوشه‌ها در نظر گرفت. اندازه‌ی دایره‌ی واژگان برابر با تعداد خوشه‌های تولید شده برابر است، به همین دلیل هر قطعه از تصویر با استفاده از خوشه‌بندی به یک کد واژه معین نگاشت می‌شود. در نتیجه می‌توان هر تصویر را توسط هیستوگرامی از واژگان بیان کرد. برای آموختن بسته‌ی ویژگی‌ها (هیستوگرام واژگان<sup>۱</sup> - بسته‌ی واژگان) و دسته‌بندی آن‌ها دو رهیافت بسیار کارا وجود دارد:

۱- استفاده از روش‌های «تمایزی<sup>۲</sup>»

۲- استفاده از روش‌های «تولیدی<sup>۳</sup>»

از روش‌های تمایزی پرکاربرد برای مشخص کردن دسته‌ی تصویر می‌توان به SVM، Adaboost و شبکه‌های عصبی مصنوعی اشاره کرد (Serre et al 2005; Csurka et al 2004). استفاده از «کرنل انطباق هرمی<sup>۴</sup>» نتایج بسیار مطلوبی را روی پایگاه داده‌ی PASCAL تولید کرده است (Zhang et al 2007). کرنل انطباق هرمی روشی است که در آن ویژگی‌های در ابعاد بالا به هیستوگرام‌های چندبُعدی و چندوضوحی نگاشته می‌شوند. یکی از مهم‌ترین مزیت‌های هیستوگرام‌های چندبُعدی قابلیت ضبط اطلاعات مکانی ویژگی‌ها توسط آن‌ها است (Grauman and Yang et al 2009)

از مهم‌ترین روش‌های تولیدی می‌توان به روش‌های Naïve Bayes و «مدل‌های بیزی سلسله مراتبی<sup>۵</sup>» اشاره کرد (Blei et al 2003; Sivic et al 2005; Hoffman 1999). پس در حالت کلی و به طور خلاصه می‌توان برای پیاده‌سازی روش بسته‌ی واژگان پنج گام اصلی زیر را در نظر گرفت:

---

<sup>۱</sup>Histogram of words

<sup>۲</sup>Discriminative

<sup>۳</sup>Generative

<sup>۴</sup>Pyramid match kernel

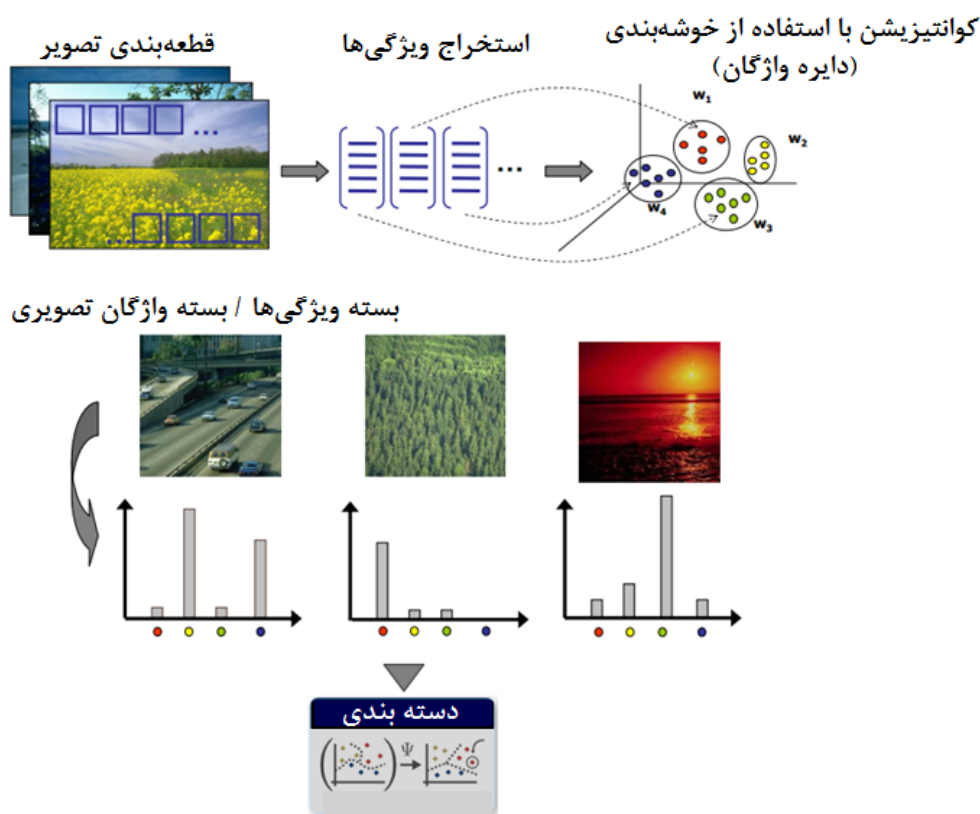
<sup>۵</sup>Hierarchical Bayesian model



- (۱) تقسیم تصویر به نواحی/قطعات
- (۲) توصیف قطعات تصویر (استخراج ویژگی‌ها)
- (۳) کوانتیزه کردن<sup>۱</sup> توصیف‌گرها برای ساختن دایره لغت با استفاده از روش‌های خوشه‌بندی
- (۴) شمارش تعداد وقوع هر لغت در هر تصویر و ساختن هیستوگرام واژگان (بسته واژگان - دایره واژگان)

(۵) آموزش دسته‌بندی‌های مختلف برای آموختن هیستوگرام لغت مربوط به هر دسته

شکل ۳-۲۵ این چهار گام را به صورت تصویری نشان می‌دهد.



شکل ۳-۲۵: مراحل تصویری پیاده‌سازی روش بسته واژگان تصویری

### ۳-۵. مسائل موجود در بازیابی تصاویر

در سال‌های اخیر مقالات متعددی در زمینه بازیابی تصاویر منتشر شده‌اند که از کارایی بسیار بالایی نیز برخوردار بوده‌اند (Alzu'bi, Amira and Ramzan 2015). با این وجود تحقیقات در این زمینه همچنان ادامه دارد و مسئله بازیابی تصاویر امروزه نیز از مباحث مهم و مسائل باز حوزه‌های پردازش

<sup>۱</sup>Quantization

تصویر، بینایی ماشین و بازیابی اطلاعات محسوب می‌شود.

مسائل مختلفی در زمینه بازیابی تصاویر وجود دارد. انسان‌ها قادر به شناسایی، دسته‌بندی و یافتن مشابهت‌های موجود آن هم به طور آنی و در یک لحظه هستند در حالی که این مسئله برای یک رایانه جزو مسائلی است که هم‌چنان حل نشده است. به عنوان مثال یک انسان به آسانی قادر به بازشناسی یک درخت چه از فاصله‌ای بسیار نزدیک و چه از فاصله‌ای بسیار دور است. همین درخت در فصل‌های مختلف سال ممکن است ظاهری متفاوت داشته باشد. برای مثال درختان به طور معمول در زمستان برگ‌ی ندارند، در پاییز رنگ برگ‌هایشان نارنجی و یا قهوه‌ای است، در تابستان سبز و میوه‌های آن‌ها در دسترس است. یک انسان تمامی این موارد و موقعیت‌ها را به راحتی بازشناسی می‌کند حال آن‌که فهم این موارد برای یک رایانه بسیار سخت و دشوار است.

در ادامه به برخی از مشکلات و چالش‌های موجود در طراحی و توسعه سیستم‌های بازیابی تصاویر، اشاره می‌شود.

## ۱. روشنایی<sup>۱</sup>

یکی از مهم‌ترین عواملی که باید در بازیابی تصاویر مشابه مدنظر قرار گرفته شود، تغییرات روشنایی است. به عنوان مثال شکل ۳-۲۲. الف تصاویر مشابهی از دسته معنایی « بزرگ راه » را نشان می‌دهند اگرچه تغییرات روشنایی در همه آن‌ها یکسان نیست. هم‌چنین شکل ۳-۲۶. ب نمونه‌ای دیگر از دسته تصاویر مشابه را نشان می‌دهد که روشنایی آن‌ها یکسان نیست.



شکل ۳-۲۶: تأثیر تغییرات روشناییالف تصاویر مشابه (بزرگ راه) با تغییرات روشنایی متفاوت ب. تصاویر

<sup>۱</sup> Illumination

### مشابه (سواحل) با تغییرات روشنایی متفاوت

یافتن چنین تصاویر مشابهی برای انسان‌ها راحت و بدیهی است ولی برای توسعهٔ یک سیستم بازیابی مقاوم باید بازشناسی اشیا و صحنه‌ها تحت روشنایی‌های متفاوت، مدنظر قرار گیرد. استفاده از توصیف‌گرهایی که در برابر تغییرات روشنایی تا حدی مقاوم هستند می‌تواند به کاهش تأثیر منفی تغییرات روشنایی در تصاویر کمک کند.

## ۲. تنوع درون دسته‌ای<sup>۱</sup>

تصاویر مشابهی که درون یک دستهٔ معنایی قرار گرفته‌اند از تنوع و گوناگونی بسیار برخوردارند. چگونگی در نظر گرفتن این تنوع خود از چالش‌های بزرگ است، به عنوان نمونه شکل ۳-۲۳. الف تصاویر مشابهی را که در دستهٔ معنایی «ساحل» قرار گرفته‌اند، نشان می‌دهد. این مطلب گویای این است که حتی تصاویر در دسته‌های معنایی خاص نیز از تنوع و گوناگونی برخوردار هستند. هم‌چنین شکل ۳-۲۷. ب تصاویر یک پرنده را در حالت‌های مختلف نشان می‌دهد.



شکل ۳-۲۷: چالش تنوع درون دسته‌ای. الف. تصاویر صحنه‌های ساحل با تنوع زیاد درون دسته‌ای ب. تصاویری از پرنده با تنوع زیاد درون دسته‌ای

## ۳. تنوع برون دسته‌ای<sup>۲</sup>

مشابه با مشکل تنوع درون دسته‌ای، چالش مهم دیگر تنوع برون دسته‌ای است. به این معنا که تصاویر مشابه در دسته‌های متفاوت باید بتوانند تا حد ممکن از یکدیگر متمایز شوند. برای مثال تصاویر مناظری مانند رودخانه و جنگل می‌توانند به هم شبیه باشند حال آن‌که از معنای متفاوتی برخوردار هستند (شکل

<sup>۱</sup> Intra-class variability

<sup>۲</sup> Inter-class variability

۲۸-۳. ب). هم‌چنین تصاویر مربوط به نوعی چنگ<sup>۱</sup> می‌تواند بسیار به تصاویر پیانو شبیه باشد. این تصاویر از نظر ظاهر و شکل بسیار به هم شبیه هستند (شکل ۳-۲۸. ب).

برای رفع مشکلات روشنایی، تنوع درون و برون دسته‌ای استفاده از توصیف‌گر سیفت می‌تواند راه حل مناسبی باشد (Liao, Liu and Hui 2013). همچنین همانطور که در فصل قبل ذکر شد استفاده از ترکیبی از ویژگی‌ها و دسته‌بندی‌های مناسب می‌تواند مشکلات مذکور را تعدیل کند.



الف. ۱. تصاویر جنگل

الف. ۲. تصاویر دشت



ب. ۱. تصاویر پیانو



ب. ۲. تصاویر نوعی چنگ

شکل ۳-۲۸: تنوع برون دسته‌ای. الف. دو تصویر جنگل ب. دو تصویر رودخانه

#### ۴. مسائل روش بسته‌واژگان تصویری

چنان‌که ذکر شد روش بسته‌واژگان تصویری از کاراترین روش‌ها در بازیابی تصاویر است و امروزه کمتر مقاله‌ای را می‌توان یافت که برای دسته‌بندی و بازیابی تصاویر از این روش بهره‌ای نبرده باشد. اما این روش دارای مشکلات و دشواری‌هایی نیز هست؛ با توجه به تحقیقات و مطالعاتی که انجام شد می‌توان مسائل اساسی این روش را چنین برشمرد (جدول ۳-۲):

#### جدول ۳-۲: مسائل موجود در روش بسته‌واژگان

| چالش                                                        | توضیح مختصر                             | چند روش نمونه پیشنهادی در منابع تحقیق | مزایا                                           | معایب                                                    | ملاحظه‌ها                                                | رهیافت رساله                                  |
|-------------------------------------------------------------|-----------------------------------------|---------------------------------------|-------------------------------------------------|----------------------------------------------------------|----------------------------------------------------------|-----------------------------------------------|
| از بین رفتن محتوای مکانی <sup>۲</sup> و معنایی <sup>۳</sup> | محتوای مکانی یعنی اجزای معنایی یک تصویر | (Lazebnik, Schmid and Ponce 2006)     | در نظر گرفتن اطلاعات مکانی و سادگی پیاده‌سازی و | اندازه یکسان قطعه‌های در نظر گرفته شده، عدم در نظر گرفتن | این روش هم‌اکنون نیز به عنوان یکی از مهم‌ترین و کاراترین | در این رساله روشی جدید برای ضبط اطلاعات مکانی |

<sup>۱</sup> harpsichord

<sup>۲</sup>Spatial Context

<sup>۳</sup>Semantic Context

|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|-------------------------------------------------------------|--------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| <p>(مثلاً شن، ماسه، کوه و دریا و...) چگونه کنار یکدیگر قرار گرفته‌اند و محتوای معنایی یعنی درک و تصور انسان از یک تصویر، در روش بسته واژگان چنین مواردی در نظر گرفته نمی‌شود.</p> | <p>سریع بودن</p>                                                              | <p>قطعات به صورت افقی و عمودی</p>                           | <p>روش‌ها برای ضبط اطلاعات مکانی در بسیاری از مقالات استفاده می‌شود.</p> | <p>تصاویر ارائه می‌شود. روش فوق با تقسیم‌بندی‌های همپوشانی‌کننده تصاویر در مقیاس‌های مختلف سعی در ضبط اطلاعات مکانی دارد</p> |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
| <p>یکی از مهم‌ترین مراحل روش بسته واژگان دسته‌بندی تصاویر است.</p>                                                                                                                | <p>دسته‌بند استفاده شده</p>                                                   | <p>دسته‌بندهای مبتنی بر SVM (Tsai, Chang and Liou 2014)</p> | <p>دقت مناسب و مدت زمان کوتاه آموزش، اجتناب از ورآموزی،</p>              | <p>انتخاب هسته مناسب تأثیر مستقیم در کارایی روش دارد، سرعت نسبتاً کم در مرحله آموزش</p>                                      |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
| <p>در این رساله شبکه‌های عصبی تابع پایه شعاعی جدیدی معرفی شده‌است که کارایی روش بسته واژگان را افزایش می‌دهد.</p>                                                                 | <p>دسته‌بندهای مبتنی بر شبکه‌های عصبی (Lu, Peng and Ip 2010)</p>              | <p>سادگی</p>                                                | <p>عدم کارایی مناسب برای داده‌هایی که نمونه‌های آموزشی کم است</p>        | <p>معمولاً دسته بندی‌کننده‌های SVM به طور گسترده‌ای در روش بسته واژگان استفاده می‌شوند.</p>                                  |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
| <p>پایداری در برابر نویز، قدرت عمومیت بخشی بالا، قابلیت بالا در حل مسائل غیر خطی پیچیده</p>                                                                                       | <p>دسته‌بندهای مبتنی بر شبکه‌های عصبی (Mukhopadhyay, Dash and Gupta 2014)</p> | <p>مدت زمان نسبتاً طولانی آموزش</p>                         | <p>مدت زمان نسبتاً طولانی آموزش</p>                                      | <p>مدت زمان نسبتاً طولانی آموزش</p>                                                                                          |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |
|                                                                                                                                                                                   |                                                                               |                                                             |                                                                          |                                                                                                                              |

|                                                    |                                                                                                                   |                                                                             |                                                                                                        |                                                    |  |
|----------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|----------------------------------------------------|--|
|                                                    |                                                                                                                   | قدرت بالا در یادگیری داده‌های با واریانس زیاد<br>قدرت بالا در برابر ورآموزی | دسته‌بندهای ترکیبی <sup>۱</sup><br>(Zhu, Chen and Zhao 2014 ; Song et al 2013 ; Marton et al 2013)     |                                                    |  |
| در این رساله از روش وزن‌دهی TF-IDF استفاده شده‌است | می‌توان با استفاده از روش‌های وزن‌دهی واژگان مشکل مذکور را تا حدی تعدیل کرد<br>( Tirilly, Claveau and Gros 2010 ) | تاکنون روش‌های خاصی برای تشخیص و حذف واژگان غیر مفید ارایه نشده‌است.        | نقاط کلیدی مشابه و واژگانی که حاوی هیچ اطلاعات مفیدی نیستند. موجب کاهش کارایی روش بسته واژگان می‌شوند. | نقاط کلیدی مشابه و واژگان تصویری بدون هیچ معنی خاص |  |

### ۶-۳. نتیجه‌گیری

در این فصل روش‌های مؤثر شناسایی و استخراج نقاط کلیدی تصویر بیان شدند و پرکاربردترین آن‌ها یعنی روش سیفت توضیح داده شد. همچنین توصیف‌گرهای مختلف تصویر مانند HOG و LBP که در حوزه‌های مختلفی نظیر پردازش تصویر، بینایی ماشینی و بازیابی تصاویر بسیار استفاده شده‌اند، مورد بررسی قرار گرفتند و چگونگی عملکرد آن‌ها توضیح داده شد.

در ادامه روش «بسته واژگان تصویری» به عنوان کاراترین روشی که امروزه از آن برای بازیابی معنایی تصاویر استفاده می‌شود، بررسی و توضیح داده شد. روش بسته واژگان تصویری روشی برگرفته از بازیابی متون است که در آن فرض بر این است که یک تصویر از تعدادی واژه که روی یک دایره واژگان تعریف شده‌اند؛ ساخته می‌شود و دو تصویر مشابه هستند اگر از واژه‌های یکسانی تشکیل شده باشند.

علاوه بر این چالش‌های حوزه بازیابی تصاویر بررسی و توضیح داده شد. از مهم‌ترین این چالش‌ها می‌توان به تغییرات روشنایی، تغییر مقیاس، تنوع‌های درون دسته‌ای و برون دسته‌ای اشاره کرد. از معایب مهم روش بسته واژگان می‌توان به از بین رفتن اطلاعات مکانی در حین ساخته شدن بردارهای ویژگی، دسته‌بند مورد استفاده و نیز وجود واژگان بی‌معنا و اضافی اشاره کرد. در فصل چهارم روش‌های پیشنهادی در رساله برای بهبود چالش‌های فوق بیان می‌شود.

<sup>۱</sup>Ensemble classifiers

## فصل چہارم

طراحی سامانۀ بازیابی تصاویر

## ۴-۱. مقدمه

همچنان که در فصل سوم اشاره شد روش بسته واژگان از مؤثرترین روش‌ها برای بازیابی و دسته‌بندی تصاویر است. در این فصل رهیافت‌های مختلفی برای بهبود کارایی روش بسته واژگان تصویری معرفی می‌شوند و سپس با استفاده از روش‌های پیشنهادی سامانه‌ای برای بازیابی تصاویر طراحی و پیاده‌سازی می‌شود.

مراحل مختلف روش بسته واژگان عبارتند از :

(۱) استخراج ویژگی‌ها

(۲) ساختن دایره واژگان

(۳) تشکیل بردارهای ویژگی با توجه به دایره واژگان ساخته شده

(۴) آموزش دسته‌بند برای یادگیری بردارهای ویژگی

در مرحله اول ویژگی‌های تصاویر پایگاه داده استخراج می‌شوند. هر چه این ویژگی‌های استخراج شده از قدرت تمایزدهی بالایی برخوردار باشند کارایی روش بسته واژگان افزایش می‌یابد. بنابر این یکی از مهم‌ترین عوامل در طراحی سامانه بازیابی تصاویر قدرت ویژگی‌های استخراج شده از تصاویر است. در ادامه روش جدیدی برای استخراج ویژگی‌های تصویر با استفاده از ویژگی‌های استخراج شده به وسیله الگوریتم سیفت ارایه می‌شود.

در مرحله دوم از روش بسته واژگان ویژگی‌های استخراج شده از تصاویر به وسیله الگوریتم‌های مختلف خوشه‌بندی مرتب و منظم شده و در خوشه‌های مختلف قرار داده می‌شوند. به این طریق دایره واژگان ساخته می‌شود. در واقع دایره واژگان بیانگر کل واژگانی است که تصاویر پایگاه داده از آن‌ها ساخته شده‌اند. روش خوشه‌بندی k-means در این مرحله از روش بسته واژگان به کارایی مناسبی منجر شده است (Wang and Huang 2015).



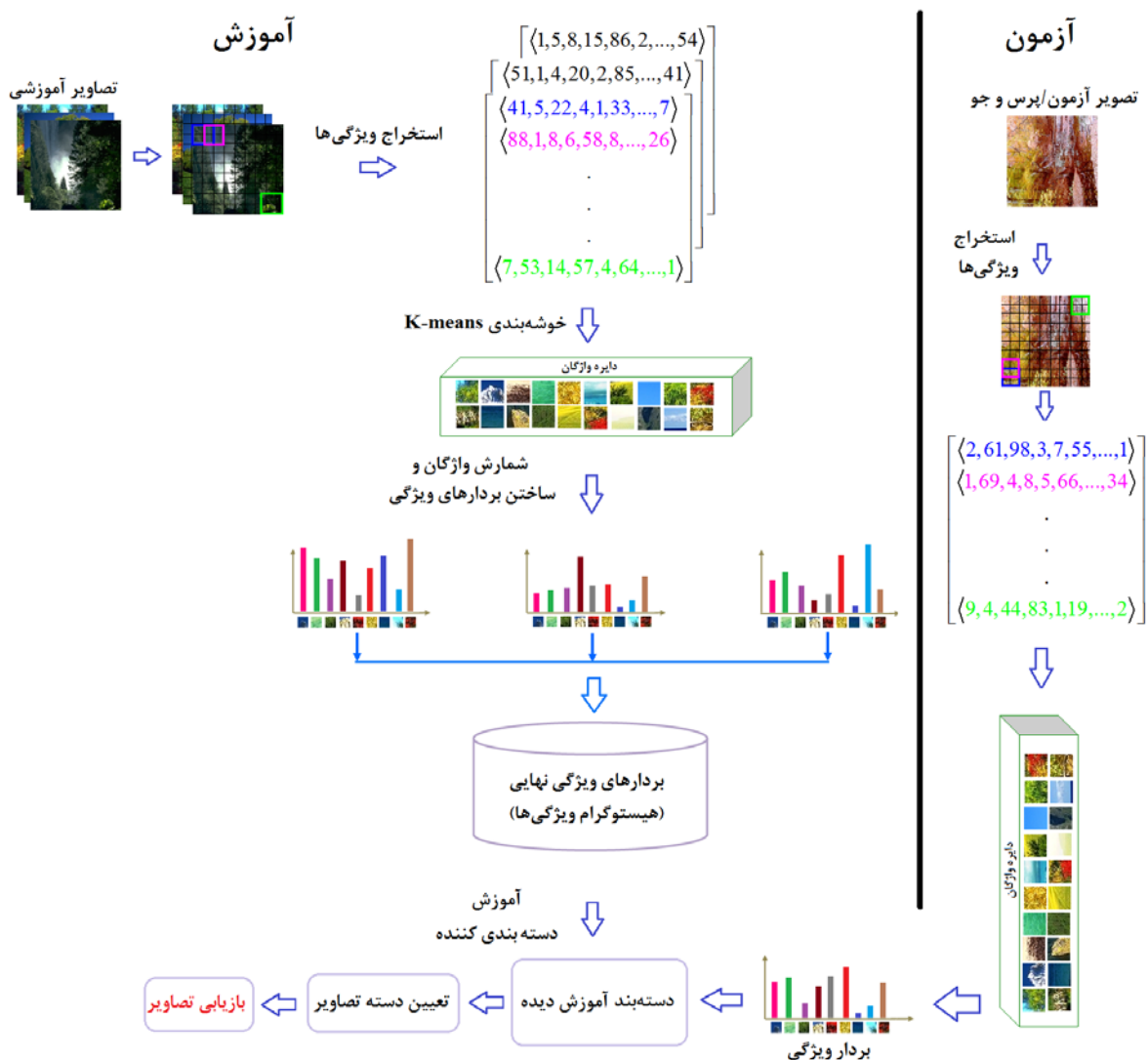
مرحله سوم از روش بسته واژگان تصویری ساخته شدن بردارهای ویژگی با توجه به دایره واژگان تشکیل شده در مرحله دوم است. طول بردار ویژگی برابر با تعداد خوشه‌ها در خوشه‌بندی انجام شده در مرحله دوم است. به عنوان مثال اگر تعداد خوشه‌ها در روش k-means برابر با ۲۵۶ در نظر گرفته شود، طول بردار ویژگی نیز ۲۵۶ خواهد بود.

مرحله آخر از روش بسته واژگان تصویری از مهم‌ترین مراحل روش بسته واژگان است که کارایی این روش را به شدت تحت تأثیر قرار می‌دهد. در این مرحله از یک دسته‌بندی‌کننده برای یادگرفتن ویژگی‌های استخراج شده از تصاویر استفاده می‌شود. در همین راستا و برای افزایش کارایی روش بسته واژگان یک دسته‌بندی‌کننده قدرتمند مبتنی بر شبکه‌های عصبی معرفی خواهد شد. چنانچه در فصل سوم ذکر شد، یکی از معایب بزرگ روش بسته واژگان از بین رفتن اطلاعات مکانی است. در این فصل روشی مؤثر برای در نظر گرفتن اطلاعات مکانی معرفی می‌شود.

## ۲-۴. مدل پیشنهادی برای بازیابی تصاویر

شکل ۴-۱ مدل پیشنهادی برای بازیابی تصاویر را نشان می‌دهد. چنانچه در این شکل ملاحظه می‌شود از روش بسته واژگان برای بازیابی معنایی تصاویر استفاده شده است. روش کار به این ترتیب است که با استفاده از تصاویری که در دسته‌های معنایی قرار دارند و روش بسته واژگان، یک دسته‌بند برای تعیین دسته تصاویر آموزش داده می‌شود. پس از آموزش دسته‌بند می‌توان از آن برای تعیین دسته تصاویر پرس‌وجو استفاده کرد. یعنی پس از آن که تصویر پرس‌وجو مشخص شد، دسته معنایی این تصویر به کمک دسته‌بندی که آموزش دیده شده است، مشخص می‌گردد. پس از تعیین دسته تصاویر پرس‌وجو شبیه‌ترین تصاویر به تصویر پرس‌وجو بازیابی می‌شوند.

به این ترتیب با استفاده از دسته‌بندی و روش‌های آموزش و یادگیری می‌توان به سطح معنایی بالایی در بازیابی تصاویر رسید. علاوه بر آن با استفاده از این روش فضای جست‌وجوی بسیار کوچک و محدود به جست‌وجو در دسته یا دسته‌هایی از تصاویر می‌شود (Alzu'bi, Amira and Ramzan 2015).



شکل ۴-۱: نمای کلی روش پیشنهادی برای بازیابی تصاویر

در ادامه هر یک از رهیافتهای پیشنهادی برای افزایش کارایی مدل ارایه شده در شکل ۴-۱ توضیح داده خواهند شد.

### ۳-۴. روش پیشنهادی برای استخراج ویژگی‌های تصاویر

هم‌چنان‌که اشاره شد امروزه اکثر روش‌های بازیابی تصاویر از ویژگی‌هایی مانند سیفت ، LBP ، HOG و ویژگی‌های استخراج شده به وسیله تبدیل موجک برای بازیابی تصاویر استفاده می‌کنند (Song and Li 2012; Liu et al 2014; Yildizer et al 2013).

بردار ویژگی‌های استخراج شده به وسیله تشخیص‌گر نقطه کلیدی سیفت از پرکاربردترین ویژگی‌های استخراج شده از تصاویر برای مقاصد مختلف نظیر بازشناسی و بازیابی اشیای مشابه به شمار می‌آیند

(Liao, Liu and Hui 2013) اما همانطور که در فصل قبل ذکر شد این ویژگی‌ها معایبی نیز دارند. در ادامه ویژگی‌های جدیدی مبتنی بر ویژگی‌های استخراج شده به وسیله تشخیص گر نقطه کلیدی سیفت معرفی می‌شوند.

### ۱-۳-۴. روش‌های پیشنهادی برای بازیابی تصاویر با استفاده از ویژگی‌های سیفت

با توجه به مطالعات انجام شده و انجام بررسی روش سیفت در فصل قبل، روش‌های مختلفی برای ارتقاء و استفاده از ویژگی‌های سیفت در بازیابی تصاویر طراحی و پیاده‌سازی شدند که در ادامه شرح مختصری از این روش‌ها ارائه می‌شود. در این روش‌ها همواره پارامترهای زمان و کارایی مد نظر بوده است. این روش‌ها در هر مرحله بهبود داده شده تا در نهایت توصیف‌گری مناسب مبتنی بر روش سیفت در حوزه CBIR به دست آید.

### الف. روش اول : خوشه‌بندی بردارهای ویژگی به ۱۲۸ خوشه

بردار ویژگی نقاط کلیدی در روش سیفت یک بردار ۱۲۸ تایی است که از ۱۶ قسمت ۸ تایی تشکیل شده است. این بردار نمایشی از اندازه ۱۲۸ برداری است که در اطراف نقطه کلیدی قرار گرفته‌اند. با تقسیم‌بندی این بردارها در ۱۲۸ خوشه سعی می‌شود بردارهای ذاتاً مشابه به یکدیگر را در یک خوشه قرار داد و تصاویر مشابه را بر همین اساس یافت. بدین ترتیب برای هر نقطه مهم یک خوشه در نظر گرفته می‌شود. بدین منظور تمام نقاط مربوط به هر خوشه در یک مجموعه قرار داده شد و از قراردادن اندازه این مجموعه‌ها در کنار هم یک بردار ویژگی عمومی با ۱۲۸ درایه ساخته شد.

فرض کنید تصویر  $A$  شامل مجموعه نقاط کلیدی  $P = \{p_1, p_2, p_3, \dots, p_k\}$  باشد و بردار ویژگی نقاط  $p_i$  با  $v_i = \{v_{i1}, v_{i2}, v_{i3}, \dots, v_{i128}\}$  نشان داده شود. بدین ترتیب برای تعریف یک بردار ویژگی عمومی، یک خوشه‌بندی با استفاده از روش  $k$ -means روی بردارهای ویژگی  $v_i$  با ۱۲۸ خوشه ساخته می‌شود. وظیفه خوشه‌بندی جمع‌آوری داده‌های همگون و نزدیک در یک خوشه است. حال نقاط کلیدی خوشه‌بندی می‌شوند و هر نقطه مهم به یک خوشه تعلق می‌گیرد. به ازای هر نقطه  $p$  گوییم  $K(p) = n$  هرگاه نقطه  $p$  با روش خوشه‌بندی تعریف شده عضو خوشه  $n$  ام باشد.  $K$  در حقیقت تابعی است که نتیجه خوشه‌بندی نقاط را نشان می‌دهد. برای ساختن بردار عمومی مجموعه‌های  $M_k, k = 1, 2, \dots, 128$  را به شکل زیر تعریف می‌کنیم :

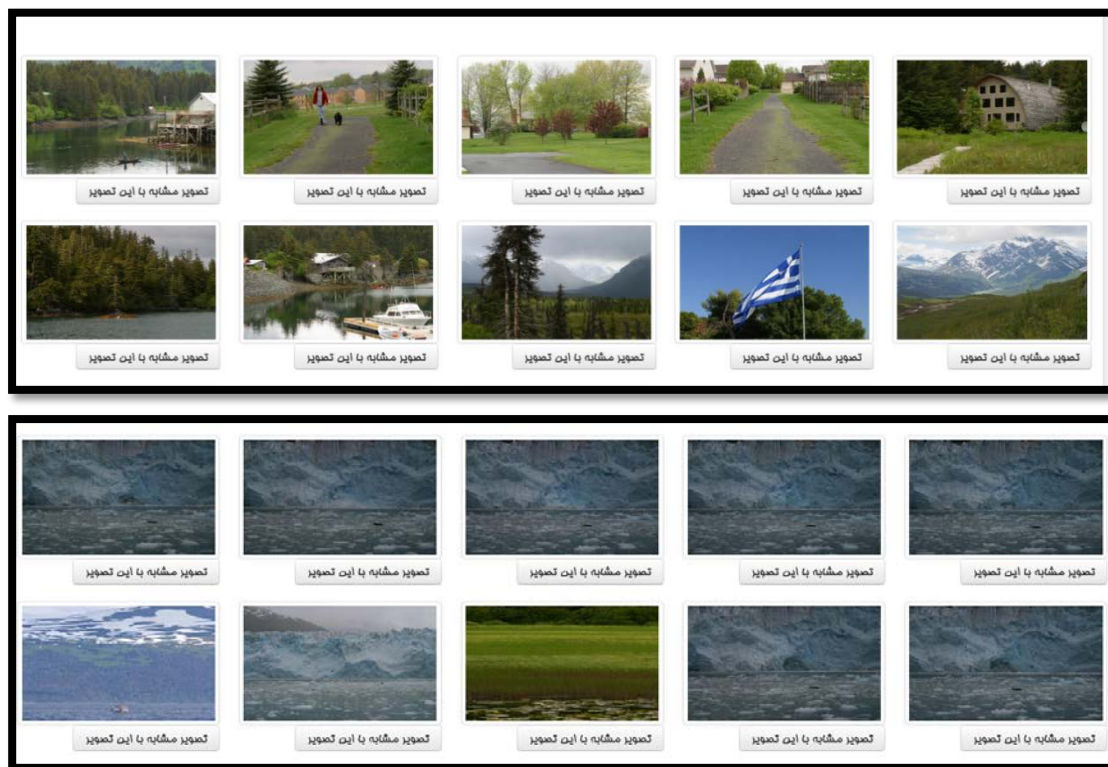
$$M_k = \{p \mid K(p) = k \text{ و } p \in P\} \quad (۴-۱)$$

بدین ترتیب بردار عمومی  $G_1$  برای هر تصویر  $A$  به شکل زیر تعریف می‌شود :

$$G_1 = [|M_1|, |M_2|, \dots, |M_{128}|] \quad (۴-۲)$$

که  $|M_i|$  تعداد اعضای مجموعه  $M_i$  است.

حال با استفاده از بردارهای ویژگی  $G_1$  برای هر تصویر مقیاس مناسبی برای اندازه گیری فواصل دو تصویر بدست می آید. در ادامه ۲ خروجی تشابه یابی تصاویر با استفاده از بردارهای ویژگی  $G_1$  نمایش داده شده است (شکل ۴-۲). هر خروجی شامل ۱۰ تصویر ابتدایی بازیابی شده است. تصویر پرس و جو در گوشه بالا سمت راست قرار گرفته است.

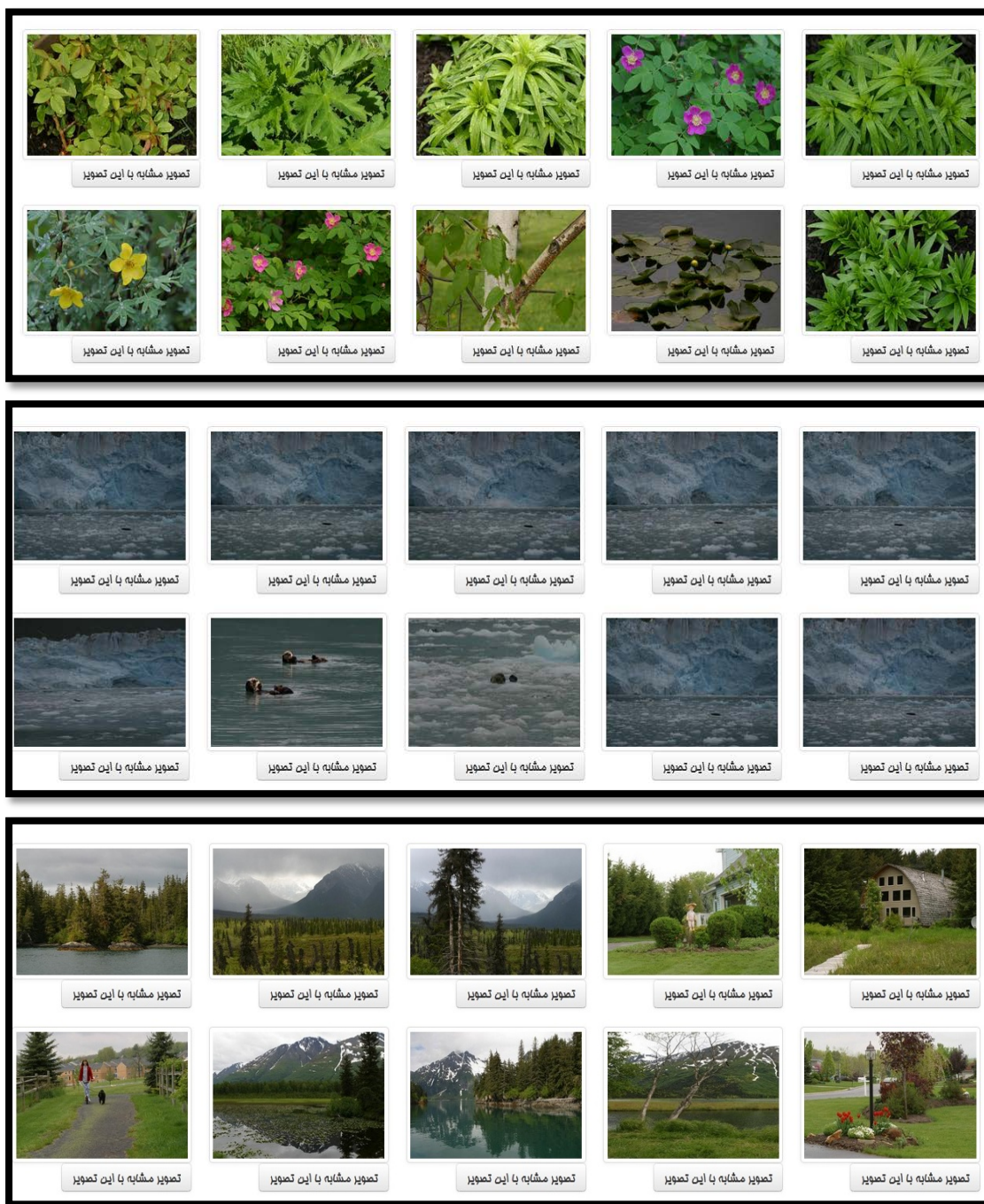


شکل ۴-۲: تشابه یابی با استفاده از روش اول ( برداری از تعداد اعضای خوشه ها)

همچنان که ملاحظه می شود نتایج جستجو بسیار امیدوار کننده است. حال با توجه به عدم استفاده از ویژگی رنگ در روش سیفت به نظر می رسد با افزودن ویژگی رنگ به عنوان معیار فرعی در صورت مسئله بتوان پاسخ های مناسب تری به دست آورد.

#### ب. روش دوم : خوشه بندی بردارهای ویژگی در فضاهای رنگی مختلف

ویژگی های استخراج شده با استفاده از الگوریتم سیفت با اعمال این الگوریتم روی تصویر خاکستری شده تصویر اصلی حاصل می شوند. در روش پیشنهادی دوم تصاویر رنگی به کانال های رنگی مختلف نظیر R، G و B، که به ترتیب مؤلفه های تصویر در کانال رنگی قرمز، سبز و آبی هستند، تجزیه می شوند و سپس روش پیشنهادی اول روی هر کانال رنگی به طور جداگانه اجرا می شود. بردار ویژگی نهایی از الحاق بردارهای ویژگی استخراج شده از هر یک از کانال ها به دست می آید و با  $G_2$  نشان داده می شود.



شکل ۴-۳: تشابه‌یابی با استفاده از روش دوم

نتایج این روش نسبت به روش اول مطلوب‌تر است که این ناشی از در نظر گرفتن اطلاعات در سه کانال رنگی متفاوت است.

ج. روش سوم : خوشه‌بندی عناصر بردارهای ویژگی با یکدیگر

این روش نیز مبتنی بر خوشه‌بندی عناصر بردار ویژگی نقاط کلیدی است. فرض کنید هر بردار ویژگی

خود از ۱۶ بردار ویژگی حاصل شده باشد. با خوشه‌بندی این مؤلفه‌های بردارهای ویژگی یک بردار جدید حاصل می‌شود. بدین ترتیب بردار  $v_j$  به صورت زیر به ۱۶ بردار تبدیل می‌شود:

$$v_j = \{v_{j1}, v_{j2}, v_{j3}, \dots, v_{j128}\} \rightarrow \{v'_{j1}, v'_{j2}, \dots, v'_{j16}\} \quad (۴-۳)$$

$$v_{j1}' = \{v_{j1}, v_{j2}, v_{j3}, \dots, v_{j8}\}, \dots, v_{j16}' = \{v_{j121}, v_{j122}, v_{j123}, \dots, v_{j128}\} \quad (۴-۴)$$

حال خوشه‌بندی  $K$ ، با هشت خوشه با استفاده از روش  $k$ -means ساخته می‌شود. بدین ترتیب  $K(v_{jl}') = n$  اگر و فقط اگر  $v_{jl}'$  عضو خوشه  $n$  ام باشد.

در نتیجه ماتریس عمومی  $G_3$  برای هر تصویر  $A$  به شکل زیر تعریف می‌شود:

$$G_3 = \begin{bmatrix} |M'_{1,1}| & \dots & |M'_{1,16}| \\ \vdots & \ddots & \vdots \\ |M'_{8,1}| & \dots & |M'_{8,16}| \end{bmatrix}_{8 \times 16} \quad (۴-۵)$$

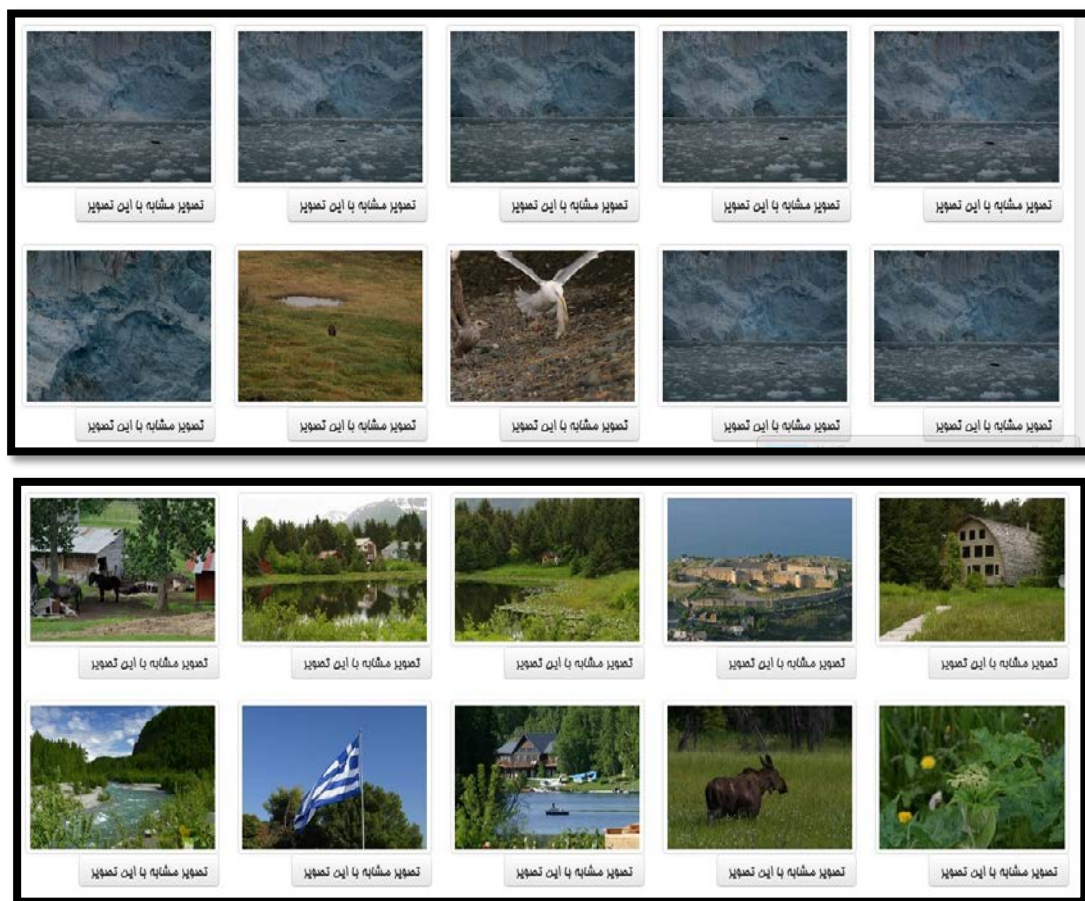
که در آن:

$$M'_{ij} = \{v'_{lj} \mid K(v'_{lj}) = i, v_l \in P\} \quad (۴-۶)$$

و  $|M'_{ij}|$  تعداد اعضای مجموعه  $M'_{ij}$  است.

در ادامه نتایج استفاده از روش مذکور در شکل ۴-۴ نمایش داده شده است.



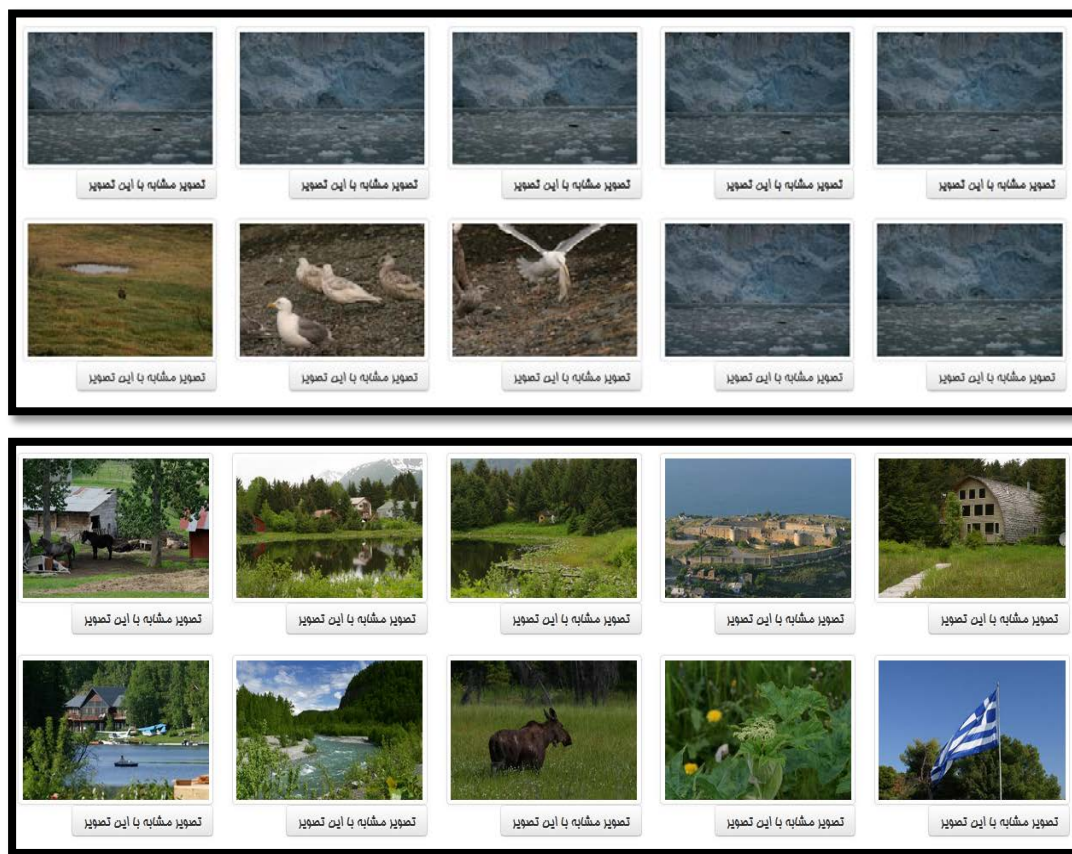


شکل ۴-۴: تشابه‌یابی با استفاده از روش سوم (ماتریس از تعداد اعضای خوشه‌ها)

#### د. روش چهارم: خوشه‌بندی عناصر بردارهای ویژگی در فضاهای رنگی مختلف

باتوجه به آنچه در روش دوم اشاره شد، می‌توان باتجزیه تصویر به کانال‌های رنگی متفاوت ویژگی‌ها را از هر یک از کانال‌ها استخراج کرد. بردار ویژگی حاصل از اعمال روش سوم بر مؤلفه‌های رنگی  $R$ ،  $G$  و  $B$ ،  $G_4$  نامیده می‌شود.

در ادامه نتیجه استفاده از این روش در تشابه‌یابی نشان داده می‌شود.



شکل ۴-۵: تشابه‌یابی با استفاده از روش چهارم

#### هـ روش پنجم: فشرده‌کردن و خوشه‌بندی بردارهای ویژگی

خوشه‌بندی بهترین راه حل برای تبدیل بردارهای محلی به بردارهای عمومی‌تر در روش سیفت به نظر می‌رسد. در روش پنجم بردارهای ویژگی را فشرده و سپس آن‌ها را خوشه‌بندی می‌کنیم. این روش را «سیفت فشرده» می‌نامیم.

برای فشرده‌کردن بردارهای محلی استخراج شده توسط روش سیفت از معنا و مفهوم روش سیفت استفاده می‌شود. به این ترتیب که برآیند هر یک از اندازه‌های بردارهای هم‌سو محاسبه و برداری جدید ساخته می‌شود. طریقه ساخت این بردار به صورت زیر است:

$$v_j = \{v_{j1}, v_{j2}, v_{j3}, \dots, v_{j128}\} \rightarrow \quad (4-7)$$

$$\bar{v}_j = \left\{ \sum_{l=0}^{15} v_{j(l*8+1)}, \sum_{l=0}^{15} v_{j(l*8+2)}, \dots, \sum_{l=0}^{15} v_{j(l*8+8)} \right\} \quad (4-8)$$

بردارهای جدید  $\bar{v}_j$  در عین محلی بودن به دلیل اینکه از عمل برآیندگیری و فشرده‌سازی حاصل شده‌اند، اطلاعات سراسری بیشتری در مقایسه با بردارهای ویژگی  $v_j$  در خود دارند. بنابر این در برابر



تغییرات مختلف تصویر پایدار هستند.

با ادامه روند مذکور می‌توان مجدداً از خوشه‌بندی برای کاهش حجم اطلاعات استفاده کرد و اطلاعاتی از یک دست را در یک خوشه قرار داد. بنابراین بردارهای حاصل را با ۱۶ خوشه به وسیله روش  $k$ -means، خوشه‌بندی می‌کنیم. بدین ترتیب  $K(\bar{v}_j) = n$  اگر و فقط اگر  $\bar{v}_j$  عضو خوشه  $n$  ام باشد.

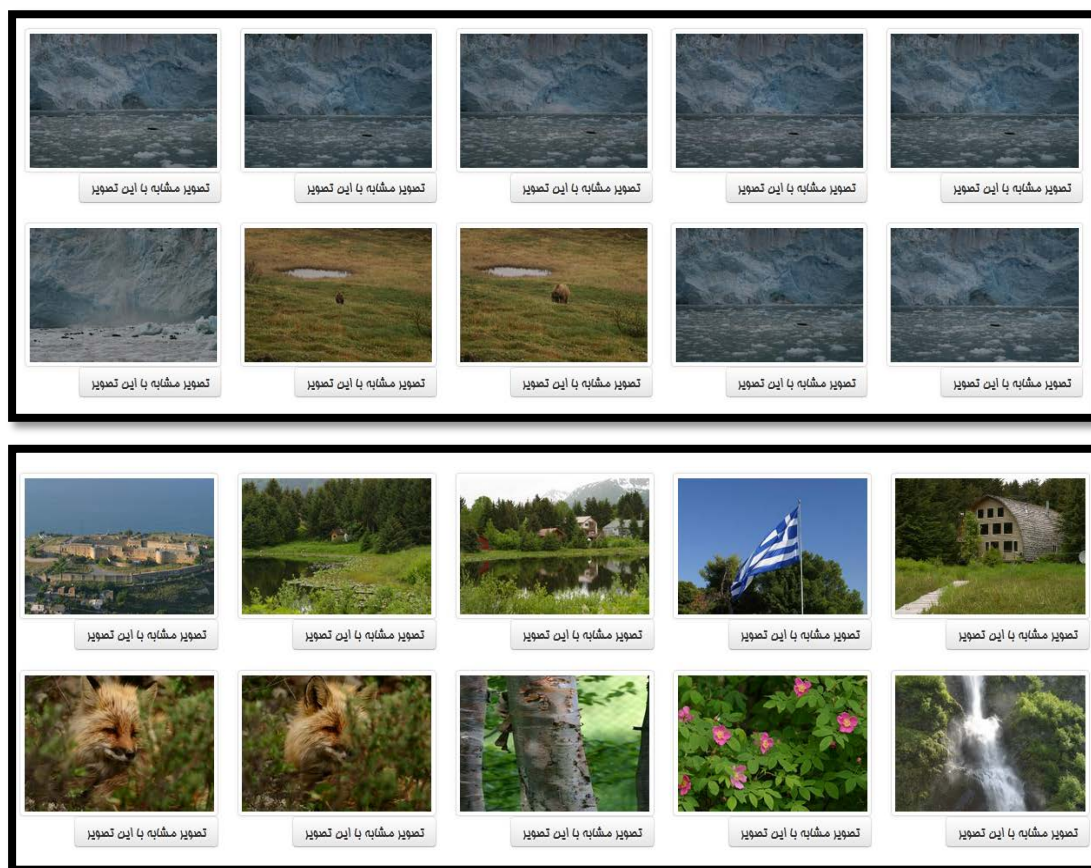
بدین ترتیب بردار عمومی  $G_5$  برای تصویر دلخواه  $A$  به شکل زیر تعریف می‌شود:

$$G_5 = [|\bar{M}_1|, |\bar{M}_2|, \dots, |\bar{M}_{16}|] \quad (۴-۹)$$

که در آن

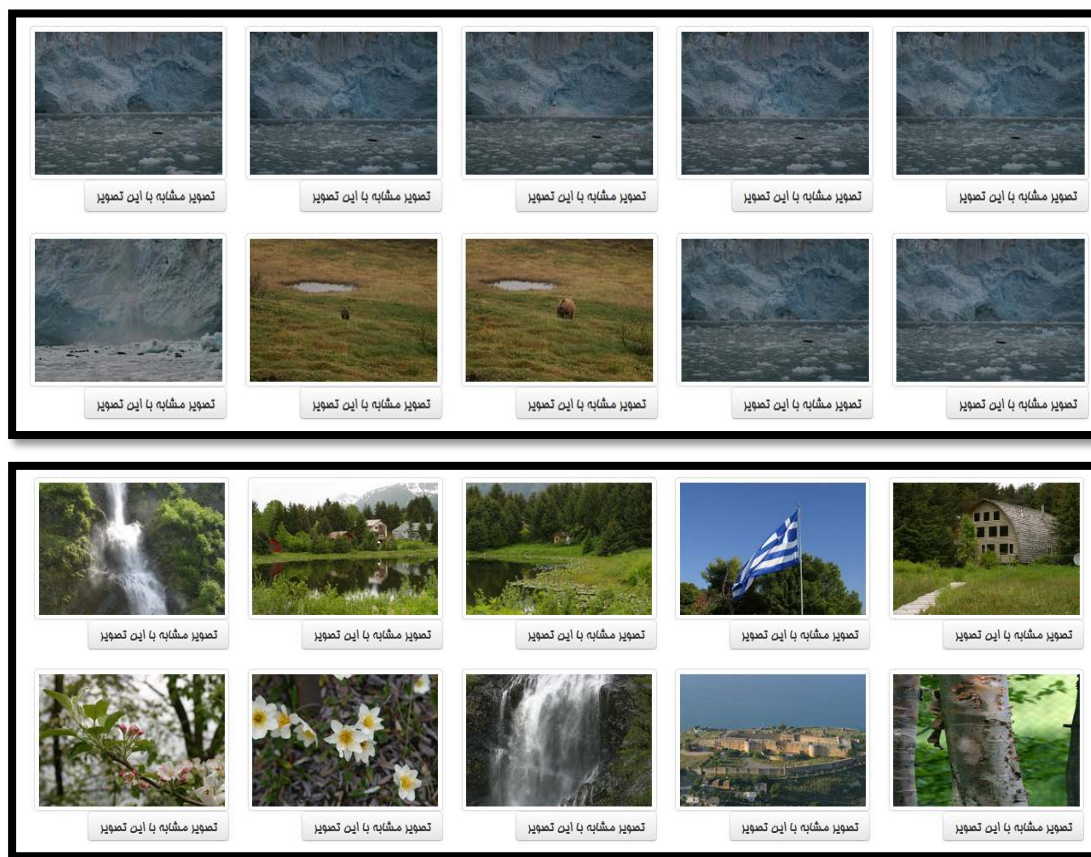
$$\bar{M}_k = \{p \mid K(\bar{v}_j) = k \text{ و } p \in P\} \quad (۴-۱۰)$$

نتیجه اجرای تشابه‌یابی با استفاده از روش چهارم در شکل ۴-۶ نمایش داده شده است.



شکل ۴-۶: تشابه‌یابی با استفاده از روش پنجم (سیفت فشرده‌شده)

و. روش ششم: فشرده‌کردن و خوشه‌بندی بردارهای ویژگی در فضاهای رنگی مختلف  
همانند آنچه در روش‌های دوم و چهارم ذکر شد، روش ششم نسخه رنگی روش پنجم است و «سیفت رنگی فشرده» نامیده می‌شود. در ادامه نتایج حاصل از ترکیب این دو بردار با هم نمایش داده می‌شود.

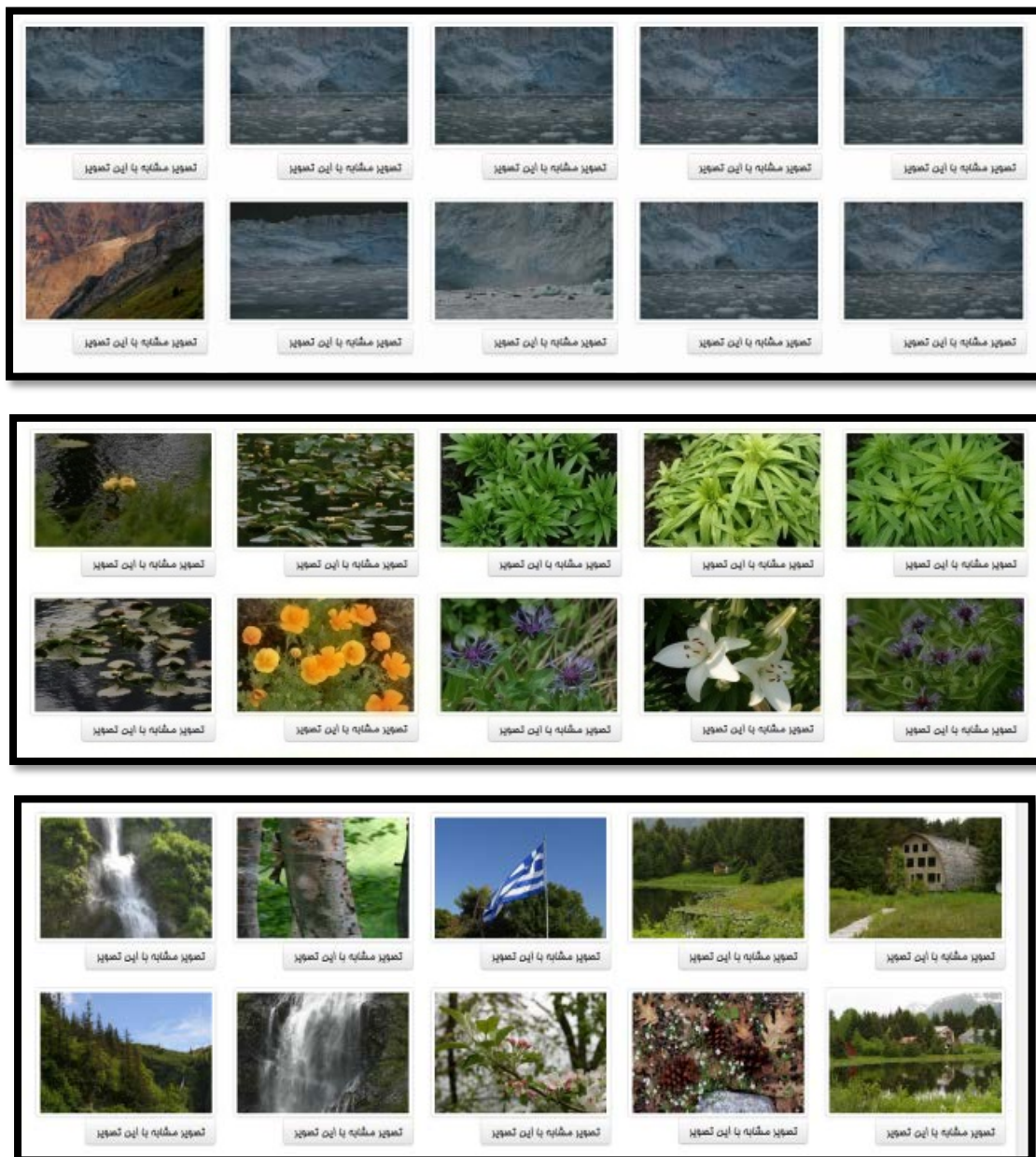


شکل ۴-۷: تشابه‌یابی با استفاده از روش ششم (سیفت رنگی فشرده‌شده)

#### ز. روش هفتم : تغییر تعداد خوشه‌ها در روش پنجم

روش‌های پنجم و ششم، کارایی مناسبی از خود نشان می‌دهند. چنین به نظر می‌رسد که با تغییر شدن تعداد خوشه‌ها کارایی روش‌ها تغییر کند. در ادامه روش پنجم با تعداد خوشه‌های بیشتر اجرا می‌شود. نتیجه این آزمایش‌ها در شکل‌های زیر نمایش داده شده است.

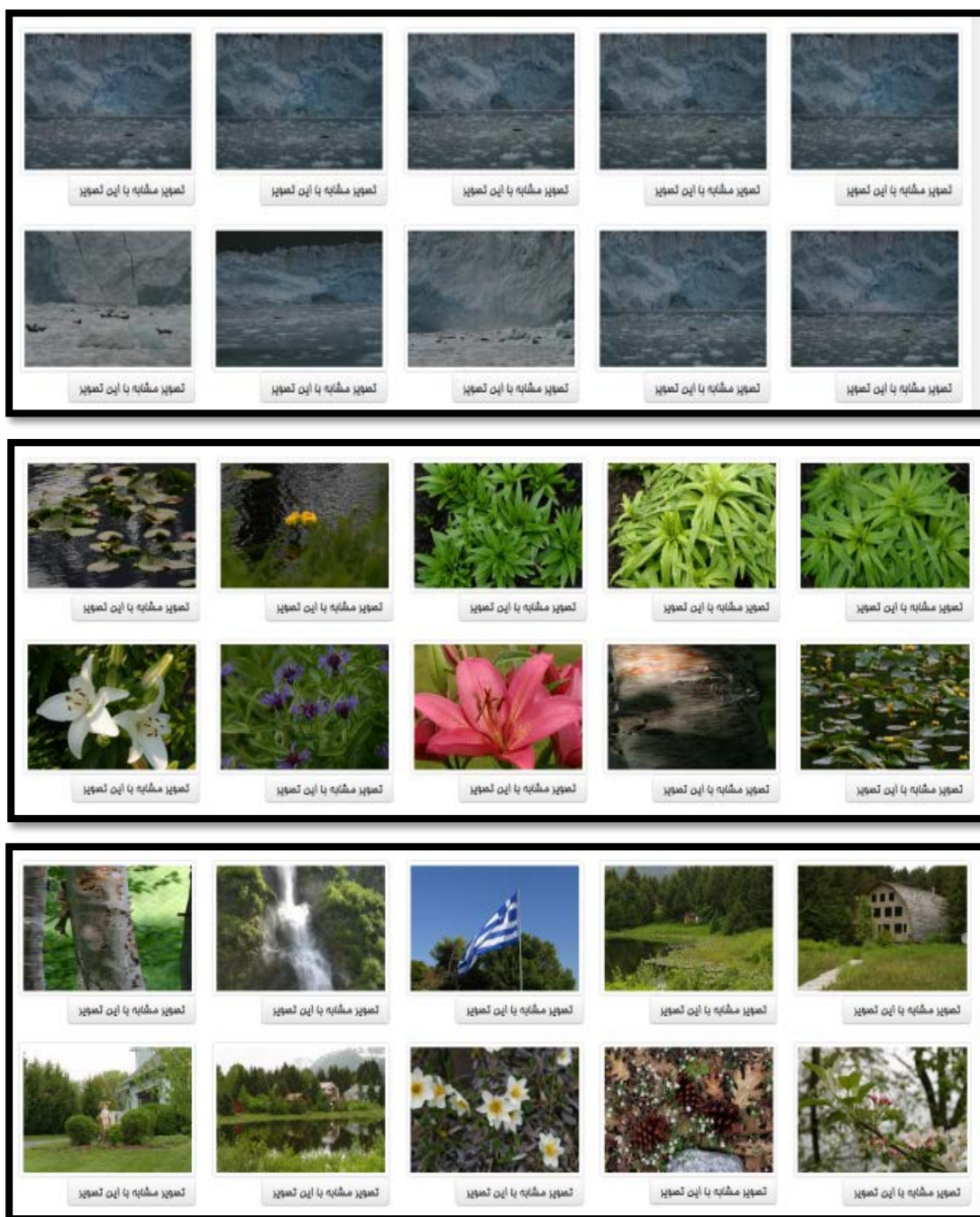
(۱) استفاده از ۳۲ خوشه



شکل ۴-۸: تشابه‌یابی با استفاده از روش پنجم و تعداد خوشه‌ها برابر ۳۲

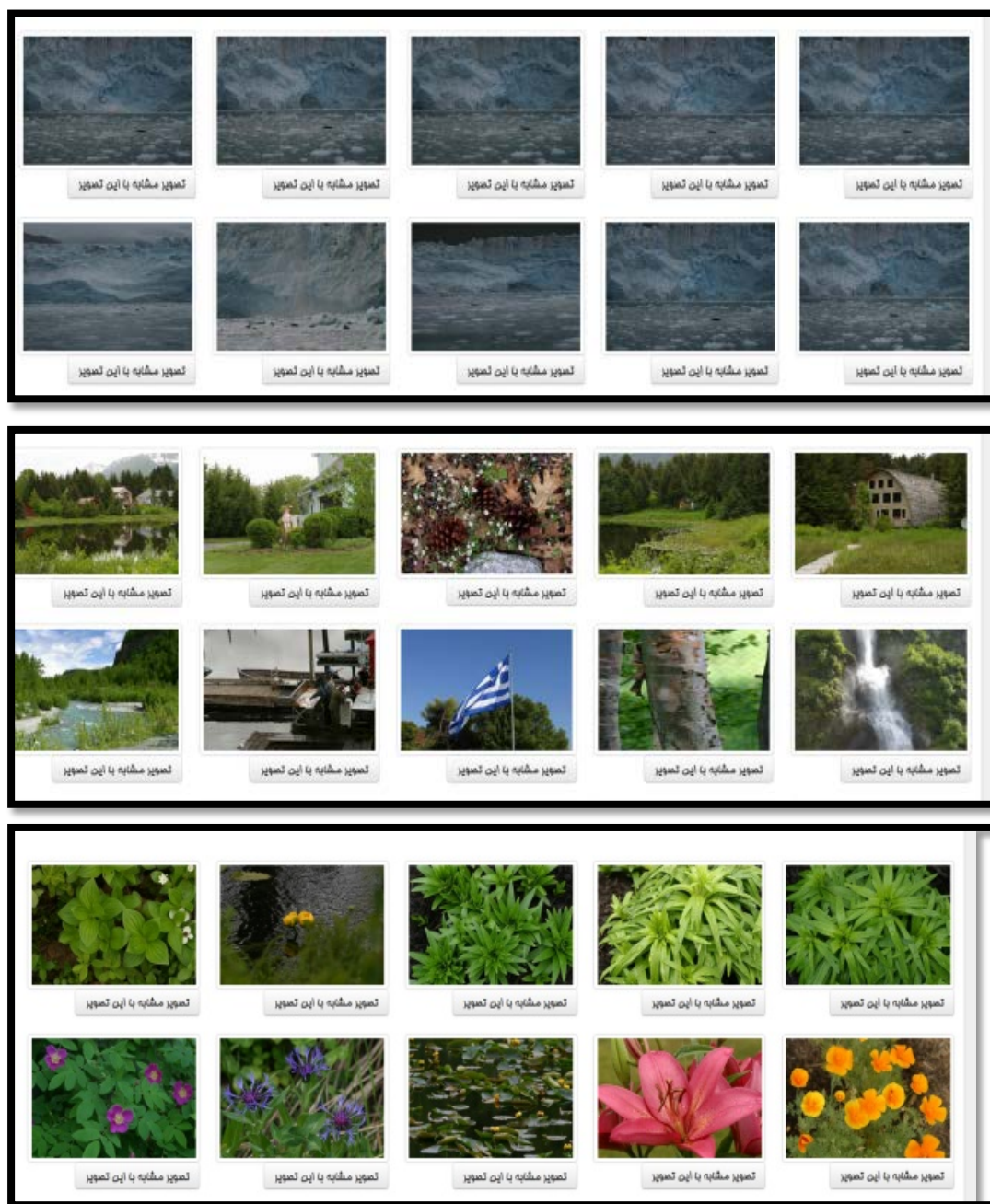


## ۲) استفاده از ۶۴ خوشه



شکل ۴-۹: تشابه‌یابی با استفاده از روش پنجم و تعداد خوشه‌ها برابر ۶۴

### ۳) استفاده از ۱۲۸ خوشه



شکل ۴-۱۰: تشابه‌یابی با استفاده از روش پنجم و تعداد خوشه‌ها برابر ۱۲۸

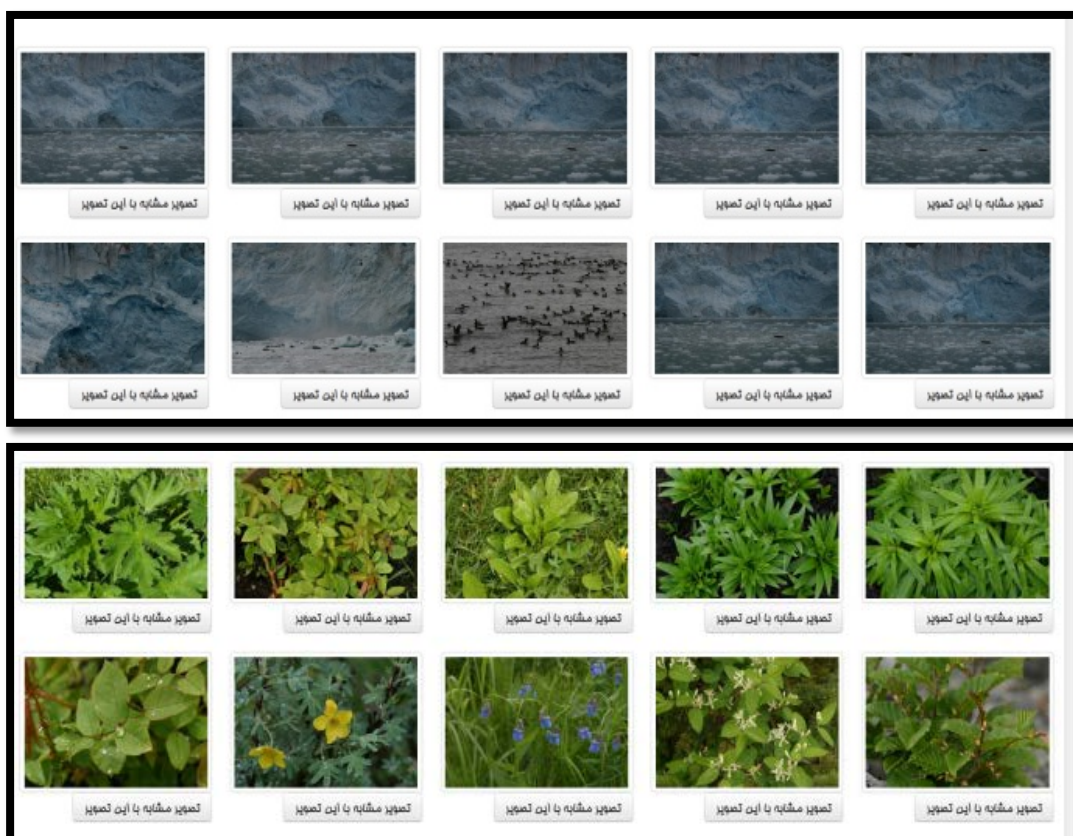
#### ح. روش هشتم : تقسیم‌بندی تصویر و استفاده از روش پنجم

یکی دیگر از نکات کلیدی درباره روش سیفت مکان قرارگیری نقاط است. برای استفاده از این نکته تصاویر به چند قسمت تقسیم می‌شوند. از آنجا که در تصاویر مشابه قسمت‌های نظیر به نظیر نیز مشابه هستند، پس بردارهای ویژگی این قسمت‌ها نیز باید به هم نزدیک باشند. از این رو می‌توان با تقسیم

تصاویر به  $n$  قسمت و اجرای روش پنجم روی هر قسمت،  $n$  بردار شانزده تایی به دست آورد. حال با مقایسه نظیر به نظیر این قسمت‌ها در هر دو تصویر می‌توان تشابه آن‌ها را اندازه‌گیری کرد. این معیاری مناسب برای تشابه‌یابی نقاط در تصاویر، مبتنی بر روش سیفت و بهره جویی از ویژگی‌های نقاط کلیدی است. در ادامه نتایج حاصل از اجرای این روش را، زمانی که تصاویر به ۴ و ۱۶ قسمت مساوی تقسیم شده باشند، نمایش داده می‌شود.

### ۱- تقسیم تصاویر به چهار قسمت

در این روش ابتدا تصویر به چهار قسمت مساوی تقسیم می‌شود و سپس روش پنجم روی هر قسمت اعمال می‌شود. با تقسیم‌بندی تصویر و استخراج اطلاعات از هر قسمت می‌توان گفت ویژگی‌های استخراج شده از ویژگی‌های سراسری به ویژگی‌های محلی تغییر می‌کنند. این امر باعث متمایزتر شدن ویژگی‌های استخراج شده و در نتیجه افزایش کارایی آن‌ها می‌شود. در شکل ۴-۱۱ نتایج حاصل از تقسیم تصویر به چهار قسمت ملاحظه می‌شود. بدیهی است که نتایج در مقایسه با روش پنجم بهبود یافته‌اند.







شکل ۴-۱۱: تشابه‌یابی با استفاده از روش پنجم و تقسیم تصویر به ۴ قسمت

## ۲- تقسیم تصاویر به ۱۶ قسمت

در این آزمایش نیز همانند آزمایش قبلی هر تصویر به قطعاتی تقسیم می‌شود. ملاحظه می‌شود که با افزایش تعداد قطعات کارایی روش پیشنهادی نیز افزایش می‌یابد. دلیل این امر آن است که ویژگی‌های استخراج شده متمایزتر می‌شوند. شکل ۴-۱۲ نتایج حاصل از این تقسیم‌بندی را نشان می‌دهد.





شکل ۴-۱۲: تشابه‌یابی با استفاده از روش پنجم و تقسیم تصویر به ۱۶ قسمت

#### ۴-۴. تعمیم روش پیشنهادی

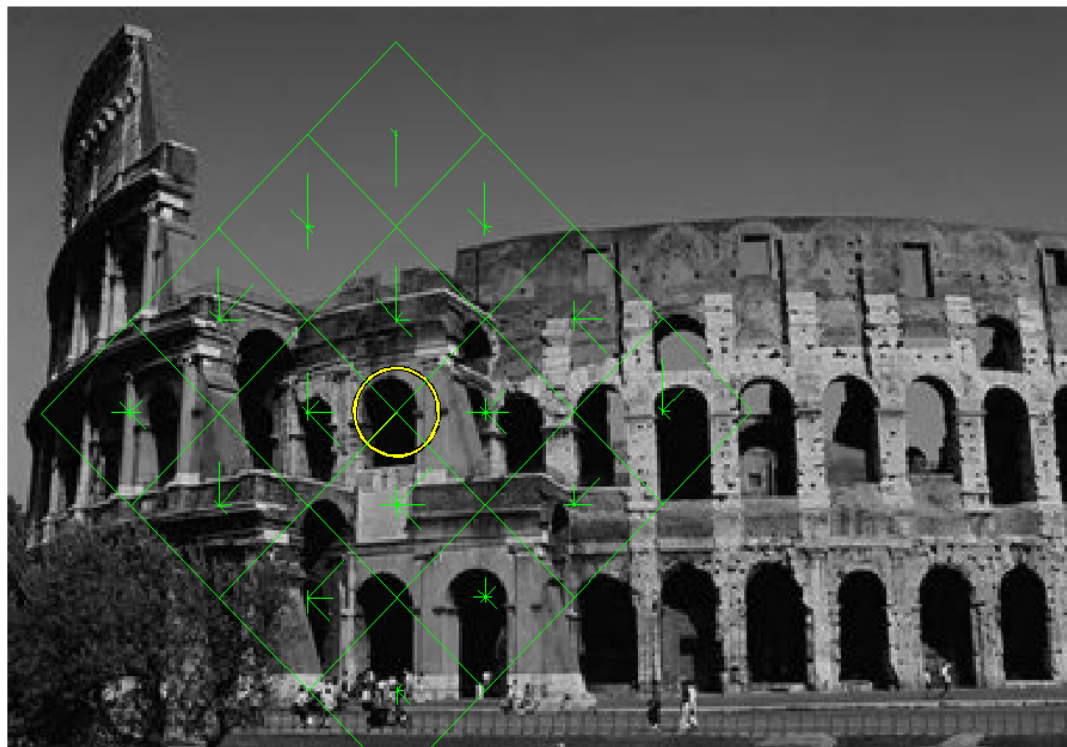
هم‌چنان‌که در ابتدای فصل ذکر شد در این رساله برای بازیابی تصاویر از روش بسته‌ واژگان استفاده می‌شود. در این قسمت دو پیشنهاد مختلف برای تعمیم روش بیان شده در قسمت قبلی ارائه می‌شود و سپس از آن در قالب روش بسته‌ واژگان استفاده می‌شود.

در ابتدا برای تعمیم توصیف‌گر پیشنهادی (سیفتِ فشرده)، این توصیف‌گر روی همهٔ پیکسل‌های تصویر اعمال و کارایی آن در جهت‌های مختلف بررسی می‌شود. همچنین برای بهبود روش بسته‌ واژگان و در نظر گرفتن اطلاعات مکانی، رهیافت جدیدی ارائه می‌شود.

##### ۴-۴-۱. بهبود توصیف‌گر سیفتِ فشرده

برای بهبود توصیف‌گر سیفتِ فشرده و استفاده از آن در قالب مذکور در شکل ۴-۱، این توصیف‌گر روی تمامی پیکسل‌های تصویر و در جهت‌های مختلف اعمال می‌شود. قابل ذکر اینکه توصیف‌گر پیشنهادی در بخش قبلی بردار گرادیان تصویر در نقاط کلیدی را در هشت جهت محاسبه می‌کرد. شکل ۴-۱۳ بردارهای گرادیان محاسبه شده در هشت جهت و در یک همسایگی  $4 \times 4$  از یک پیکسل دلخواه را نشان می‌دهد.





شکل ۴-۱۳: بردارهای گرادیان محاسبه شده در هشت جهت در نقطه‌ای دلخواه از تصویر

برای افزایش کارایی سیفِ فشرده به نظر می‌رسد با محاسبه بردارهای گرادیان در جهت‌های بیشتر بتوان اطلاعات بیشتری از تصاویر استخراج کرد و به این ترتیب قابلیت تمایزدهی توصیف‌گر پیشنهادی را افزایش داد. به همین دلیل و برای یافتن پارامترهای بهینه روش پیشنهادی، با استفاده از پایگاه‌ونگ گرادیان تصویر در جهت‌های مختلف محاسبه شد. به همین منظور از هر دسته از تصاویر ۲۰ تصویر برای آموزش و ۱۰ تصویر برای آزمون استفاده شد. جدول ۴-۱ نتایج این آزمایش‌ها را نشان می‌دهد.

جدول ۴-۱: نتایج محاسبه بردارهای گرادیان در جهت‌های مختلف

| طول بردار ویژگی | تعداد بین‌ها | تعداد جهت‌ها | کارایی | زمان (ثانیه) |
|-----------------|--------------|--------------|--------|--------------|
| ۳۰۰             | ۱۶           | ۱۶           | ۷۷٪    | ۲۳۵          |
| ۳۰۰             | ۱۶           | ۳۶           | ۷۸٪    | ۳۴۸          |
| ۳۰۰             | ۳۶           | ۳۶           | ۸۴٪/۵  | ۳۵۱/۷۶       |
| ۳۰۰             | ۳۶           | ۶۴           | ۷۷٪    | ۵۰۹,۹        |
| ۱۰۰۰            | ۱۶           | ۱۶           | ۷۴٪    | ۳۱۵          |
| ۱۰۰۰            | ۱۶           | ۳۶           | ۷۰٪    | ۴۴۸/۱۶       |
| ۱۰۰۰            | ۱۶           | ۶۴           | ۶۸٪    | ۶۱۵/۸        |
| ۱۰۰۰            | ۳۶           | ۳۶           | ۸۰٪    | ۴۳۷          |
| ۱۰۰۰            | ۳۶           | ۶۴           | ۸۱٪    | ۶۰۳          |
| ۱۰۰۰            | ۶۴           | ۶۴           | ۸۱٪    | ۶۲۲/۲۲       |

این جدول نشان می‌دهد با تغییر تعداد جهتهایی که گرادیان تصویر در یک نقطه محاسبه می‌شود، می‌توان قدرت تمایز توصیف‌گر پیشنهادی را افزایش داد. لازم به ذکر است در جدول مذکور منظور از «تعداد بین‌ها» تعداد جهتهای در نظر گرفته شده برای کوانتیزه کردن<sup>۱</sup> جهتهایی است که گرادیان تصویر در آن‌ها محاسبه شده‌است. برای مثال اگر تعداد جهتها برابر با ۳۶ و تعداد بین‌ها برابر با ۱۶ باشد، بردارهای گرادیان در جهتهای  $0^\circ, 10^\circ, 20^\circ, \dots, 350^\circ$  محاسبه خواهند شد و این جهتها در هیستوگرامی با ۱۶ بین قرار خواهند گرفت. این ۱۶ بین عبارتند از:  $0^\circ$  تا  $22/5^\circ$ ،  $22/5^\circ$  تا  $45^\circ$  ... و  $337/50^\circ$  تا  $360^\circ$ .

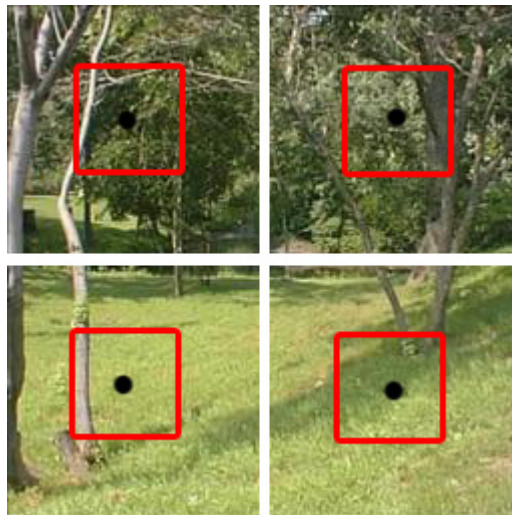
آزمایش‌های انجام شده نشان می‌دهد با افزایش تعداد جهتها و نیز افزایش طول بردارهای ویژگی، مدت زمان اجرای روش بسته واژگان با استفاده از توصیف‌گر پیشنهادی، افزایش می‌یابد که این امر طبیعی است. با توجه به جدول فوق می‌توان نتیجه‌گیری کرد که با انتخاب تعداد جهت و تعداد بین به صورت (۱۶، ۱۶) و (۳۶، ۳۶) می‌توان به بازده قابل قبولی رسید.

## ۲-۴-۴. استخراج اطلاعات مکانی

هم‌چنان‌که در فصل سوم ذکر شد، یکی از مهم‌ترین معایب روش بسته واژگان تصویری از بین رفتن اطلاعات مکانی است. در واقع روش بسته واژگان قادر به تشخیص شکل و تفکیک اشیا از پس‌زمینه تصویر نیست. غلبه بر چنین محدودیت‌هایی برای ساختن یک مدل قدرتمند که قادر به ضبط اطلاعات چیدمانی تصویر باشد از مسائل بسیار سخت در پردازش سطح بالای تصویر به شمار می‌رود. این مشکل به خصوص در طراحی سیستم‌هایی که با تصاویری با پس‌زمینه‌های شلوغ سروکار دارند و یا از اشیایی تشکیل شده‌اند که با یکدیگر همپوشانی دارند، بسیار جلوه می‌کند. برای حل این مشکل در این رساله روش جدیدی معرفی می‌شود که با ضبط اطلاعات مکانی به طور مؤثری باعث افزایش کارایی روش بسته واژگان می‌شود.

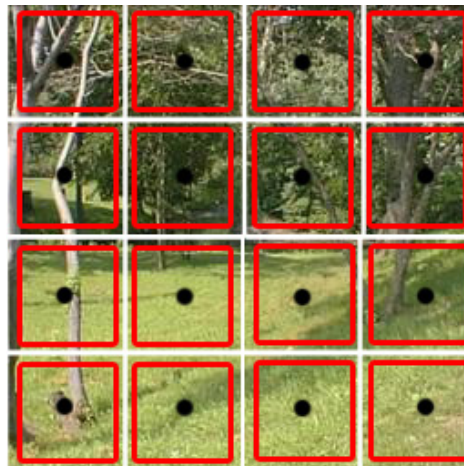
در روش پیشنهادی برای ضبط اطلاعات مکانی از تقسیم‌بندی‌های متفاوت تصویر در مقیاس‌های مختلف استفاده می‌شود. روش کار به این ترتیب است که تصویر در مقیاس‌های مختلف بلوک‌بندی می‌شود. پس از بلوک‌بندی تصویر، در مرکز هر بلوک پنجره‌ای با اندازه  $ws \times ws$  قرار داده می‌شود و ویژگی‌های هر بلوک با توجه به پنجره مذکور استخراج می‌شوند. شکل ۴-۱۴ تصویر  $I$  و تقسیم آن به چهار بلوک و پنجره‌هایی که در بلوک‌های مختلف گذاشته شده است را نشان می‌دهد.

<sup>۱</sup>Quantization

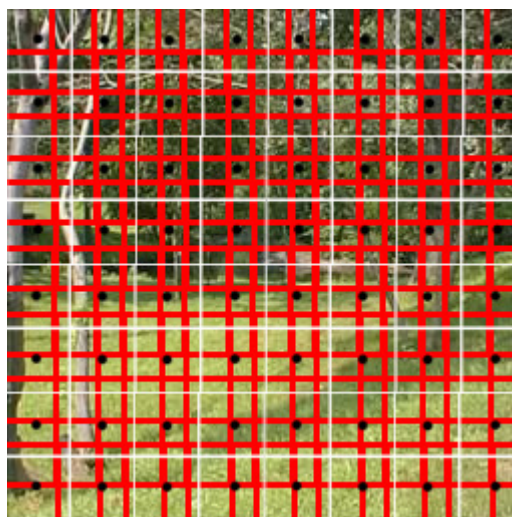


شکل ۴-۱۴: تصویری دلخواه و تقسیم آن به چهار بلوک

شکل ۴-۱۵ نمونه دیگری از این تقسیم‌بندی را نشان می‌دهد. در این شکل، تصویر  $I$  با استفاده از یک تقسیم‌بندی  $4 \times 4$  به ۱۶ بلوک تقسیم شده است. شکل ۴-۱۶ نیز یک بلوک‌بندی  $8 \times 8$  را نشان می‌دهد.

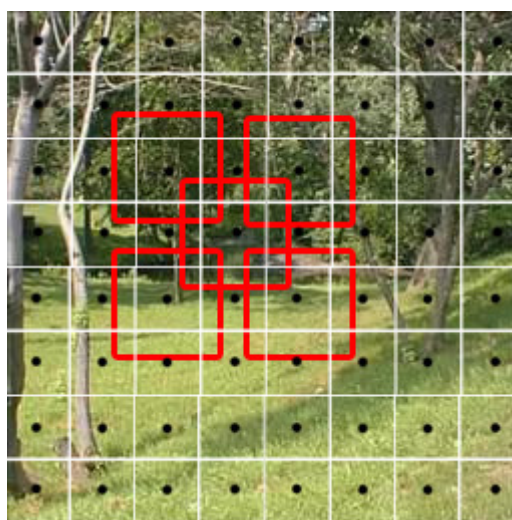


شکل ۴-۱۵: بلوک‌بندی  $4 \times 4$



شکل ۴-۱۶: بلوک‌بندی ۸×۸

چنان‌که ملاحظه می‌شود در همهٔ بلوک‌بندی‌ها اندازهٔ پنجره ثابت باقی می‌ماند. این مسئله مزیت بسیار مهمی است. زیرا با توجه به افزایش تعداد بلوک‌ها، پنجره‌ها با یکدیگر تداخل پیدا می‌کنند و این امر باعث استخراج اطلاعات مکان می‌شود. علاوه بر این، روند مذکور برای تصویر  $I$  در مقیاس‌های مختلف انجام می‌شود که این امر نیز باعث افزایش اطلاعات مکانی می‌شود. شکل ۴-۱۷ تصویر بلوک‌بندی شدهٔ  $I$  و پنجره‌هایی را که در هر بلوک قرار گرفته و با یکدیگر تداخل پیدا کرده‌اند، نشان می‌دهد.



شکل ۴-۱۷: تصویر بلوک‌بندی شدهٔ  $I$  و تداخل پنجره‌ها

اگر فرض شود  $S$  مجموعهٔ مقیاس‌ها و  $B$  مجموعهٔ بلوک‌بندی‌ها باشند که به صورت  $S = \{s_1, s_2, \dots, s_n\}$  و  $B = \{B_1, B_2, \dots, B_n\}$  تعریف شده‌اند، آنگاه شبه کد نشان داده شده در شکل ۴-۴

۱۸ روش پیشنهادی را برای ضبط اطلاعات مکانی نشان می‌دهد.

$S = \{s_1, s_2, \dots, s_n\} = \{1, 0/75, 0/5, 0/35, 0/25\}$

$B = \{B_1, B_2, \dots, B_k\}$

For  $s = s_1 : s_n$

For  $B = B_1 : B_k$

$I = \text{imresize}(I, s)$

Partition  $I$  to  $B$  block ( $I_B$ )

Compute feature vector for each window centered at block  $B$

end

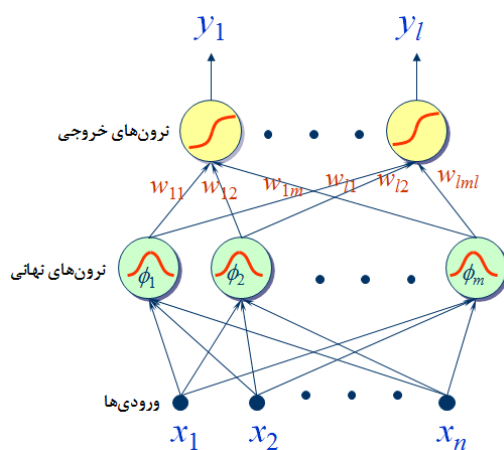
end

شکل ۴-۱۸: شبه کد ضبط اطلاعات مکانی

با اجرای الگوریتم فوق در کانال‌های رنگی متفاوت و نیز اجرای آن روی تصویر خاکستری حاصل از تصویر اصلی می‌توان اطلاعات متمایزتری از تصاویر استخراج کرد و به این ترتیب ویژگی‌های مختلف را از کانال‌های رنگی مختلف استخراج کرد.

#### ۴-۵. دسته‌بند پیشنهادی

شبکه‌های RBF جزو شبکه‌های نیمه‌نظارتی هستند که از یک لایه نهانی در معماری خود استفاده می‌کنند. در لایه نهانی با استفاده از توابع پایه گوسی، شبکه نگاشتی از فضای ورودی به فضایی با ابعاد بالاتر انجام می‌دهد. دو کاربرد اصلی و مهم شبکه‌های RBF در «دسته‌بندی الگوها» و «تقریب توابع» است (Montazer, Sabzevari and Khatir 2007). در شکل ۴-۱۹ معماری شبکه‌های استفاده شده برای دسته‌بندی الگوها ملاحظه می‌شود.



شکل ۴-۱۹: شبکه عصبی RBF برای دسته‌بندی الگوها

به طور کلی کارایی یک شبکه عصبی RBF تحت تأثیر سه عامل مهم قرار دارد: ۱- روش آموزش شبکه برای تعیین وزن‌های بهینه، ۲- روش تعیین مراکز توابع گوسی و ۳- روش تعیین پهناهای توابع گوسی.

در شبکه عصبی پیشنهادی با استفاده از الگوریتم بهینه‌سازی PSO روش جدیدی برای تعیین مراکز خوشه‌های توابع گوسی لایه نهانی پیشنهاد شده‌است. روش پیشنهادی در مقایسه با سایر روش‌ها نتایج مطلوب‌تری را تولید می‌کند.

#### ۱-۵-۴. تعیین وزن‌های لایه نهانی

تاکنون روش‌های بسیاری برای آموزش شبکه‌های عصبی RBF (تعیین وزن‌های لایه نهانی) معرفی شده‌است. منتظر و همکاران روش‌های جدیدی را برای آموزش شبکه عصبی RBF ارائه کرده‌اند که از کارایی بالایی برخوردار است (Montazer, Sabzevari and Khatir 2007). روش‌های ارائه شده نسخه‌های بهبود یافته‌ای از روش‌های «بیشترین کاهش»<sup>۱</sup> و «پخش سریع»<sup>۲</sup> هستند. نویسندگان در این مقاله برای بهبود کارایی روش بیشترین کاهش، رابطه جدیدی برای به روز کردن «آهنگ یادگیری»<sup>۳</sup> شبکه معرفی می‌کنند. برای تعیین وزن‌های خروجی روش یادگیری بیشترین کاهش بهینه (OSD)<sup>۴</sup> معرفی شده‌است. روش فوق در هر تکرار از الگوریتم آموزش، آهنگ یادگیری بهینه را محاسبه و از آن استفاده می‌کند. با استفاده از این روش وزن‌های شبکه به صورت زیر به روز می‌شوند:

$$\Delta W_{\text{opt}} = \lambda_{\text{opt}} \Delta W - \frac{B}{2C} = \frac{(E\Phi)(E\Phi)^T E\Phi}{(E\Phi\Phi^T)(E\Phi\Phi^T)^T} \quad (4-11)$$

بنابر این وزن‌های جدید برابر خواهند بود با:

$$W_{\text{new}} = W_{\text{old}} + \frac{(E\Phi)(E\Phi)^T E\Phi}{(E\Phi\Phi^T)(E\Phi\Phi^T)^T} \quad (4-12)$$

که مقدار اولیه  $W$  به صورت تصادفی انتخاب می‌شود.

#### ۲-۵-۴. الگوریتم بهینه‌سازی PSO

مطالعات بسیاری روی رفتار جمعی حیوانات و شبیه‌سازی حرکات هماهنگ آن‌ها صورت گرفته است که از این میان شبیه‌سازی پرواز پرندگان بسیار مورد توجه واقع شده‌است. یکی از روش‌های

---

<sup>۱</sup> Steepest Decent (SD)

<sup>۲</sup> Quick Propagation (QP)

<sup>۳</sup> Learning Rate

<sup>۴</sup> Optimum Steepest Descend (OSD)

بهینه‌سازی که براساس پرواز گروهی پرندگان و برای شبیه‌سازی جست‌وجوی غذای پرندگان طراحی شده‌است، الگوریتم بهینه‌سازی PSO نام دارد که توسط کندی و ابره‌ارت<sup>۱</sup> در سال ۱۹۹۵ معرفی شد. این نظریه بر این اصل استوار است که پرندگان در گروه، حرکات موفق همسایگان را تقلید می‌کنند. رفتار جمعی ناشی از این عمل، کشف نواحی بهینه در فضای جستجو است (Poli, Kennedy and Blackwell 2007).

الگوریتم PSO، گروهی از ذرات (پرندگان) را که هر ذره در آن یک پاسخ احتمالی مسئله است، ضبط و نگهداری می‌کند. در این الگوریتم، به هر ذره سرعتی نسبت داده می‌شود. ذرات در یک فضای چند بُعدی پرواز می‌کنند و موقعیت هر ذره با توجه به تجربه خودش و همسایگانش تنظیم می‌شود.

تغییر مکان ذرات در الگوریتم اولیه PSO به صورت زیر تعریف می‌شود:

$$v_{pd}^{new} = v_{pd}^{old} + c_1 r_1 (pbest_{pd} - x_{pd}^{old}) + c_2 r_2 (gbest_d - x_{pd}^{old}) \quad (4-13)$$

$$x_{pd}^{new} = x_{pd}^{old} + v_{pd}^{new} \quad (4-14)$$

در رابطه (۴-۱۳)  $c_1$  و  $c_2$  دو ثابت مثبت هستند که به ترتیب شاخص‌های شناختی<sup>۲</sup> و اجتماعی<sup>۳</sup> خوانده می‌شوند و  $r_1$  و  $r_2$  دو مقدار تصادفی تصادفی در بازه‌ی  $[0, 1]$  هستند.

بردار  $d$  بعدی  $x_{pd} = (x_{i1}, x_{i2}, \dots, x_{id})$  و  $pbest_{pd} = (p_{i1}, p_{i2}, \dots, p_{id})$  به ترتیب بیانگر موقعیت ذره<sup>۴</sup>  $i$ ام و بهترین مکان قبلی ذره  $i$ ام هستند. بردار  $gbest_d$  بیانگر مکان بهترین ذره در میان جمعیت و  $v_{pd} = (v_{i1}, v_{i2}, \dots, v_{id})$  نشان‌دهنده سرعت ذره  $i$ ام است. رابطه‌های (۴-۱۴) و (۴-۱۵) بیانگر چگونگی حرکت دسته ذرات هستند.

رابطه (۴-۱۳) از سه قسمت تشکیل شده است. قسمت اول بیانگر ممان لحظه‌ای<sup>۴</sup> است و نشان‌دهنده آن است که سرعت نمی‌تواند به طور ناگهانی تغییر یابد بلکه از سرعت فعلی شروع به تغییر می‌کند. قسمت دوم، دانش تجربی یک ذره است که عموماً به آن جزء شناختی گفته می‌شود و با فاصله ذره از بهترین مکانی که تا کنون در آن بوده است متناسب است. بهترین مکان هر ذره، از لحظه آغازین در نظر گرفته می‌شود. اطلاعاتی که بین ذره و همسایگانش مبادله می‌شود و در واقع بیانگر یادگیری از پرواز گروهی است، جزء سوم رابطه (۴-۱۳) را تشکیل می‌دهد و به جزء اجتماعی، شهرت دارد (Montazer and Giveki 2015).

الگوریتم بهینه‌سازی PSO طی سه گام برای حل مسائل بهینه‌سازی کار خود را انجام می‌دهد:

<sup>۱</sup>Eberhart & Kennedy

<sup>۲</sup>Cognitive

<sup>۳</sup>Social

<sup>۴</sup> Momentum

الف. با در نظر گرفتن گروهی از ذرات به صورت تصادفی راه حل‌ها، مقادیر سرعت و همچنین ذرات تعیین می‌گردند. سپس الگوریتم با انجام تکرارهایی جستجو را برای یافتن نقطه بهینه آغاز می‌کند.

ب. طی هر تکرار مقدار تابع هدف به ازای هر ذره محاسبه می‌شود، اگر مقدار بدست آمده برای ذره بهترین مقدار باشد، مکان ذره به عنوان  $pbest^1$  ذخیره می‌شود.

ج. در هر تکرار بین همه ذرات، ذره‌ای که بهترین مقدار تابع هدف را به خود اختصاص داده‌باشد انتخاب و مکان ذره به عنوان  $gbest^2$  ذخیره می‌شود.

د. براساس موقعیت‌های  $pbest$  و  $gbest$  سرعت‌ها تصحیح می‌شوند.

ه. موقعیت ذرات تثبیت می‌شود.

و. اگر برخی از شرایط مانند حداکثر تکرار و یا حداقل خطا حاصل شود، عملیات خاتمه می‌یابد، در غیر این‌صورت به گام دوم برمی‌گردیم (Montazer and Giveki 2015).

روابط بیان شده در قسمت قبل اولین روابط الگوریتم گروه ذرات را نشان می‌دهند. در این روابط هیچ ساز و کاری برای کنترل سرعت ذرات وجود ندارد، لذا لازم بود که یک مقدار حداکثر برای بیشترین سرعت آن‌ها در نظر گرفته شود، که در صورت خروج از این حد سرعت، لحاظ شود. مقدار این پارامتر برای فرایند بهینه‌سازی تعیین کننده است زیرا ممکن است به ازای مقادیر بالای سرعت، ذرات از پاسخ بهینه رد شوند و یا با سرعت کم از اکتشاف بهینه سراسری در فضای جستجو جلوگیری شود. لذا عدم وجود یک فرایند کنترل برای مقادیر سرعت، منجر به کاهش کارایی الگوریتم در مقابل سایر الگوریتم‌های تکاملی می‌شود، این مسئله باعث شد که یک پارامتر وزنی برای سرعت ذرات در نظر گرفته شود و روابط اولیه به صورت زیر اصلاح شوند :

$$v_{pd}^{new} = \omega v_{pd}^{old} + c_1 r_1 (pbest_{pd} - x_{pd}^{old}) + c_2 r_2 (gbest_d - x_{pd}^{old}) \quad (4-15)$$

$$x_{pd}^{new} = x_{pd}^{old} + v_{pd}^{new} \quad (4-16)$$

در روابط بالا  $\omega$  اینرسی<sup>۳</sup> وزنی نامیده می‌شود که نقش مهمی در کارایی الگوریتم PSO برعهده دارد. این پارامتر به صورت زیر به روز می‌شود :

---

<sup>۱</sup>Particle best

<sup>۲</sup>Global best

<sup>۳</sup>Inertia weight



$$\omega = \omega_{max} - (\omega_{max} - \omega_{min}) \frac{k}{k_{max}} \quad (۴-۱۷)$$

در رابطه (۴-۱۷)  $\omega_{max}$  برابر است با بیشترین میزان  $\omega$ ،  $\omega_{min}$  برابر است با کمترین میزان  $\omega$ ،  $k$  برابر است با تعداد تکرار فعلی و  $k_{max}$  برابر با حداکثر تعداد تکرارها است.

نقش اینرسی وزنی در معادله (۴-۱۷) به عنوان شاخصی مؤثر بر همگرایی PSO اثبات شده است (Liang et al 2006). اینرسی وزنی برای کنترل تأثیر سرعت‌های پیشین بر سرعت‌های جاری مورد استفاده قرار می‌گیرد. این پارامتر توانایی دسته را در اکتشاف مناطق بهینه سراسری و محلی و نیز استخراج پاسخ بهینه تنظیم می‌کند (Clerc 2010).

مقدار مناسب پارامترهای  $c_1$  و  $c_2$  باعث افزایش سرعت همگرایی و بهبود پاسخ بهینه می‌شوند. این ضرایب که ضرایب شتاب‌دهنده نامیده می‌شوند، در بازه  $[0,2]$  انتخاب می‌شوند.

پارامترهای  $r_1$  و  $r_2$  برای حفظ تنوع گروه به کار می‌روند و اعدادی تصادفی در بازه  $[0,1]$  با توزیع یکنواخت هستند. این پارامترها، اجازه حرکت ذرات در گام‌های تصادفی، در محدوده نقاط بهینه سراسری و محلی را می‌دهند.

با وجود اینکه روابط جدید (۴-۱۵) و (۴-۱۷) به افزایش کارایی PSO کمک می‌کنند، همچنان از کاستی‌های برخوردار هستند که این موجب کند شدن سرعت همگرایی الگوریتم PSO و نیز در گیر افتادن در کمینه محلی می‌شود.

تغییر  $\omega$  در رابطه (۴-۱۷) به صورت خطی است که موجب کاهش کارایی الگوریتم PSO می‌شود. همچنین تغییر مقدار  $\omega$  تنها به تعداد دفعات تکرار الگوریتم وابسته است و تنها با توجه به این عامل تعیین می‌شود که این باعث می‌شود الگوریتم PSO در حل مسایلی که بسیار غیرخطی و پیچیده‌اند به مشکل برخورد کند و نتواند کارایی مطلوبی را از خود نشان دهد. برای رفع مشکلات مذکور روش جدیدی برای تعیین مقدار مناسب  $\omega$  پیشنهاد می‌شود. روابط جدید به صورت زیر نوشته می‌شوند:

$$v_{id}^{k+1} = \omega_{id}^k v_{id}^k + c_1 r_1 (pbest_{id}^k - x_{id}^k) + c_2 r_2 (gbest_{id}^k - x_{id}^k) \quad (۴-۱۸)$$

$$x_{id}^{k+1} = x_{id}^k + \mu_i^k v_{id}^{k+1} \quad (۴-۱۹)$$

چنان‌که ملاحظه می‌شود در روابط جدید اینرسی وزنی و سرعت ذره به صورت تطبیق پذیری به روز می‌شوند. در حالت کلی مطلوب است آن است که ذرات در تکرارهای ابتدایی الگوریتم از قدرت جست‌وجوی سراسری بالا و در تکرارهای انتهایی از قدرت جست‌وجوی محلی مناسبی برخوردار باشند. این امر باعث افزایش کارایی الگوریتم و نیز همگرایی سریع‌تر آن می‌شود. بنابر آنچه گفته شد روابط زیر برای به روز کردن مقادیر  $\omega$  و  $\mu$  پیشنهاد می‌شوند:

طراحی سامانه بازیابی تصاویر □ ۸۶

$$\omega_i^k = k_1 h_i^k + k_2 b_i^k + \omega_0 \quad (۴-۲۰)$$

$$h_i^k = |(max\{F_{id}^k, F_{id}^{k-1}\} - min\{F_{id}^k, F_{id}^{k-1}\})/f_1| \quad (۴-۲۱)$$

$$b_i^k = 1/n * \sum_{i=1}^n (F_i^k - F^{avg})/f_2 \quad (۴-۲۲)$$

$$\mu_i^k = \begin{cases} \left(v_{max}/v_i^k\right) e^{-\left(k/k_{max}\right)^2} & \text{if } v_i^k > v_{max} \\ 1 & \text{if } v_{min} < v_i^k < v_{max} \\ \left(v_{min}/v_i^k\right) e^{-\left(k/k_{max}\right)^2} & \text{if } v_i^k < v_{min} \end{cases} \quad (۴-۲۳)$$

در روابط بالا  $\omega_0 \in (0,1]$  مقدار اولیه اینرسی وزنی است که اثر سرعت‌های قبلی دیده‌شده را تنظیم می‌کند در انجام آزمایش این مقدار برابر با ۱ در نظر گرفته شده‌اند. در رابطه (۴-۲۰) ضرایب  $k_1$  و  $k_2$  به طور تجربی و در بازه  $[0,1]$  انتخاب می‌شوند.

در رابطه (۴-۲۳) پارامتر  $\mu$  به طور تطبیق‌پذیری مقدار  $v^{k+1}$  را با در نظر گرفتن مقدار  $v^k$  و نیز تعداد تکرارها به روز می‌کند. در روابط (۴-۲۰) تا (۴-۲۳) سایر پارامترها به صورت زیر تعریف می‌شوند:

$h_i^k$  سرعت تکامل تدریجی فرایند جست‌وجو را تعیین می‌کند.

$b_i^k$  متوسط واریانس تابع هدف<sup>۱</sup> دسته ذرات است.

$F_{id}^k$  مقدار تابع هدف به ازای  $pbest_{id}^k$  است.

$F_{id}^{k-1}$  مقدار تابع هدف به ازای  $pbest_{id}^{k-1}$  است.

$f_1$  تابع هنجارکننده محلی است که به صورت زیر محاسبه می‌شود:

$$f_1 = \max \{ \Delta F_1, \Delta F_2, \dots, \Delta F_n \}, \Delta F_i = |F_{id}^k - F_{id}^{k-1}| \quad (۴-۲۴)$$

$n$  اندازه دسته ذرات است.

$F^{avg}$  برابر با مقدار میانگین توابع هدف همه ذرات در تکرار  $k$ ام است.

$f_2$  تابع هنجارکننده سراسری است که به صورت:

$$= \max \{ |F_1^k - F^{avg}|, |F_2^k - F^{avg}|, \dots, |F_n^k - F^{avg}| \} f_2 \quad (۴-۲۵)$$

محاسبه می‌شود.

در نهایت از الگوریتم معرفی شده (iPSO<sup>۲</sup>) برای تعیین مراکز توابع گوسی در لایه نهانی شبکه عصبی RBF استفاده می‌شود. روش ارایه شده تلاش می‌کند  $N$  نمونه در فضای  $\mathcal{R}^n$  را طوری در  $K$  خوشه قرار دهد که مجموع فواصل اقلیدسی هر نمونه تا مرکز خوشه کمینه شود. شکل ۴-۲۰ شبه کد محاسبه

<sup>۱</sup> Fitness function

<sup>۲</sup> Improved PSO

موارد فوق را نشان می‌دهد.

|                                                                                                             |
|-------------------------------------------------------------------------------------------------------------|
| <b>For each Particle [i] Do</b>                                                                             |
| Initialize Position vector $X[i]$ in the range of maximum and minimum of dataset patterns                   |
| Initialize Velocity vector $V[i]$ in the range of $[-a, a]$ ( $a = \max(\text{data}) - \min(\text{data})$ ) |
| Put initial Particle[i] into $pbest_{id}[i]$                                                                |
| <b>Nexti</b>                                                                                                |
| <b>While</b> maximum iteration is not attained <b>Do</b>                                                    |
| <b>For each Particle[i] Do</b>                                                                              |
| <b>For each Cluster[j] Do</b>                                                                               |
| Compute the Fitness Function using Equations (۴-۱۸), (۴-۱۹) and (۴-۲۰)                                      |
| <b>Nextj</b>                                                                                                |
| <b>Next i</b>                                                                                               |
| <b>if</b> run number is greater than 1 <b>Then</b>                                                          |
| <b>For each Particle[i] Do</b>                                                                              |
| <b>For each Cluster[j] Do</b>                                                                               |
| <b>if</b> Fitness Function of Particle[i]'S Cluster[j] is better than Fitness Function of                   |
| $pbest_{id}[i]$ 'S Cluster [j] <b>Then</b>                                                                  |
| Put Cluster[j] of Particle[i] into Cluster[j] of $pbest_{id}[i]$                                            |
| <b>Endif</b>                                                                                                |
| <b>Nextj</b>                                                                                                |
| <b>Nexti</b>                                                                                                |
| <b>Endif</b>                                                                                                |
| <b>For each Cluster[j] Do</b>                                                                               |
| <b>For each Particle[i] Do</b>                                                                              |
| Put the best of $pbests$ in terms of Fitness Function into $gbest$                                          |
| <b>Nexti</b>                                                                                                |
| <b>Nextj</b>                                                                                                |
| Compute inertia weight $\omega$ using Equation (۴-۲۰)                                                       |
| Compute $\mu$ using Equation (۴-۲۳)                                                                         |
| <b>For each Particle [i] Do</b>                                                                             |
| <b>Update</b> Velocity vector $V[i]$ using Equation (۴-۱۸)                                                  |
| <b>Update</b> Position vector $X[i]$ using Equation (۴-۱۹)                                                  |
| <b>Nexti</b>                                                                                                |
| <b>Endwhile</b>                                                                                             |
| Having processes algorithm, RBF unit centers are adjusted with $gbest$                                      |

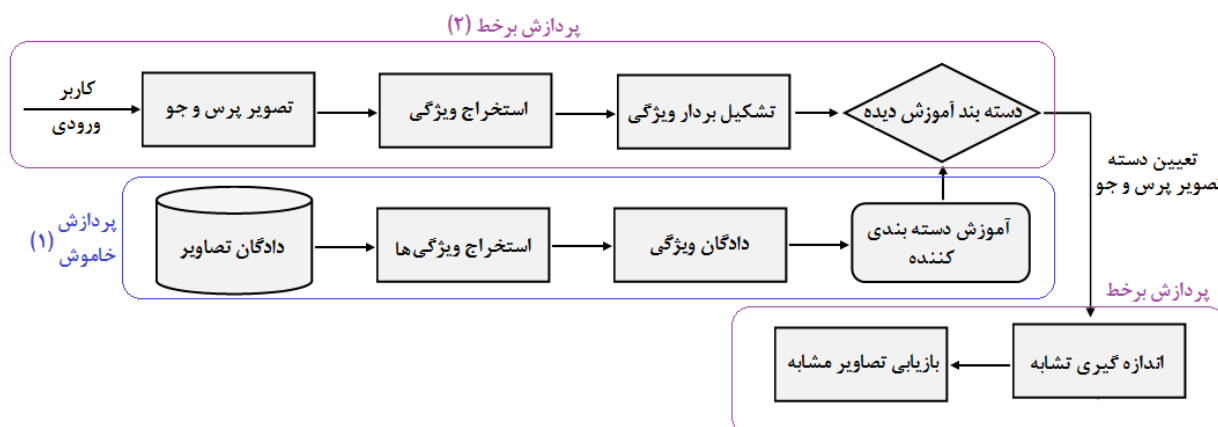
شکل ۴-۲۰: شبه‌کد روش محاسبه مراکز توابع گوسی با استفاده از الگوریتم iPSO

پس از یافتن مراکز توابع گوسی پهنای آن‌ها با استفاده از روش نزدیک‌ترین  $p$  همسایه<sup>۱</sup> محاسبه می‌شود. برای آموزش شبکه و یافتن وزن‌های بهینه نیز از روش «بیشترین کاهش بهینه» استفاده می‌شود.

<sup>۱</sup> $p$ -nearest neighbour

## ۴-۶. سامانه طراحی شده

در این قسمت به بررسی سامانه طراحی شده برای بازیابی تصاویر پرداخته می‌شود. سامانه طراحی شده الگوریتم‌های پیشنهادی در قسمت‌های قبل را به عنوان مغز موتور جویش، در خود جای داده است.



شکل ۴-۲۱: اجزای اصلی تشکیل دهنده موتور جویشگر طراحی شده

در شکل ۴-۲۱ اجزای موتور جویش تصویری نمایش داده می‌شود، ملاحظه می‌شود که پردازش‌های انجام شده به وسیله سامانه به دو دسته کلی تقسیم می‌شوند :

۱. پردازش‌های برون خط<sup>۱</sup>

۲. پردازش‌های برخط<sup>۲</sup>.

در سامانه طراحی شده، ابتدا پردازش‌های خاموش انجام می‌شود. این پردازش‌ها شامل استخراج ویژگی از کل تصاویر پایگاه داده و آموزش دسته‌بندهای RBF و SVM است. ویژگی‌های استخراج شده عبارتند از ویژگی‌های سیفت فشرده بهبودیافته و نیز الگوی دودویی محلی مکانی که به ترتیب در بخش‌های ۴-۴-۱ و ۴-۴-۲ توضیح داده شدند.

پردازش‌های برخط شامل استخراج بردار ویژگی تصویر پرس‌وجو، تعیین دسته تصویر پرس‌وجو و اندازه‌گیری تشابه تصویر پرس‌وجو با سایر تصاویر و مرتب‌سازی نتایج است. در پردازش‌های برخط ابتدا کاربر تصویر پرس‌وجو را انتخاب می‌کند و سپس بردار ویژگی این تصویر استخراج می‌شود. بردار ویژگی استخراج شده به دسته‌بند آموزش یافته وارد می‌شود تا دسته‌ای را که تصویر پرس‌وجو به آن تعلق دارد، تعیین کند. پس از تعیین دسته تصویر پرس‌وجو، شبیه‌ترین تصاویر به تصویر پرس‌وجو تعیین و به طور

<sup>۱</sup> offline

<sup>۲</sup> online

مرتب شده به کاربر نشان داده می‌شود.

شکل ۴-۲۲ صفحه اصلی سامانه طراحی شده را نشان می‌دهد. چنانچه در این شکل ملاحظه می‌شود:

۱. سامانه تحت وب پیاده‌سازی شده است (قسمت (۱)).

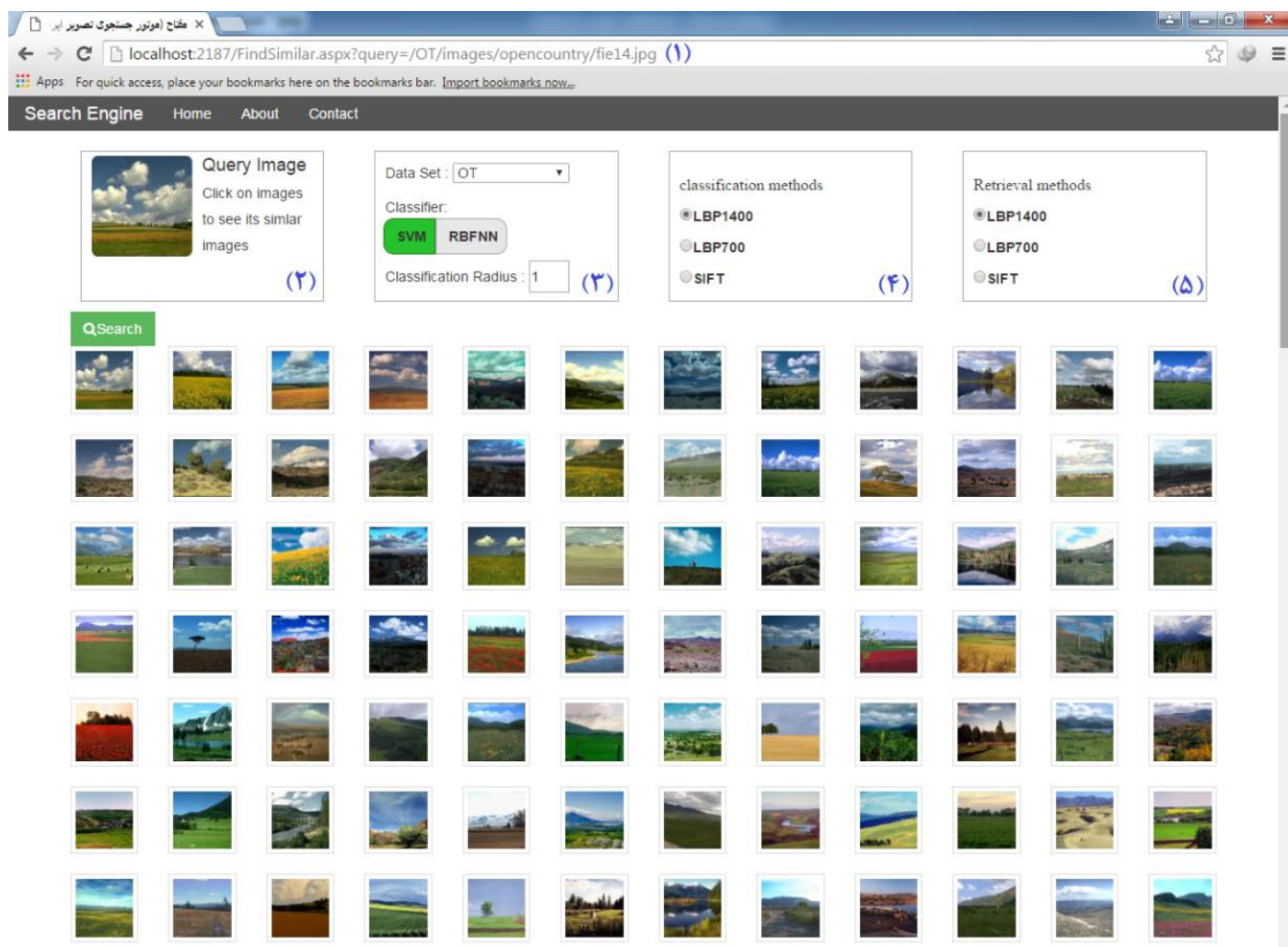
۲. برای کار با سامانه کاربر می‌تواند تنظیمات مختلفی را در نظر گرفت. در قسمت (۲) کاربر می‌تواند تصویر پرس‌وجو را انتخاب کند.

۳. در قسمت (۳) تنظیمات مختلفی وجود دارد که کاربر می‌تواند به دلخواه خود آن‌ها را اعمال کند. در این قسمت کاربر می‌تواند دسته‌بند مورد نظر خود را انتخاب کند. برای ایشان دو انتخاب متصور است: دسته‌بند RBF و دسته‌بند SVM. علاوه بر آن کاربر می‌تواند پایگاه داده مورد نظر خود را انتخاب کند. یکی دیگر از ویژگی‌های مهم سامانه، شعاع دسته‌بندی<sup>۱</sup> است. این پارامتر نقش مهمی در تولید نتایج نهایی جست‌وجو ایفا می‌کند به همین دلیل در ادامه نقش این پارامتر توضیح داده خواهد شد.

انتخاب پایگاه داده یکی دیگر از مواردی است که در این قسمت تعبیه شده است. کاربر می‌تواند از میان چندین پایگاه داده موجود یکی را به دلخواه خود انتخاب کند.

۴. در قسمت (۴) کاربر ویژگی را که با آن دسته معنایی تصاویر مشخص می‌شود، انتخاب می‌کند. چنانچه در شکل ۴-۲۲ ملاحظه می‌شود کاربر می‌تواند از ویژگی‌های الگوی دودویی محلی مکانی (با طول بردارهای ویژگی ۷۰۰ و ۱۴۰۰) و نیز سیفت فشرده بهبود یافته (با طول بردار ویژگی ۱۵۰۰) استفاده کند.

۵. در قسمت (۵) ویژگی‌هایی که با استفاده از آن‌ها عمل مرتب‌سازی تصاویر صورت می‌گیرد، انتخاب می‌شوند. این ویژگی‌ها در شکل ۴-۲۲ مشخص شده‌اند.

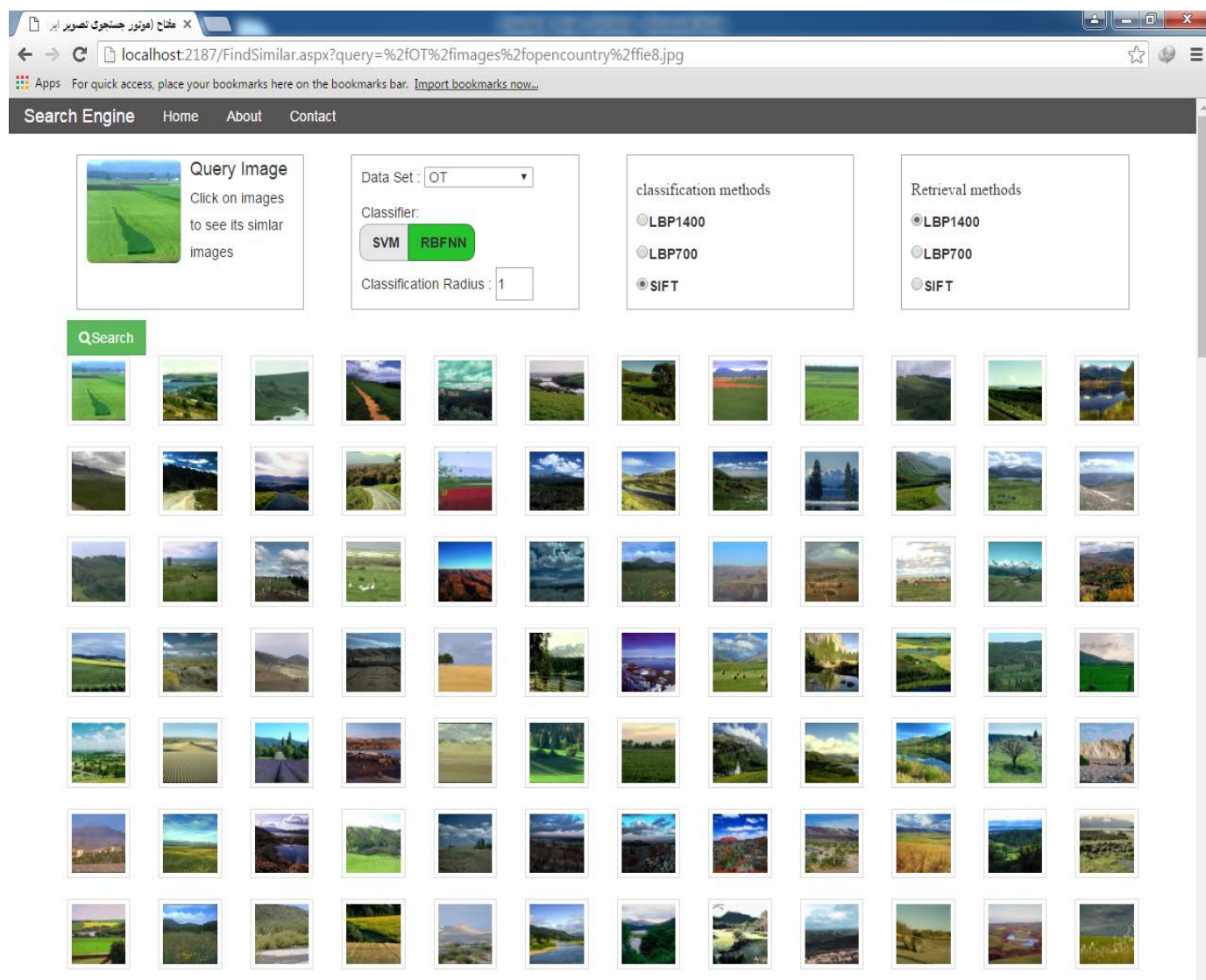


شکل ۴-۲۲: نمای کلی سامانه بازیابی تصاویر مشابه

برای نمونه شکل ۴-۲۳ نتایج تصاویر مشابه با تصویر پرسوجوی دشت از پایگاه داده OT را نشان می‌دهد. در این پرسوجو برای یافتن تصاویر مشابه از دسته‌بندی‌کننده iPSO-OSD، شعاع دسته‌بندی برابر با ۱، ویژگی‌های سift فشرده بهبود یافته برای تعیین دسته معنایی تصویر پرسوجو و ویژگی‌های الگوی دودویی محلی مکانی برای مرتب‌سازی نتایج، استفاده شده‌است.

هم‌چنان‌که گفته شد پس از وارد کردن تصویر پرسوجو هر یک از نرون‌های لایه خروجی دسته‌بندی‌کننده iPSO-OSD مقداری را به عنوان خروجی نرون تولید می‌کنند و نرونی خروجی که بیشترین مقدار را داشته باشد به عنوان نرون برنده تشخیص داده می‌شود. این نرون دسته‌ای را مشخص می‌کند که تصویر پرسوجو به آن تعلق دارد.

## طراحی سامانه بازیابی تصاویر □ ۹۱



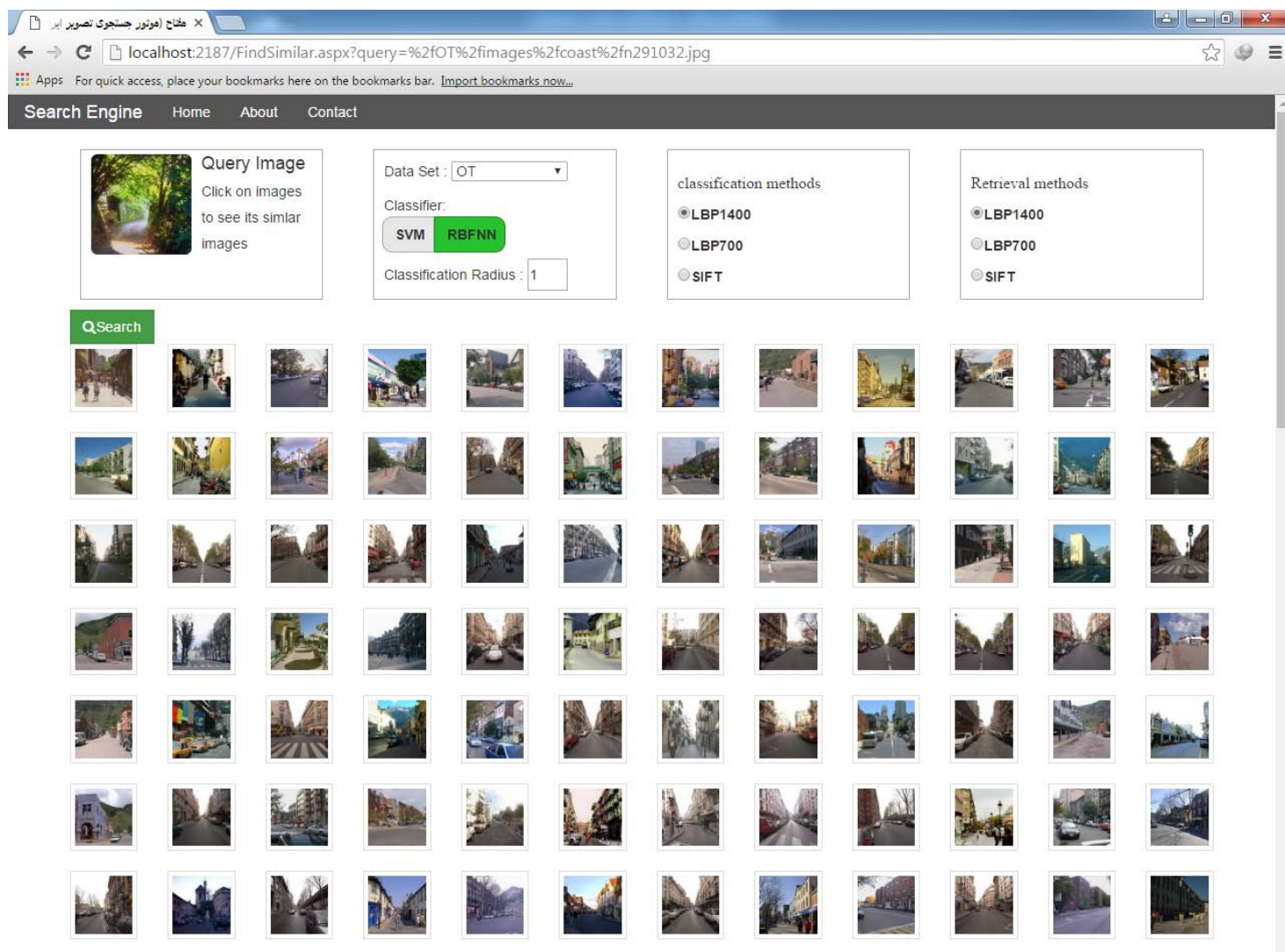
شکل ۴-۲۳: تصاویر مشابه با تصویر دشت

اگر دسته‌بند دسته تصویر پرس‌وجو را درست تشخیص داده باشد همه تصاویر بازیابی شده شبیه به تصویر پرس‌وجو خواهند بود و بهترین نتایج حاصل می‌شود. اما اگر دسته‌بند نادرست تشخیص دهد، نتایج رضایت بخش نخواهند بود. برای حل این مشکل پارامتری به نام شعاع دسته‌بندی در سامانه طراحی شده‌است. اگر کاربر نتایج مورد نظر خود را دریافت نکند، با تغییر این پارامتر می‌تواند تصاویر شبیه به هم را بازیابی کند. اگر شعاع بازیابی برابر با ۱ در نظر گرفته شود، بدان معنا است که کاربر مایل است تصاویر مشابه را تنها از دسته‌ای که نرون مربوط به آن بزرگ‌ترین خروجی را تولید کرده‌است، بازیابی کند. اگر شعاع بازیابی برابر با ۳ در نظر گرفته شود، سامانه تصاویر مشابه را از میان ۳ دسته‌ای که نرون‌های مربوط به آن‌ها بزرگ‌ترین مقدار خروجی را تولید کرده‌اند، بازیابی می‌کند. به این ترتیب کاربر در صورت عدم رضایت از نتایج حاصل شده، می‌تواند با تغییر شعاع دسته‌بندی به نتایج مطلوب دست یابد.



طراحی سامانه بازیابی تصاویر ۹۲

برای نمونه اگر تصویر پرسوجوی مشخص شده در شکل ۴-۲۴ به عنوان پرسوجو انتخاب شود، هم‌چنان‌که ملاحظه می‌شود نتایج بسیار ناامیدکننده خواهند بود. لازم به ذکر است که در این حالت از دسته‌بندی‌کننده iPSO-OSD، شعاع دسته‌بندی ۱ و ویژگی‌های الگوی دودویی محلی مکانی برای دسته‌بندی و نیز مرتب‌سازی تصاویر استفاده شده‌است.

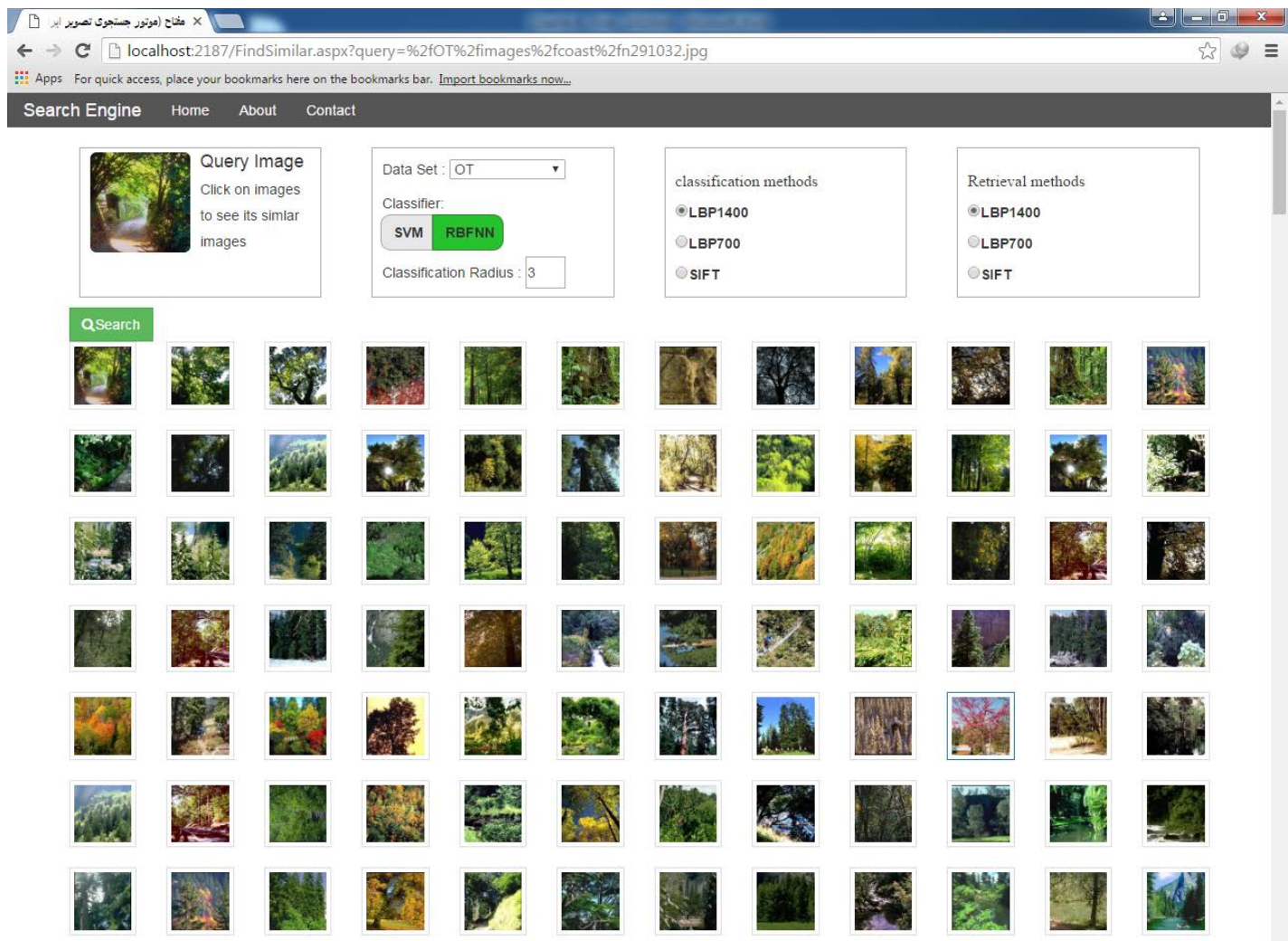


شکل ۴-۲۴: نتایج جست‌وجوی تصویر جنگل با شعاع دسته‌بندی برابر با ۱

حال با تغییر شعاع دسته‌بندی از ۱ به ۳ نتایج بازیابی به صورتی خواهد شد که در شکل ۴-۲۵ نشان داده شده‌است.



## طراحی سامانه بازیابی تصاویر □ ۹۳

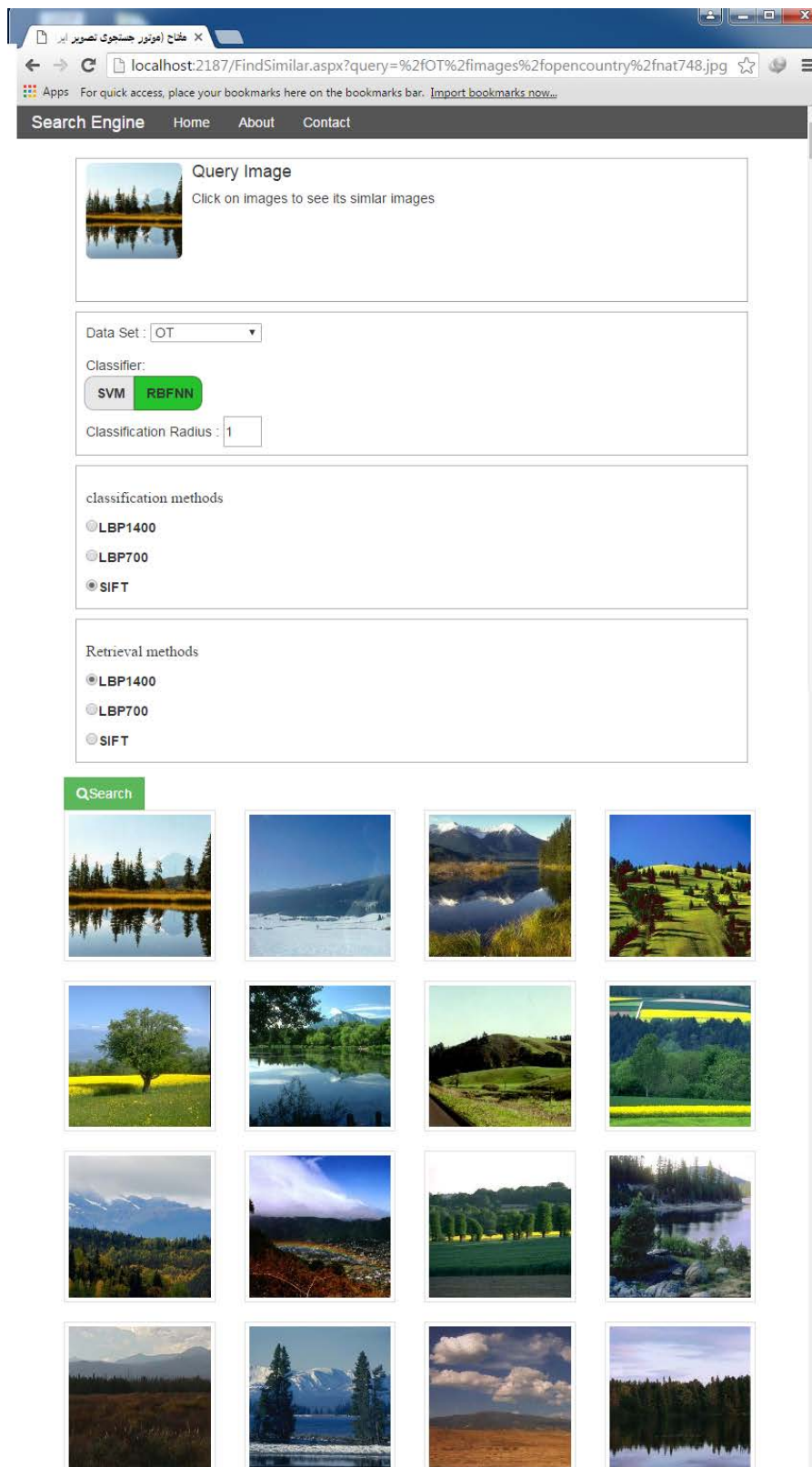


شکل ۴-۲۵: بهبود نتایج جست‌وجوی تصویر جنگل با تغییر شعاع دسته‌بندی ۱ به ۳

یکی دیگر از ویژگی‌های سامانه، کنش‌گرا<sup>۱</sup> بودن آن است. این قابلیت این ویژگی را به سامانه اضافه می‌کند که می‌توان به راحتی از آن در تلفن‌های همراه و تبلت‌ها استفاده کرد. شکل ۴-۲۶ این قابلیت را با تغییر اندازه مرورگر نشان می‌دهد.

<sup>۱</sup>Responsive

## طراحی سامانه بازیابی تصاویر □ ۹۴



شکل ۴-۲۶: قابلیت کنش گرا بودن سامانه طراحی شده

## ۷-۴. نتیجه گیری

در این فصل دستاوردهای مهم رساله معرفی شدند و سپس با استفاده از الگوریتم‌های پیشنهادی سامانه‌ای برای یافتن تصاویر مشابه طراحی و پیاده‌سازی شد. ابتدا روشی جدید برای استخراج ویژگی‌های تصویر معرفی شد. این روش مبتنی بر ویژگی‌های استخراج شده به وسیله گرادیان تصویر در جهت‌های مختلف، محاسبه برآیند و خوشه‌بندی این ویژگی‌ها است.

در ادامه روش جدیدی برای ضبط اطلاعات مکانی و بهبود روش بسته واژگان تصویری ارائه شد. در روش مذکور تصویر به بلوک‌هایی تقسیم می‌شود. سپس با تغییر اندازه بلوک‌بندی و با ثابت نگه‌داشتن پنجره‌ای که ویژگی‌ها از آن استخراج می‌شود و نیز با در نظر گرفتن تصویر در مقیاس‌های متفاوت، اطلاعات مکانی به خوبی ضبط می‌شوند.

یکی دیگر از دستاوردهای رساله ارائه یک دسته‌بند جدید با استفاده از شبکه‌های عصبی RBF است. در این دسته‌بند برای تعیین مراکز و پهنای توابع گوسی لایه نهانی از الگوریتم جدیدی مبتنی بر روش PSO استفاده شده‌است. در نهایت همه الگوریتم‌های پیشنهادی در قالب سامانه‌ای تحت وب و با قابلیت کنش‌گرا بودن پیاده‌سازی و استفاده شدند.

## فصل پنجم

جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده

## ۱-۵. مقدمه

در این فصل ارزیابی روش پیشنهادی و مقایسه آن با سایر روش‌ها و نیز جمع‌بندی و نتیجه‌گیری رساله آورده شده است. همچنین نوآوری تحقیق، محدودیتهای تحقیق و پیشنهاد برای تحقیقات آتی بیان خواهد شد.

## ۲-۵. معیارهای ارزیابی در حوزه بازتابی تصاویر

معیارهای مختلفی در حوزه بازتابی و دسته‌بندی تصاویر برای ارزیابی روش‌های پیشنهادی وجود دارد. از مهم‌ترین این معیارها می‌توان به «دقت<sup>۱</sup>»، «بازخوانی<sup>۲</sup>» و «متوسط میانگین دقت<sup>۳</sup>» اشاره کرد (Alzu'bi, Amira and Ramzan 2015). برای محاسبه معیارهای فوق از ماتریس «درهم‌ریختگی<sup>۴</sup>» به صورت زیر استفاده می‌شود:

جدول ۵-۱: جدول ارزیابی

|       |      | پیش‌بینی شده |      |
|-------|------|--------------|------|
|       |      | منفی         | مثبت |
| واقعی | منفی | a            | b    |
|       | مثبت | c            | d    |

در این صورت خواهیم داشت:

---

<sup>۱</sup> Precision

<sup>۲</sup> Recall

<sup>۳</sup> mean Average Precision (mAP)

<sup>۴</sup> Confusion matrix

$$recall = TP = \frac{d}{c + d} \quad (۵-۱)$$

$$FP = \frac{b}{a + b} \quad (۵-۲)$$

$$TN = \frac{a}{a + b} \quad (۵-۳)$$

$$FN = \frac{c}{c + d} \quad (۵-۴)$$

$$Precision = P = \frac{d}{b + d} \quad (۵-۵)$$

معمولاً دقت فقط برای اولین  $r$  تصویر بازیابی شده محاسبه می‌شود زیرا که کاربر به اولین نتایج بازیابی شده به عنوان پاسخ مورد نظرش علاقه بیشتری دارد. معمولاً در ارزیابی‌ها فرض بر این است که کاربر به اولین ۲۰، ۳۰ یا حداکثر ۵۰ تصویر بازیابی شده علاقه مند است. البته با توجه به پایگاه داده‌های استفاده شده، عدد  $r$  می‌تواند تغییر کند.

در این رساله با توجه به بازیابی تصاویر با استفاده از دسته‌بندی آن‌ها از ماتریس درهم‌ریختگی که اطلاعاتی را دربارهٔ قدرت دسته‌بند به دست می‌دهد، استفاده می‌شود. این ماتریس مشخص می‌کند برای هر دسته از تصاویر پایگاه داده، دسته‌بند نمونه‌های آزمون را چگونه تشخیص داده‌است. چه نمونه‌هایی را درست و چه نمونه‌هایی را نادرست تشخیص داده‌است. همچنین این ماتریس مشخص می‌کند که نمونه‌هایی که به نادرستی دسته‌بندی شده‌اند، در کدام دسته‌ها قرار گرفته‌اند. بنابراین می‌توان با استفاده از این ماتریس به قدرت دسته‌بندی‌کننده در تمایز دسته‌های مختلف پی برد.

میانگین دقت<sup>۱</sup> (AP) به عنوان میانگین دقت‌های محاسبه شده برای هر تصویر مرتبط بازیابی شده محاسبه می‌شود. سپس mAP به عنوان مقدار متوسط AP روی مجموعه‌ای از پرس‌وجوها محاسبه می‌شود. در واقع دقت، اطلاعاتی دربارهٔ با میزان مرتبط بودن اولین اسناد بازیابی شده به دست می‌دهد در حالی که mAP اطلاعاتی دربارهٔ کل رتبه‌بندی به دست می‌دهد. : در این رتبه‌بندی هر چه تصاویر مرتبط دیرتر ظاهر شوند (رتبهٔ کمتری داشته باشند)، mAP عددی کوچکتر خواهد بود. در آزمایش‌های انجام شده در هر جا که لازم بوده‌است و مطابق با مقالات بین‌المللی معتبر داده‌های استفاده شده به داده‌های آموزش، اعتبارسنجی و آزمون تقسیم شده‌اند.

## ۳-۵. نتایج ارزیابی روش‌های پیشنهادی روی پایگاه داده‌های معیار

### ۱-۳-۵. ارزیابی روش سیفت فشرده

---

<sup>۱</sup> Average Precision (AP)



برای ارزیابی روش پیشنهادی سift فشرده از پایگاه داده ونگ که پایگاه داده‌ای معیار است، استفاده شده‌است. این پایگاه داده حاوی ۱۰۰۰ تصویر در ۱۰ دسته صدتایی است (Wang, Li and Wiederhold 2001). تصاویر به طور معمول در اندازه‌های ۲۵۶×۳۸۴ و ۳۸۴×۲۵۶ هستند. شکل ۵-۱ تعدادی از این تصاویر را نشان می‌دهد.



شکل ۵-۱: تعدادی از تصاویر پایگاه داده ونگ

۱۰ دسته موجود در این پایگاه داده عبارتند از: روستایان افریقایی، ساحل، ساختمان‌ها، اتوبوس‌ها، دایناسورها، گل‌ها، کوهستان و یخ‌ها، اسب‌ها و غذا. جدول ۵-۲ نتایج حاصل از پیاده‌سازی روش پیشنهادی پنجم با تقسیم تصویر به ۱۶ قطعه و ۱۲۸ خوشه و مقایسه آن با سایر روش‌ها را نشان می‌دهد.

جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □۱۰۰

### جدول ۵-۲: متوسط دقت حاصل از روش سیفت فشرده در مقایسه با سایر روش‌ها

| دسته / روش                                       | روستایان آفریقایی | ساحل  | ساختمان‌ها | اتوبوس‌ها | دایناسورها | فیل‌ها | گل‌ها | اسب‌ها | کوهستان و یخ‌ها | غذا   | میانگین |
|--------------------------------------------------|-------------------|-------|------------|-----------|------------|--------|-------|--------|-----------------|-------|---------|
| روش پیشنهادی در رساله                            | 65/41             | 58/36 | 67/31      | 95/83     | 99/35      | 76/9   | 89/3  | 88/5   | 61/27           | 79/4  | ۷۸/۱۶   |
| روش پیشنهادی در (Jhanwar et al 2004)             | 42/25             | 39/75 | 37/35      | 74/10     | 91/45      | 30/40  | 85/15 | 53/80  | 29/25           | 36/95 | 52/64   |
| روش پیشنهادی در (Huang and Dai 2003)             | 42/40             | 44/55 | 41         | 85/15     | 58/65      | 42/55  | 89/75 | 58/90  | 26/80           | 42/65 | 53/24   |
| روش پیشنهادی در (Lin, Chen and Chan 2009)        | 68/30             | ۵۴    | 56/15      | 88/80     | 99/25      | 65/80  | 89/10 | 80/25  | 52/15           | 73/25 | 72/7    |
| روش پیشنهادی در (Tian et al 2014)                | 74/6              | 37/8  | 53/9       | 96/۷      | ۹۹         | 65/9   | 91/2  | 86/9   | 58/5            | 62/2  | 72/7    |
| روش پیشنهادی در (ElAlami 2011)                   | ۷۰                | ۵۶    | ۵۷         | ۸۸        | ۹۹         | ۶۸     | ۹۱    | ۸۳     | ۵۴              | ۷۳    | ۷۳/۹    |
| روش پیشنهادی در (Banerjee , Kundu and Maji 2009) | ۶۱                | ۵۶    | ۶۳         | ۷۲        | ۹۵         | ۷۷     | ۸۳    | ۹۵     | ۶۸              | ۵۷    | ۷۲/۷    |
| روش پیشنهادی در (Qi and Han 2005)                | ۶۸                | ۶۵    | ۵۲         | ۹۷        | ۹۹         | ۵۲     | ۸۳    | ۷۸     | ۶۱              | 73    | ۷۲/۸    |
| روش پیشنهادی در (Irtaza et al ۲۰۱۴)              | ۶۵                | ۶۰    | ۶۲         | 85        | ۹۳         | ۶۵     | ۹۴    | ۷۷     | ۷۳              | 81    | ۷۵/۵    |
| روش پیشنهادی در (ElAlami ۲۰۱۴)                   | ۷۲/۶              | ۵۹/۳  | ۵۸/۷       | ۸۹/۱      | ۷۷/۲       | ۹۹/۳   | ۷۰/۲  | ۹۲/۸   | ۸۵/۶            | ۵۶/۲  | ۷۶/۱    |
| روش پیشنهادی در (Subrahmanyam 2001)              | ۷۷/۷              | ۴۹/۴  | ۵۹/۴       | ۹۰/۷۰     | ۹۹         | ۶۱/۲   | ۹۱/۴  | ۹۲/۲   | ۴۶/۴            | ۷۲/۸  | ۷۴      |



جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □ ۱۰۱

چنان‌که جدول نشان می‌دهد روش پیشنهادی در مقایسه با سایر روش‌ها از کارایی بیشتری برخوردار است. نتایج نشان می‌دهد روش پیشنهادی برای دسته‌های ساحل و کوهستان و یخ‌ها نسبت به سایر دسته‌ها از کارایی کمتری برخوردار بوده است. دلیل این امر ماهیت تصاویر موجود در این دسته‌ها است. تصاویر موجود در این دسته‌ها تصاویری بافتاری هستند. از طرفی ویژگی‌های استخراج شده از تصاویر به وسیله روش پیشنهادی، مبتنی بر گرادیان تصویر هستند. از آنجا که ویژگی گرادیان برای تصاویر بافتاری آنچنان کارا و متمایز نیست، روش پیشنهادی در بازیابی این تصاویر آنچنان کارا نبوده است. از سوی دیگر روش پیشنهادی برای دسته تصاویر اتوبوس‌ها، دایناسورها، گل‌ها، اسب‌ها، غذا، ساختمان‌ها و روستاییان آفریقایی از کارایی مطلوبی برخوردار است. دلیل این امر آن است که ویژگی‌های استخراج شده از این تصاویر به طور مؤثری می‌توانند محتوای تصویر را توصیف کنند.

### ۲-۳-۵. ارزیابی توصیف‌گر سift فشرده بهبود یافته

در این قسمت نتایج حاصل از بهبود توصیف‌گر سift فشرده روی پایگاه داده ونگ ارایه می‌شود. در این آزمایش‌ها کارایی روش پیشنهادی با استفاده از روش بسته واژگان و با تغییر طول بردار ویژگی (اندازه دایره واژگان) سنجیده می‌شود. جدول‌های ۳-۵ و ۴-۵ نتایج این آزمایش‌ها را نشان می‌دهند. لازم به ذکر است که در همه آزمایش‌های انجام شده ۴۰٪ تصاویر پایگاه داده برای آموزش و باقی تصاویر به عنوان آزمون استفاده شده‌اند. در پیاده‌سازی روش بسته واژگان از دسته‌بند SVM استفاده شده است.

جدول ۳-۵: نتایج محاسبه بردارهای گرادیان در جهت‌های مختلف

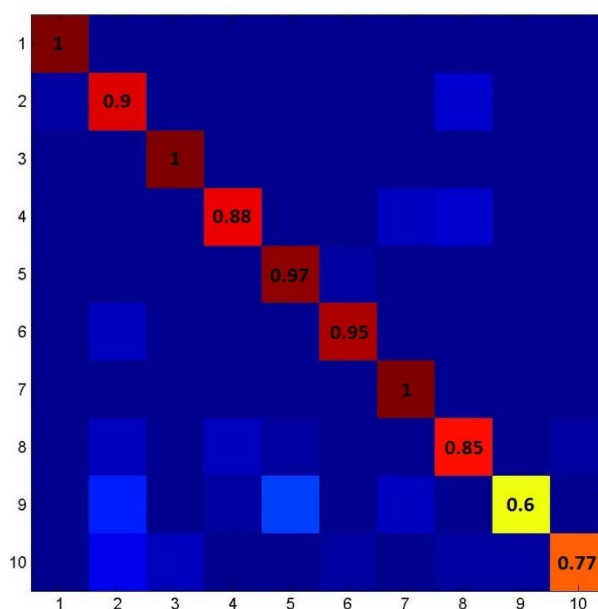
| طول بردار ویژگی | تعداد بین‌ها | تعداد جهت‌ها | کارایی | زمان (ثانیه) |
|-----------------|--------------|--------------|--------|--------------|
| ۱۰۰             | ۱۶           | ۱۶           | ۸۴/۵   | ۶۹۹          |
| ۲۰۰             | ۱۶           | ۱۶           | ۸۶/۵   | ۷۲۳          |
| ۳۰۰             | ۱۶           | ۱۶           | ۸۵/۵   | ۷۵۶          |
| ۴۰۰             | ۱۶           | ۱۶           | ۸۲/۵   | ۷۸۲          |
| ۵۰۰             | ۱۶           | ۱۶           | ۸۲/۲۵  | ۸۰۵          |
| ۶۰۰             | ۱۶           | ۱۶           | ۸۷/۵   | ۸۳۱          |
| ۷۰۰             | ۱۶           | ۱۶           | ۸۷/۷۵  | ۸۵۲          |
| ۸۰۰             | ۱۶           | ۱۶           | ۷۵     | ۸۷۴          |
| ۹۰۰             | ۱۶           | ۱۶           | ۸۶     | ۸۹۶          |
| ۱۰۰۰            | ۱۶           | ۱۶           | ۸۷     | ۹۳۷          |
| ۱۲۰۰            | ۱۶           | ۱۶           | ۷۱     | ۹۹۴          |
| ۱۴۰۰            | ۱۶           | ۱۶           | ۸۹/۲۵  | ۱۰۶۲         |
| ۱۱۹۱            | ۸۷/۵         | ۱۶           | ۱۶     | ۱۸۰۰         |

### ادامه جدول ۵-۳: نتایج محاسبه بردارهای گرادیان در جهت‌های مختلف

| طول بردار ویژگی | تعداد بین‌ها | تعداد جهت‌ها | کارایی | زمان (ثانیه) |
|-----------------|--------------|--------------|--------|--------------|
| ۱۲۸۰            | ۸۷/۲۵        | ۱۶           | ۱۶     | ۲۰۰۰         |
| ۱۲۵۶            | ۸۸/۵         | ۱۶           | ۱۶     | ۲۴۰۰         |
| ۱۵۱۱            | ۸۸           | ۱۶           | ۱۶     | ۲۸۰۰         |
| ۱۷۱۶            | ۸۶/۷۸        | ۱۶           | ۱۶     | ۳۲۰۰         |
| ۱۹۸۶            | ۸۹           | ۱۶           | ۱۶     | ۳۶۰۰         |
| ۲۱۵۱            | ۸۸           | ۱۶           | ۱۶     | ۳۸۰۰         |
| ۲۲۵۶            | ۸۸/۷۸        | ۱۶           | ۱۶     | ۴۰۰۰         |

با توجه به جدول ۵-۳، روش پیشنهادی با پارامترهای (تعداد جهت، تعداد بین) = (۱۶، ۱۶) و طول بردار ویژگی برابر با ۱۴۰۰ به کارایی برابر با ۸۹/۲۵ می‌رسد. شکل ۵-۲ ماتریس درهم‌ریختگی را برای روش فوق نشان می‌دهد. در ماتریس مذکور دسته‌های (۱) تا (۱۰) به ترتیب عبارتند از اتوبوس‌ها، ساحل، دایناسورها، فیل‌ها، گل‌ها، غذا، اسب‌ها، کوهستان، بومیان آفریقایی و ساختمان‌ها. هم‌چنان‌که ملاحظه می‌شود با استفاده از روش پیشنهادی می‌توان به کارایی بالایی در بازیابی معنایی تصاویر با استفاده از دسته‌بندی آن‌ها رسید.

ex-Wang-Compressed Dense SIFT 1400 Words - mean accuracy: 0.892500



شکل ۵-۲: ماتریس درهم‌ریختگی برای روش سیفت فشرده بهبود یافته با طول بردار ویژگی ۱۴۰۰

هم‌چنان‌که ملاحظه می‌شود روش مذکور برای دسته‌های اتوبوس‌ها، دایناسورها و اسب‌ها به بالاترین کارایی دست یافته‌است. همچنین روش پیشنهادی برای دسته‌های ساحل و کوهستان نیز دارای کارایی

جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □ ۱۰۳

بسیار مطلوبی است. این امر نشان می‌دهد که تغییر اعمال شده در روش سیفت فشرده و به کارگیری آن در قالب روش بسته‌ واژگان تصویری می‌تواند برای تصاویر بافتی نیز به طور مؤثری به کارا واقع شود. کم‌ترین کارایی روش پیشنهادی به دسته‌های (۹) و (۱۰) (بومیان افریقایی و ساختمان‌ها) مربوط می‌شود. کارایی روش مذکور برای این دسته‌ها به ترتیب برابر است با: ۰/۶ و ۰/۷۷.

همان‌طور که ماتریس درهم‌ریختگی نشان می‌دهد، برای دسته‌ بومیان افریقایی بیشترین سردرگمی با دسته‌های (۲) و (۵) (ساحل و گل‌ها) رخ داده‌است. دلیل این امر آن است که بخش زیادی از تصاویر این دسته پس‌زمینه‌هایی مانند تصاویر ساحل دارند. همچنین در بسیاری از تصاویر این دسته بومیان افریقایی تاج‌هایی گل مانند برسر دارند و این امر باعث تشابه این دو دسته به یکدیگر می‌شود.

با توجه به ماتریس درهم‌ریختگی، دسته‌ (۱۰) (ساختمان‌ها) بیشترین سردرگمی را با دسته‌ (۲) (ساحل) دارد. دلیل این امر آن است که بخش وسیعی از تصاویر این دو دسته از بافت آسمان و ابر تشکیل شده‌است.

**جدول ۵-۴: نتایج محاسبه‌ بردارهای گرادیان در جهت‌های مختلف**

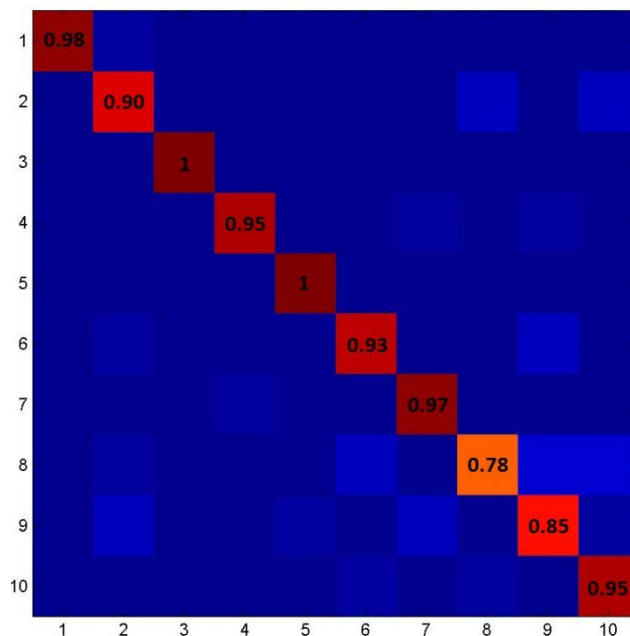
| طول بردار ویژگی | تعداد بین‌ها | تعداد جهت‌ها | کارایی | زمان (ثانیه) |
|-----------------|--------------|--------------|--------|--------------|
| ۱۰۰             | ۳۶           | ۳۶           | ۸۴     | ۸۵۱          |
| ۲۰۰             | ۳۶           | ۳۶           | ۸۶/۲۵  | ۹۳۴          |
| ۳۰۰             | ۳۶           | ۳۶           | ۸۶     | ۱۰۷۵         |
| ۴۰۰             | ۳۶           | ۳۶           | ۸۳/۷   | ۱۱۳۱         |
| ۵۰۰             | ۳۶           | ۳۶           | ۸۳/۷۵  | ۱۱۶۳         |
| ۶۰۰             | ۳۶           | ۳۶           | ۸۷/۵   | ۱۱۸۷         |
| ۷۰۰             | ۳۶           | ۳۶           | ۸۶/۷   | ۱۲۱۳         |
| ۸۰۰             | ۳۶           | ۳۶           | ۸۳/۵   | ۱۲۴۵         |
| ۹۰۰             | ۳۶           | ۳۶           | ۸۵/۵   | ۱۲۶۹         |
| ۱۰۰۰            | ۳۶           | ۳۶           | ۸۷/۵   | ۱۲۹۱         |
| ۱۲۰۰            | ۳۶           | ۳۶           | ۸۷/۵   | ۱۳۵۱         |
| ۱۴۰۰            | ۳۶           | ۳۶           | ۸۷     | ۱۲۶۵         |
| ۱۶۰۰            | ۳۶           | ۳۶           | ۸۴/۲۵  | ۱۴۳۴         |
| ۱۸۰۰            | ۳۶           | ۳۶           | ۸۹     | ۱۴۷۷         |
| ۲۰۰۰            | ۳۶           | ۳۶           | ۸۷/۵   | ۱۵۲۳         |
| ۲۴۰۰            | ۳۶           | ۳۶           | ۸۹/۵   | ۱۷۶۲         |
| ۲۸۰۰            | ۳۶           | ۳۶           | ۹۰,۵   | ۱۹۵۲         |
| ۳۲۰۰            | ۳۶           | ۳۶           | ۹۳     | ۲۲۲۶         |
| ۳۶۰۰            | ۳۶           | ۳۶           | ۸۸/۷۸  | ۲۴۵۳         |
| ۳۸۰۰            | ۳۶           | ۳۶           | ۹۰/۲۵  | ۲۵۵۱         |
| ۴۰۰۰            | ۳۶           | ۳۶           | ۸۷/۵   | ۲۸۴۱         |

جدول ۵-۴ نتایج حاصل از اجرای روش پیشنهادی با پارامترهای (تعداد جهت، تعداد بین) = (۳۶، ۳۶)

جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □۱۰۴

را نشان می‌دهد. با توجه به این جدول روش سیفت بهبود یافته با بردار ویژگی به طول ۳۲۰۰، به کارایی برابر با ۰/۹۳ می‌رسد. شکل ۵-۳ ماتریس درهم‌ریختگی روش پیشنهادی با بردار ویژگی به طول ۳۲۰۰ را نشان می‌دهد. هم‌چنان‌که ملاحظه می‌شود با افزایش طول بردار ویژگی می‌توان به نتایج مناسبی رسید.

ex-Wang-bovw-DenseSIFT-3200 - mean accuracy: 0.930000



شکل ۵-۳: ماتریس درهم‌ریختگی سیفت فشرده بهبود یافته با طول بردار ویژگی ۳۲۰۰

### ۳-۳-۵. ارزیابی روش پیشنهادی برای ضبط اطلاعات مکانی

در این بخش نتایج حاصل از شبیه‌سازی روش پیشنهادی برای ضبط اطلاعات مکانی با استفاده از ویژگی‌های الگوی دودویی محلی گزارش می‌شود. لازم به ذکر است که آزمایش‌های انجام شده در این بخش روی پایگاه دادهٔ معیار OT انجام شده‌اند. پایگاه دادهٔ OT شامل ۲،۶۸۸ تصویر در ۸ دسته به شرح زیر است (Oliva and Torralba 2002).

جدول ۵-۵: شرح پایگاه دادهٔ OT

| وضوح تصاویر | تعداد تصاویر موجود در دسته | دسته تصویر       |
|-------------|----------------------------|------------------|
| ۲۵۰×۲۵۰     | ۳۶۰                        | ساحل             |
| ۲۵۰×۲۵۰     | ۳۲۸                        | جنگل             |
| ۲۵۰×۲۵۰     | ۳۷۴                        | کوهستان          |
| ۲۵۰×۲۵۰     | ۴۱۰                        | دشت              |
| ۲۵۰×۲۵۰     | ۲۶۰                        | بزرگ راه         |
| ۲۵۰×۲۵۰     | ۳۰۸                        | داخل شهر         |
| ۲۵۰×۲۵۰     | ۳۵۶                        | ساختمان‌های بلند |
| ۲۵۰×۲۵۰     | ۲۹۲                        | خیابان‌ها        |

جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □ ۱۰۵

شکل ۴-۵ نمونه‌هایی از تصاویر موجود در این پایگاه داده را نشان می‌دهد. در این پایگاه داده تقریباً همه تصاویر در بردارنده آسمان هستند. از چالش‌های مهم این پایگاه داده تنوع درون و برون دسته‌ای زیاد آن است.



شکل ۴-۵: برخی تصاویر پایگاه داده OT

جدول ۶-۵ نتایج حاصل از شبیه‌سازی روش پیشنهادی برای ضبط اطلاعات مکانی را با استفاده از پارامترهای مختلف روی پایگاه داده OT نشان می‌دهد.

**جدول ۵-۶: نتایج حاصل از اجرای روش پیشنهادی با پارامترهای مختلف**

| کارایی (%) | زمان (ثانیه) | WS×WS | تعداد مقیاس‌ها<br>(S) | طول بردار ویژگی |
|------------|--------------|-------|-----------------------|-----------------|
| ۷۹/۶       | ۱۴۲۵         | ۳۵×۳۵ | [1,2,3]               | ۱۰۰             |
| ۷۷/۸       | ۱۸۳۴         | ۴۰×۴۰ | [1,2,3]               | ۱۰۰             |
| ۸۲         | ۱۵۲۳         | ۳۵×۳۵ | [1,2,3]               | ۲۰۰             |
| ۸۴/۵       | ۲۰۳۶         | ۴۰×۴۰ | [1,2,3]               | ۲۰۰             |
| ۸۲/۶       | ۲۰۴۷         | ۴۰×۴۰ | [1,2,3]               | ۳۰۰             |
| ۸۳/۷       | ۲۳۳۶         | ۴۵×۴۵ | [1,2,3]               | ۳۰۰             |
| ۸۶         | ۱۴۵۳         | ۳۵×۳۵ | [1,2,3]               | ۴۰۰             |
| ۸۵         | ۲۱۵۶         | ۴۰×۴۰ | [1,2,3]               | ۴۰۰             |
| ۸۵/۲       | ۲۳۷۴         | ۴۵×۴۵ | [1,2,3]               | ۴۰۰             |
| ۸۶/۴       | ۱۴۵۵         | ۳۵×۳۵ | [1,2,3]               | ۵۰۰             |
| ۸۵/۳       | ۲۲۱۹         | ۴۰×۴۰ | [1,2,3]               | ۵۰۰             |
| 86/5       | ۱۵۹۲         | ۳۵×۳۵ | [1,2,3]               | ۶۰۰             |
| 85/8       | ۱۹۹۶         | ۴۰×۴۰ | [1,2,3]               | ۶۰۰             |
| 86         | 1558         | ۳۵×۳۵ | [1,2,3]               | ۷۰۰             |
| 87/5       | 1958         | ۴۰×۴۰ | [1,2,3]               | ۷۰۰             |
| 86/2       | 2574         | ۴۵×۴۵ | [1,2,3]               | ۷۰۰             |
| 87/2       | 2018         | ۴۰×۴۰ | [1,2,3]               | ۸۰۰             |
| 87/6       | 2620         | ۴۵×۴۵ | [1,2,3]               | ۸۰۰             |
| 87/5       | 1257         | ۳۵×۳۵ | [1,2,3]               | ۹۰۰             |
| 87/5       | 1344         | ۳۰×۳۰ | [1,2,3]               | ۱۰۰۰            |
| 87         | 1722         | ۳۵×۳۵ | [1,2,3]               | ۱۰۰۰            |
| 86/5       | 17400        | ۳۵×۳۵ | [1,2,3]               | ۱۱۰۰            |
| 86         | 1413         | ۳۰×۳۰ | [1,2,3]               | ۱۲۰۰            |
| 88         | 1910         | ۳۵×۳۵ | [1,2,3]               | ۱۲۰۰            |
| 87/8       | 2622         | ۴۰×۴۰ | [1,2,3]               | ۱۲۰۰            |
| ۸۹         | 2589         | ۴۰×۴۰ | [1,2,3]               | ۱۴۰۰            |
| 86/8       | 2900         | ۴۵×۴۵ | [1,2,3]               | ۱۴۰۰            |
| 86         | 1797         | ۳۰×۳۰ | [1,2,3]               | ۱۶۰۰            |
| 87/2       | 2205         | ۳۵×۳۵ | [1,2,3]               | ۱۶۰۰            |
| 85/7       | 27681        | ۴۰×۴۰ | [1,2,3]               | ۱۶۰۰            |
| 87/6       | 2633         | ۴۰×۴۰ | [1,2,3]               | ۱۸۰۰            |
| 88/2       | 3219         | ۴۵×۴۵ | [1,2,3]               | ۱۸۰۰            |
| 88         | 2048         | ۳۰×۳۰ | [1,2,3]               | ۲۰۰۰            |
| 87/5       | 2410         | ۳۵×۳۵ | [1,2,3]               | ۲۰۰۰            |
| 88         | 2737         | ۴۰×۴۰ | [1,2,3]               | ۲۰۰۰            |
| 88/2       | 2461         | ۳۰×۳۰ | [1,2,3]               | ۲۲۰۰            |
| 88         | 2664         | ۳۵×۳۵ | [1,2,3]               | ۲۲۰۰            |

### ادامه جدول ۵-۶: نتایج حاصل از اجرای روش پیشنهادی با پارامترهای مختلف

| کارایی (%) | زمان (ثانیه) | WS×WS | تعداد مقیاس‌ها (S) | طول بردار ویژگی |
|------------|--------------|-------|--------------------|-----------------|
| 89         | 2747         | ۴۰×۴۰ | [1,2,3]            | 2200            |
| 88/8       | 2681         | ۳۵×۳۵ | [1,2,3]            | 2400            |
| 88/5       | 3463         | ۴۰×۴۰ | [1,2,3]            | 2400            |
| 87/1       | 3741         | ۴۵×۴۵ | [1,2,3]            | 2400            |

باتوجه به جدول ۵-۶، اندازه بهینه برای پنجره‌هایی که ویژگی‌ها از آن‌ها استخراج می‌شوند، می‌تواند ۴۰×۴۰ و یا ۳۵×۳۵ در نظر گرفته شود.

با توجه به آزمایش‌های انجام شده روش پیشنهادی برای ضبط اطلاعات مکانی با بردارهای ویژگی به طول‌های متفاوت، نتایج مطلوبی را به دست می‌دهد. جدول ۵-۷ نتایج مقایسه روش پیشنهادی را با سایر روش‌ها روی پایگاه داده OT نشان می‌دهد. ملاحظه می‌شود روش پیشنهادی در رساله با طول بردار ویژگی کوچک‌تر منجر به تولید نتایج بهتری می‌شود.

### جدول ۵-۷: مقایسه نتایج حاصل از روش پیشنهادی با سایر روش‌ها

| روش                   | تعداد تصاویر آموزش | تعداد تصاویر آزمون | طول بردار ویژگی | کارایی |
|-----------------------|--------------------|--------------------|-----------------|--------|
| روش پیشنهادی در رساله | ۸۰۰                | ۱۸۸۸               | ۱۴۰۰            | ۸۹     |
| PHOG                  | ۸۰۰                | ۱۸۸۸               | ۳۴۰۰            | ۷۶/۴   |
| GIST                  | ۸۰۰                | ۱۸۸۸               | ۵۱۲             | ۸۰/۴   |
| Gabor LBP             | ۸۰۰                | ۱۸۸۸               | ۴۰۰۰            | ۷۹/۹   |
| Gradient LBP          | ۸۰۰                | ۱۸۸۸               | ۵۶۰۰            | ۷۳/۳   |
| SIFT                  | ۸۰۰                | ۱۸۸۸               | ۱۵۰۰            | ۸۲     |
| SIFT+pLSA             | ۸۰۰                | ۱۸۸۸               | ۱۵۰۰            | ۸۳/۵   |
| SIFT+SP               | ۸۰۰                | ۱۸۸۸               | ۱۵۰۰            | ۸۷     |
| SIFT+pLSA + SP        | ۸۰۰                | ۱۸۸۸               | ۱۵۰۰            | ۸۶/۸   |

روش PHOG از ترکیب روش‌های HOG و انطباق هرم مکانی<sup>۱</sup> حاصل شده است. در این روش ویژگی‌های تصویر با استفاده از توصیف‌گر هیستوگرام گرادیان‌های جهت‌دار استخراج می‌شود و برای افزودن اطلاعات مکانی به این ویژگی‌ها، از روش انطباق هرم مکانی استفاده می‌شود. روش GIST روشی سراسری است که مطابق آنچه در فصل دوم توضیح داده شد از آن برای توصیف برخی ویژگی‌های تصاویر طبیعت استفاده می‌شود. روش‌های Gabor LBP و Gradient LBP نیز روش‌هایی مبتنی بر ویژگی الگوی دودویی محلی

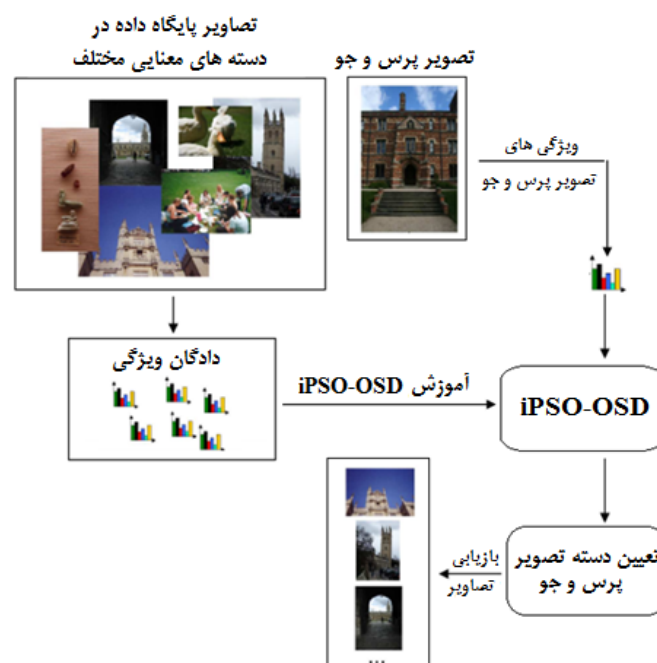
<sup>۱</sup> Spatial Pyramid Matching (SPM)

جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □ ۱۰۸

هستند. در روش Gabor LBP ویژگی‌های الگوی دودویی محلی با اعمال فیلترهای گابور و تقسیم‌بندی تصویر به منظور ضبط اطلاعات مکانی حاصل می‌شوند. در روش Gradient LBP ویژگی‌های الگوی دودویی محلی از تصویر گرادیان استخراج می‌شوند. در بقیه روش‌ها از توصیف‌گر سیفت برای استخراج ویژگی‌های تصویر استفاده می‌شود و با استفاده از ترکیب روش‌های گوناگون مانند pLSA و انطباق هرم مکانی سعی در ضبط اطلاعات مکانی و افزایش معنا می‌شود.

### ۴-۳-۵. ارزیابی iPSO-OSD در روش بسته واژگان

هم‌چنان‌که در فصل ۴ ذکر شد یکی از ارکان اصلی روش بسته واژگان دسته‌بندی است که از آن برای یافتن دسته معنایی تصاویر مشابه استفاده می‌شود. برای ارزیابی کارایی iPSO-OSD در بازیابی معنایی تصاویر از این دسته‌بند در قالب روش بسته واژگان استفاده شده‌است. شکل ۵-۵ شمای کلی روش پیشنهادی را نشان می‌دهد.



شکل ۵-۵: بازیابی تصاویر با استفاده از iPSO-OSD

شکل ۵-۵ نشان می‌دهد ابتدا ویژگی‌های همه تصاویر پایگاه داده از آن‌ها استخراج و بردارهای ویژگی ساخته می‌شود. ویژگی‌های استخراج شده از تصاویر می‌تواند سیفت، LBP یا HOG باشد. در این آزمایش‌ها از ویژگی‌های سیفت فشرده‌بهبود یافته در قالب روش بسته واژگان استفاده شده‌است. از ویژگی‌های استخراج شده برای آموزش و آزمون شبکه عصبی iPSO-OSD استفاده می‌شود. پس



جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □ ۱۰۹

از آموزش iPSO-OSD، این دسته‌بند می‌تواند دسته تصاویر پرس‌وجو را تعیین کند. با تعیین دسته تصاویر پرس‌وجو می‌توان شبیه‌ترین تصاویر به تصاویر پرس‌وجو را به عنوان خروجی نهایی بازیابی کرد. باتوجه به این نکات روش پیشنهادی برای بازیابی تصاویر روی سه پایگاه داده معروف OT، FP و LSP اجرا شده است که در ادامه نتایج حاصل از این پیاده‌سازی‌ها بیان می‌شود. قبل از بیان نتایج آزمایش‌ها توضیحی مختصر درباره پایگاه داده‌های FP و LSP بیا می‌شود. این پایگاه داده‌ها نسخه‌های تعمیم یافته‌ای از پایگاه داده OT هستند. پایگاه داده FP (Fei-Fei and Perona, 2005) از 3,864 تصویر خاکستری که در ۱۳ دسته قرار گرفته‌اند، تشکیل شده است. جدول ۵-۸ اطلاعات مربوط به این پایگاه داده را بیان می‌کند.

جدول ۵-۸: شرح پایگاه داده FP

| تعداد تصاویر موجود در دسته | دسته تصویر       |
|----------------------------|------------------|
| ۳۶۰                        | ساحل             |
| ۳۲۸                        | جنگل             |
| ۳۷۴                        | کوهستان          |
| ۴۱۰                        | دشت              |
| ۲۶۰                        | بزرگراه          |
| ۳۰۸                        | داخل شهر         |
| ۳۵۶                        | ساختمان‌های بلند |
| ۲۹۲                        | خیابان‌ها        |
| ۲۴۱                        | حومه شهر         |
| ۱۷۴                        | اتاق خواب        |
| ۱۵۱                        | آشپزخانه         |
| ۲۸۹                        | اتاق پذیرایی     |
| ۲۱۶                        | اداره            |

وجود تصاویر در دسته‌هایی مانند اتاق خواب، اداره، اتاق پذیرایی و ... باعث می‌شود تشخیص تشابه در چنین دسته‌هایی بسیار سخت باشد. اندازه تصاویر در این پایگاه داده به طور متوسط  $۲۵۰ \times ۳۰۰$  است. شکل ۵-۶ تعدادی از تصاویر موجود در این پایگاه داده را نشان می‌دهد.

پایگاه داده LSP حاوی 4,485 تصویر خاکستری است که در ۱۵ دسته قرار گرفته‌اند. اندازه متوسط تصاویر در این پایگاه داده  $۲۵۰ \times ۳۰۰$  است. جدول ۵-۹ اطلاعات مربوط به این پایگاه داده را نشان می‌دهد.

جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □ ۱۱۰



شکل ۵-۶: برخی تصاویر پایگاه داده FP

جدول ۵-۹: اطلاعات مربوط به پایگاه داده LSP

| تعداد تصاویر موجود در دسته | دسته تصویر       |
|----------------------------|------------------|
| ۳۶۰                        | ساحل             |
| ۳۲۸                        | جنگل             |
| ۳۷۴                        | کوهستان          |
| ۴۱۰                        | دشت              |
| ۲۶۰                        | بزرگ راه         |
| ۳۰۸                        | داخل شهر         |
| ۳۵۶                        | ساختمان‌های بلند |
| ۲۹۲                        | خیابان‌ها        |
| ۲۴۱                        | حومه شهر         |
| ۱۷۴                        | اتاق خواب        |
| ۱۵۱                        | آشپزخانه         |
| ۲۸۹                        | اتاق پذیرایی     |
| ۲۱۶                        | اداره            |
| ۳۱۵                        | فروشگاه          |
| ۳۱۱                        | صنعت             |

شکل ۵-۷ ماتریس درهم‌ریختگی حاصل از اجرای روش پیشنهادی را برپایگاه داده OT نشان می‌دهد. در ماتریس مذکور دسته‌های (۱) تا (۸) به ترتیب عبارتند از ساحل، جنگل، بزرگ‌راه، داخل شهر، کوهستان، دشت، خیابان و ساختمان‌های بلند.

| Output Class | 1             | 2             | 3             | 4             | 5             | 6             | 7             | 8             |                |
|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
|              | 265<br>14.1%  | 0<br>0.0%     | 4<br>0.2%     | 0<br>0.0%     | 1<br>0.1%     | 6<br>0.3%     | 0<br>0.0%     | 0<br>0.0%     | 96.0%<br>4.0%  |
|              | 0<br>0.0%     | 240<br>12.8%  | 0<br>0.0%     | 0<br>0.0%     | 2<br>0.1%     | 1<br>0.1%     | 0<br>0.0%     | 0<br>0.0%     | 98.8%<br>1.2%  |
|              | 3<br>0.2%     | 0<br>0.0%     | 194<br>10.3%  | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.1%     | 0<br>0.0%     | 98.0%<br>2.0%  |
|              | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.1%     | 221<br>11.7%  | 0<br>0.0%     | 0<br>0.0%     | 3<br>0.2%     | 1<br>0.1%     | 97.8%<br>2.2%  |
|              | 3<br>0.2%     | 7<br>0.4%     | 2<br>0.1%     | 0<br>0.0%     | 232<br>12.3%  | 9<br>0.5%     | 0<br>0.0%     | 0<br>0.0%     | 91.7%<br>8.3%  |
|              | 13<br>0.7%    | 12<br>0.6%    | 3<br>0.2%     | 0<br>0.0%     | 3<br>0.2%     | 243<br>12.9%  | 1<br>0.1%     | 0<br>0.0%     | 88.4%<br>11.6% |
|              | 0<br>0.0%     | 1<br>0.1%     | 4<br>0.2%     | 8<br>0.4%     | 0<br>0.0%     | 1<br>0.1%     | 184<br>9.8%   | 0<br>0.0%     | 92.9%<br>7.1%  |
|              | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 7<br>0.4%     | 1<br>0.1%     | 0<br>0.0%     | 1<br>0.1%     | 204<br>10.8%  | 95.8%<br>4.2%  |
|              |               |               |               |               |               |               |               |               |                |
|              | 93.3%<br>6.7% | 92.3%<br>7.7% | 93.3%<br>6.7% | 93.6%<br>6.4% | 97.1%<br>2.9% | 93.5%<br>6.5% | 96.8%<br>3.2% | 99.5%<br>0.5% | 94.7%<br>5.3%  |
| Target Class |               |               |               |               |               |               |               |               |                |
|              | 1             | 2             | 3             | 4             | 5             | 6             | 7             | 8             |                |

|              |    |              |               |                |               |                |                |               |                |               |               |                |                |                |                |
|--------------|----|--------------|---------------|----------------|---------------|----------------|----------------|---------------|----------------|---------------|---------------|----------------|----------------|----------------|----------------|
| Output Class | 1  | 173<br>6.9%  | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 1<br>0.0%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%      | 99.4%<br>0.6%  |                |
|              | 2  | 0<br>0.0%    | 231<br>9.2%   | 0<br>0.0%      | 10<br>0.4%    | 0<br>0.0%      | 2<br>0.1%      | 19<br>0.8%    | 0<br>0.0%      | 0<br>0.0%     | 1<br>0.0%     | 0<br>0.0%      | 0<br>0.0%      | 37.8%<br>12.2% |                |
|              | 3  | 0<br>0.0%    | 0<br>0.0%     | 228<br>9.1%    | 0<br>0.0%     | 0<br>0.0%      | 2<br>0.1%      | 4<br>0.2%     | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%      | 1<br>0.0%      | 37.0%<br>3.0%  |                |
|              | 4  | 0<br>0.0%    | 1<br>0.0%     | 0<br>0.0%      | 156<br>6.2%   | 1<br>0.0%      | 0<br>0.0%      | 1<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%      | 38.1%<br>1.9%  |                |
|              | 5  | 5<br>0.2%    | 0<br>0.0%     | 0<br>0.0%      | 3<br>0.1%     | 174<br>6.9%    | 0<br>0.0%      | 0<br>0.0%     | 4<br>0.2%      | 6<br>0.2%     | 0<br>0.0%     | 4<br>0.2%      | 6<br>0.2%      | 35.7%<br>14.3% |                |
|              | 6  | 1<br>0.0%    | 4<br>0.2%     | 8<br>0.3%      | 3<br>0.1%     | 0<br>0.0%      | 214<br>8.5%    | 5<br>0.2%     | 1<br>0.0%      | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%      | 2<br>0.1%      | 39.9%<br>10.1% |
|              | 7  | 1<br>0.0%    | 23<br>0.9%    | 5<br>0.2%      | 5<br>0.2%     | 0<br>0.0%      | 4<br>0.2%      | 228<br>9.1%   | 0<br>0.0%      | 1<br>0.0%     | 0<br>0.0%     | 1<br>0.0%      | 0<br>0.0%      | 0<br>0.0%      | 35.1%<br>14.9% |
|              | 8  | 0<br>0.0%    | 0<br>0.0%     | 0<br>0.0%      | 5<br>0.2%     | 11<br>0.4%     | 0<br>0.0%      | 1<br>0.0%     | 152<br>6.1%    | 2<br>0.1%     | 0<br>0.0%     | 3<br>0.1%      | 0<br>0.0%      | 0<br>0.0%      | 37.4%<br>12.6% |
|              | 9  | 0<br>0.0%    | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 1<br>0.0%      | 0<br>0.0%      | 0<br>0.0%     | 5<br>0.2%      | 209<br>8.3%   | 1<br>0.0%     | 0<br>0.0%      | 1<br>0.0%      | 2<br>0.1%      | 35.4%<br>4.6%  |
|              | 10 | 0<br>0.0%    | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 120<br>4.8%   | 6<br>0.2%      | 1<br>0.0%      | 3<br>0.1%      | 32.3%<br>7.7%  |
|              | 11 | 1<br>0.0%    | 0<br>0.0%     | 0<br>0.0%      | 2<br>0.1%     | 1<br>0.0%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 4<br>0.2%     | 100<br>4.0%    | 3<br>0.1%      | 16<br>0.6%     | 78.7%<br>21.3% |
|              | 12 | 0<br>0.0%    | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 4<br>0.2%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 2<br>0.1%     | 2<br>0.1%      | 109<br>4.3%    | 8<br>0.3%      | 37.2%<br>12.8% |
|              | 13 | 2<br>0.1%    | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 1<br>0.0%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%     | 9<br>0.4%     | 22<br>0.9%     | 12<br>0.5%     | 147<br>5.9%    | 76.2%<br>23.8% |
|              |    |              | 94.5%<br>5.5% | 89.2%<br>10.8% | 94.6%<br>5.4% | 84.8%<br>15.2% | 89.7%<br>10.3% | 96.4%<br>3.6% | 88.4%<br>11.6% | 93.8%<br>6.2% | 95.9%<br>4.1% | 97.6%<br>12.4% | 72.5%<br>27.5% | 82.6%<br>17.4% | 81.7%<br>18.3% |
|              |    | 1            | 2             | 3              | 4             | 5              | 6              | 7             | 8              | 9             | 10            | 11             | 12             | 13             |                |
|              |    | Target Class |               |                |               |                |                |               |                |               |               |                |                |                |                |

شکل ۵-۸: ماتریس درهم‌ریختگی روش پیشنهادی با دسته‌بند iPSO-OSD روی پایگاه داده FP

در ماتریس درهم‌ریختگی شکل ۵-۸ دسته‌های (۱) تا (۱۳) به ترتیب عبارتند از حومه شهر، ساحل، جنگل، بزرگراه، داخل شهر، کوهستان، دشت، خیابان، ساختمان‌های بلند، اداره، اتاق خواب و اتاق پذیرایی.

در ماتریس درهم‌ریختگی شکل ۵-۹ دسته‌های (۱) تا (۱۵) به ترتیب عبارتند از حومه شهر، ساحل، جنگل، بزرگراه، داخل شهر، کوهستان، دشت، خیابان، ساختمان‌های بلند، اداره، اتاق خواب، محل‌های صنعتی (کارخانه‌ها)، آشپزخانه، اتاق پذیرایی و فروشگاه.

جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □۱۱۳

|              |              |             |               |               |               |                |                |               |               |               |               |               |                |                |                |                |                |
|--------------|--------------|-------------|---------------|---------------|---------------|----------------|----------------|---------------|---------------|---------------|---------------|---------------|----------------|----------------|----------------|----------------|----------------|
| Output Class | 1            | 165<br>5.7% | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.0%      | 0<br>0.0%      | 1<br>0.0%      | 98.8%<br>1.2%  |                |
|              | 2            | 0<br>0.0%   | 206<br>7.1%   | 0<br>0.0%     | 9<br>0.3%     | 0<br>0.0%      | 1<br>0.0%      | 12<br>0.4%    | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.0%      | 0<br>0.0%      | 0<br>0.0%      | 90.0%<br>10.0% |                |
|              | 3            | 0<br>0.0%   | 0<br>0.0%     | 205<br>7.0%   | 0<br>0.0%     | 0<br>0.0%      | 3<br>0.1%      | 4<br>0.1%     | 0<br>0.0%     | 1<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%      | 3<br>0.1%      | 94.9%<br>5.1%  |                |
|              | 4            | 0<br>0.0%   | 0<br>0.0%     | 0<br>0.0%     | 160<br>5.5%   | 0<br>0.0%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.0%     | 0<br>0.0%      | 0<br>0.0%      | 0<br>0.0%      | 98.8%<br>1.2%  |                |
|              | 5            | 0<br>0.0%   | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.0%     | 165<br>5.7%    | 0<br>0.0%      | 0<br>0.0%     | 4<br>0.1%     | 3<br>0.1%     | 1<br>0.0%     | 3<br>0.1%     | 5<br>0.2%      | 10<br>0.3%     | 2<br>0.1%      | 11<br>0.4%     | 90.5%<br>9.5%  |
|              | 6            | 0<br>0.0%   | 0<br>0.0%     | 3<br>0.1%     | 2<br>0.1%     | 1<br>0.0%      | 217<br>7.4%    | 2<br>0.1%     | 1<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.0%      | 1<br>0.0%      | 0<br>0.0%      | 2<br>0.1%      | 94.3%<br>5.7%  |
|              | 7            | 0<br>0.0%   | 14<br>0.5%    | 6<br>0.2%     | 3<br>0.1%     | 0<br>0.0%      | 15<br>0.5%     | 230<br>7.9%   | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%      | 0<br>0.0%      | 0<br>0.0%      | 95.8%<br>4.2%  |
|              | 8            | 0<br>0.0%   | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.0%     | 4<br>0.1%      | 0<br>0.0%      | 0<br>0.0%     | 186<br>6.4%   | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%      | 1<br>0.0%      | 0<br>0.0%      | 1<br>0.0%      | 96.4%<br>3.6%  |
|              | 9            | 0<br>0.0%   | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 3<br>0.1%      | 0<br>0.0%      | 0<br>0.0%     | 6<br>0.2%     | 216<br>7.4%   | 0<br>0.0%     | 0<br>0.0%     | 10<br>0.3%     | 0<br>0.0%      | 1<br>0.0%      | 7<br>0.2%      | 98.9%<br>1.1%  |
|              | 10           | 0<br>0.0%   | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.0%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 149<br>5.1%   | 1<br>0.0%     | 3<br>0.1%      | 2<br>0.1%      | 1<br>0.0%      | 2<br>0.1%      | 93.7%<br>6.3%  |
|              | 11           | 0<br>0.0%   | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 110<br>3.8%   | 0<br>0.0%      | 2<br>0.1%      | 4<br>0.1%      | 0<br>0.0%      | 94.8%<br>5.2%  |
|              | 12           | 0<br>0.0%   | 0<br>0.0%     | 0<br>0.0%     | 2<br>0.1%     | 3<br>0.1%      | 0<br>0.0%      | 0<br>0.0%     | 1<br>0.0%     | 4<br>0.1%     | 0<br>0.0%     | 4<br>0.1%     | 158<br>5.4%    | 0<br>0.0%      | 5<br>0.2%      | 2<br>0.1%      | 98.3%<br>11.7% |
|              | 13           | 0<br>0.0%   | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.0%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 4<br>0.1%     | 4<br>0.1%     | 1<br>0.0%      | 114<br>3.9%    | 10<br>0.3%     | 3<br>0.1%      | 93.2%<br>16.8% |
|              | 14           | 2<br>0.1%   | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 2<br>0.1%      | 0<br>0.0%      | 0<br>0.0%     | 0<br>0.0%     | 0<br>0.0%     | 7<br>0.2%     | 35<br>1.2%    | 1<br>0.0%      | 11<br>0.4%     | 154<br>5.3%    | 5<br>0.2%      | 71.0%<br>29.0% |
|              | 15           | 0<br>0.0%   | 0<br>0.0%     | 0<br>0.0%     | 1<br>0.0%     | 8<br>0.3%      | 0<br>0.0%      | 0<br>0.0%     | 2<br>0.1%     | 0<br>0.0%     | 0<br>0.0%     | 2<br>0.1%     | 11<br>0.4%     | 1<br>0.0%      | 8<br>0.3%      | 161<br>5.5%    | 93.0%<br>17.0% |
|              |              |             | 98.8%<br>1.2% | 93.6%<br>6.4% | 95.8%<br>4.2% | 99.4%<br>10.6% | 97.8%<br>12.2% | 91.9%<br>8.1% | 92.7%<br>7.3% | 93.0%<br>7.0% | 96.4%<br>3.6% | 92.5%<br>7.5% | 68.8%<br>31.3% | 92.7%<br>17.3% | 79.7%<br>20.3% | 93.2%<br>16.8% | 90.9%<br>19.1% |
|              | 1            | 2           | 3             | 4             | 5             | 6              | 7              | 8             | 9             | 10            | 11            | 12            | 13             | 14             | 15             |                |                |
|              | Target Class |             |               |               |               |                |                |               |               |               |               |               |                |                |                |                |                |

شکل ۵-۹: ماتریس درهم‌ریختگی روش پیشنهادی با دسته‌بند iPSO-OSD روی پایگاه داده LSP

جدول ۵-۱۰: نتایج روش پیشنهادی برای بازیابی معنایی تصاویر را با سایر روش‌های موجود مقایسه

می‌کند.

#### جدول ۵-۱۰: مقایسه نتایج حاصل از روش پیشنهادی با سایر روش‌ها

| LSP   | FP   | OT   | پایگاه داده                                       |
|-------|------|------|---------------------------------------------------|
|       |      |      | روش                                               |
| 89/1  | 89/4 | 94/7 | روش پیشنهادی در رساله                             |
| -     | -    | 83/7 | روش پیشنهادی در (Oliva and Torralba 2001)         |
| -     | 65/2 | -    | روش پیشنهادی در (Fei-Fei and Perona 2005)         |
| 81/4  | 74/7 |      | روش پیشنهادی در (Lazebnik, Schmid and Ponce 2006) |
| 77    | -    | -    | روش پیشنهادی در (Van Gemert et al 2008)           |
| 85/6  | -    | 91/4 | روش پیشنهادی در (Boloivinou et al 2013)           |
| 84    | 86   | 91/1 | روش پیشنهادی در (Zhang et al 2014)                |
| ۸۴/۲  | ۸۶/۱ | 89.2 | روش پیشنهادی در (Zhou, Zhou and Hu 2013)          |
| ۸۴/۲  | -    | 88/1 | روش پیشنهادی در (Meng, Wang and Wu 2012)          |
| ۸۳/۳۲ | -    | -    | روش پیشنهادی در (Yang et al 2015)                 |
| 85/8  | -    | -    | روش پیشنهادی در (Zou et al 2016)                  |

#### ۴-۵. محدودیت‌های پژوهش

محدودیت‌های پژوهش در دو دسته محدودیت‌های مرتبط با پژوهش و محدودیت‌ها و موانع پیش‌روی پژوهشگر به شرح زیر بیان می‌شوند:

##### الف. محدودیت‌های پژوهش

۱- مسئله تشابه‌یابی تصویری مسئله‌ای ذاتاً دشوار است زیرا تصاویر از تنوع بسیاری برخوردارند و روش‌های موجود در مقایسه با توانمندی انسان در تشخیص تصاویر مشابه از کارایی کمتری برخوردارند. بنابراین تحقیق و پژوهش در این حوزه همچنان جای کار بسیاری دارد.

۲- رویکردها و روش‌های متعددی برای بازیابی محتوای محور تصاویر وجود دارد؛ انتخاب بهینه رویکرد با توجه به پایگاه داده‌ای که مورد نظر است، صورت می‌گیرد.

۳- با وجود تلاش برای رسیدن به نتایج مطلوب در حوزه بازیابی تصاویر، مشکل «فاصله معنایی» همچنان پابرجا است (شکل ۵-۱۰).



جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □ ۱۱۵



شکل ۵-۱۰: تصاویری از حیواناتی که بسیار به هم شبیه هستند

شکل ۵-۱۰ نمونه‌ای از تصاویری را نشان می‌دهد که بسیار به هم شبیه هستند و ممکن است یک سیستم رایانه‌ای قادر به تشخیص آن‌ها از یکدیگر نباشد این در حالی است که انسان‌ها به راحتی می‌توانند این تصاویر را از یکدیگر متمایز سازند.

### ب. محدودیت‌های پژوهش‌گر

۱- عدم در دسترس بودن سخت‌افزار مناسب برای پیاده‌سازی بسیاری از روش‌ها و ایده‌هایی که به ذهن می‌رسید.

۲- حجم زیاد داده‌های برای آزمون در محیط واقعی نیاز به سخت‌افزارهای قوی داشته که با توجه به محدودیت‌های مالی، این آزمون‌ها در شرایط محدود انجام گرفته‌است.

### ۵-۵. پیشنهادهایی برای تحقیقات آتی

پیشنهاد‌های آتی پژوهش در دو دسته پیشنهاد‌های پژوهشی و پیشنهاد‌های کاربردی مطرح می‌شود:

#### الف. پیشنهاد‌های پژوهشی

۱- چنانچه ذکر شد استفاده از دسته‌بند‌های قوی می‌تواند موجب افزایش کارایی سامانه‌های بازیابی تصاویر شود. بنابر این پژوهش دربارهٔ یافتن روش‌های یادگیری کاراتر و یا بهبود دسته‌بند‌های موجود می‌تواند موضوعی فراخور پژوهش برای ارزیابی و تحلیل بیشتر روش‌های پیشنهادی باشد. برای مثال استفاده از روش‌های یادگیری عمیق می‌تواند موضوعی قابل توجه باشد. همچنین معرفی روش‌های یادگیری کاراتر برای دسته‌کننده‌های مختلف نظیر شبکه‌های عصبی پرسپترون چند لایه و شبکه‌های عصبی تابع پایه شعاعی می‌تواند موضوع‌ای پژوهشی مناسبی باشند.

۲- از آنجا که مدل اصلی رساله برای بازیابی معنایی تصاویر بر روش روش بسته‌واژگان استوار است،

جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □ ۱۱۶

تحقیق و پژوهش دربارهٔ مراحل مختلف روش بستهٔ واژگان می‌تواند از اهمیت بالایی در این حوزه برخوردار باشد. به عنوان نمونه یکی از مراحل مهم روش بستهٔ واژگان خوشه‌بندی واژگان استخراج شده است. بنابراین پژوهش دربارهٔ روش‌های جدید خوشه‌بندی قابل اعمال روی داده‌های بزرگ<sup>۱</sup> می‌تواند از دیگر زمینه‌های تحقیقاتی در این حوزه به شمار رود. روش بستهٔ واژگان در تصاویر از روش روش بستهٔ واژگان در متون نشأت گرفته‌است. همچنان که در یک متن همهٔ واژگان مهم و تأثیرگذار نمی‌باشند، در تصویر نیز ممکن است بسیاری از واژگان در مشخص کردن محتوای معنایی تصویر تأثیری نداشته باشند. تحقیق و تفحص در خصوص روش‌های مختلف انتخاب واژگان مهم یا حذف واژگان زائد می‌تواند یکی دیگر از زمینه‌های تحقیقاتی به شمار رود.

۳- هر تصویر از مفاهیم و اشیای خاصی تشکیل شده‌است. به جای قطعه‌قطعه کردن تصویر و استخراج ویژگی‌ها از این قطعات، می‌توان با استفاده از الگوریتم‌های بخش‌بندی<sup>۲</sup> تصویر را به قسمت‌های معناداری تقسیم کرد و سپس ویژگی‌ها را از این قطعات استخراج کرد. این امر به شدت فاصلهٔ معنایی را کاهش می‌دهد. اما چالش بزرگ در اینجا عدم وجود روش بخش‌بندی با کارایی بالا روی انواع مختلف تصاویر است. بنابر این تحقیق در خصوص تولید الگوریتم‌های جدید بخش‌بندی تصویر و استفاده از آن‌ها در مرحلهٔ استخراج ویژگی‌ها در روش بستهٔ واژگان، می‌تواند به عنوان تحقیقات آتی بیان شود.

## ب. پیشنهاد‌های کاربردی

۱- سامانهٔ پیاده‌سازی شده پایه‌ای مناسب برای راه اندازی موتورهای جستجوگر است. در این سامانه از کدهای سطح بالای متلب استفاده شده است که به زبان C# کامپایل شده اند. بنابراین روش مشابهت‌سنجی و مقایسهٔ تصاویر و حتی دسته بند استفاده شده آن به سادگی قابل تغییر هستند و می‌توان با استفاده از سخت‌افزاری مناسب و افزودن یک خزش‌گر به سیستم فوق پژوهش‌های لازم برای تولید یک موتور جست‌وجو در ابعاد بزرگ را انجام داد.

۲- سامانهٔ طراحی شده می‌تواند در موارد خاصی مانند زیر به صورت عملی استفاده شود :

الف- تشابه‌یابی تصاویر چهره

ب- یافتن تصاویر مستهجن و فیلترسازی آن‌ها

ج- به کاربردن سیستم طراحی شده در تشخیص‌های پزشکی<sup>۳</sup>

د- استفاده از الگوریتم طراحی شده در مجموعه‌های هنری<sup>۱</sup> های تصاویر چهره بخش استنتاج فازی

---

<sup>۱</sup>Big data

<sup>۲</sup>Segmentation

<sup>۳</sup> Medical Diagnosis



جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □ ۱۱۷

سامانه به صورت جداگانه پیاده‌سازی شده است و در فرایندی نیمه‌برنامه‌نویسی شده نتایج آن به سامانه منتقل شده است. خودکارسازی این روند موجب افزایش سرعت شناسایی سامانه می‌شود.

## ۶-۵. نتیجه‌گیری

با توجه به اهداف اصلی پژوهش که بازیابی تصاویر مشابه بوده است، نتایج و یافته‌های پژوهش نیز در همین راستا مورد بررسی قرار می‌گیرد.

در فصل اول رساله کلیات پژوهش و ساختار رساله بیان شد. مبانی نظری پژوهش در فصل دوم به دقت مورد بررسی قرار گرفت و مشخص شد که روند بازیابی محتوای محور تصاویر به سوی استفاده از دسته‌بندی‌های مختلف در حال رشد است.

در فصل سوم روش‌های اصلی استفاده شده در رساله، مزایا و معایب هر یک از این روش‌ها توضیح داده شد. از مهم‌ترین عوامل مؤثر بر کارایی بازیابی محتوای محور تصاویر توصیف‌گر استفاده شده است. در این فصل توصیف‌گر سیفت به دقت تحلیل شد و نقاط ضعف و قوت این توصیف‌گر برای بازیابی تصاویر بررسی شد. همانطور که ذکر شد توصیف‌گر سیفت در برابر تغییرات روشنایی، هم‌پوشانی، چرخش و تغییر مقیاس تصاویر پایدار است. اما برای استفاده مستقیم از این توصیف‌گر در بازیابی تصاویر مشکلاتی نظیر سرعت پایین، حجم حافظه مصرفی بالا، حساس بودن به نویز و دقت کم ملاحظه شد. در این رساله روش جدیدی مبتنی بر توصیف‌گر سیفت برای بازیابی تصاویر پیشنهاد شد. همچنین در این فصل روش «بسته‌واژگان» به عنوان روش پایه‌ای استفاده شده در رساله توضیح داده شد. از مهم‌ترین معایب روش بسته‌واژگان از بین رفتن اطلاعات مکانی است. در این رساله روش جدیدی برای ضبط اطلاعات مکانی پیشنهاد شده است که موجب افزایش کارایی روش بسته‌واژگان می‌شود. همچنین در این به فصل تأثیر مهم دسته‌بند استفاده شده در روش بسته‌واژگان اشاره شد. در این رساله دسته‌بند جدیدی مبتنی بر شبکه‌های عصبی تابع پایه شعاعی معرفی شد.

در فصل چهارم روش‌های مختلف پیشنهادی در رساله توضیح داده شدند. در این فصل روش‌های مختلف بهبود توصیف‌گر سیفت شرح داده شد. آزمایش‌های انجام شده نشان دادند که با فشرده‌سازی و خوشه‌بندی ویژگی‌های استخراج شده به وسیله تشخیص‌گر نقطه کلیدی سیفت می‌توان به نتایج مطلوبی در بازیابی تصاویر رسید. برای بهبود بیشتر روش پیشنهادی اعمال توصیف‌گر روی همه پیکسل‌های تصویر بررسی شد. همچنین نشان داده شد که استفاده از مؤلفه‌های رنگی مختلف موجب افزایش قدرت تمایز توصیف‌گر پیشنهادی می‌شود. در ادامه چگونگی ضبط اطلاعات مکانی با استفاده از قرار دادن

جمع‌بندی نتایج و پیشنهاد برای پژوهش‌های آینده □ ۱۱۸

پنجره‌های همپوشانی‌کننده با اندازه‌های متفاوت و در مقیاس‌های مختلف نشان داده شد. پس از آن روش جدیدی برای آموزش شبکه عصبی تابع پایه شعاعی با استفاده از نسخه بهبود یافته الگوریتم PSO بیان شد. در آخر سامانه طراحی شده برای بازیابی تصاویر با استفاده از روش‌های پیشنهادی معرفی شد. در فصل پنجم نتایج ارزیابی روش‌های پیشنهاد شده در رساله روی دادگان تصویری مختلف ارائه شدند. روش سیف فشرده روی دادگان ونگ متوسط دقت بازیابی را در مقایسه با بهترین روش موجود ۴۰٪/۲ افزایش داد. همچنین روش سیف فشرده بهبود یافته روی دادگان ونگ متوسط دقت بازیابی را به ۸۹٪/۵ رساند یعنی در مقایسه با بهترین روش موجود کارایی به میزان ۱۴٪ افزایش یافت. با ضبط اطلاعات مکانی روی پایگاه داده OT دقت روش در مقایسه با بهترین روش‌های موجود ۳٪ افزایش یافت. همچنین با استفاده از دسته‌بندی‌کننده پیشنهادی دقت روش در مقایسه با بهترین روش‌های موجود روی پایگاه داده‌های OT، FP و LSP به ترتیب ۳۰٪/۶، ۳۰٪/۱ و ۳۰٪/۳ افزایش یافته‌است.

## فهرست منابع

پیروزمند، مریم. ۱۳۹۱. خصوصیات وب ایران. نخستین کنفرانس ملی مدیریت منابع اطلاعاتی وب. تهران

Abbasi, S. Mokhtarian, F. and Kittler. J. 1999. "Curvature scale space image in shape similarity retrieval". *Multimedia Systems*. Pp. 467-476.

Aksoy, S., Koperski, K., Tusk, C., Marchisio, G., & Tilton, J. C. (2005). Learning Bayesian classifiers for scene classification with a visual grammar. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(3), 581-589.

Allili. M. S. 2012. "Wavelet modeling using finite Mixtures of Generalized Gaussian Distributions: Application to Texture Discrimination and Retrieval", *IEEE Transactions on Image Processing*, 21 (4), pp. 1452-1464.

Alzu'bi, A., Amira, A., & Ramzan, N. (2015). Semantic content-based image retrieval: A comprehensive study. *Journal of Visual Communication and Image Representation*, 32, 20-54.

Arica, N. and Vural. F. T. Y. 2003. "BAS: A Perceptual Shape Descriptor Based on the Beam Angle Statistics". *Pattern Recognition Letters*, 24(9-10), pp: 1627–1639.

Bashar, M. K. Ohnishi, N. Matsumoto, T. Takeuchi, Y. Kudo, H. and Agusa. K. 2005. Image retrieval by pattern categorization using wavelet domain perceptual features with LVQ neural network. *Pattern Recognition Letters*. Pp. 2315–2335.

Barnard, K., Duygulu, P., Forsyth, D., De Freitas, N., Blei, D. M., & Jordan, M. I. (2003). Matching words and pictures. *The Journal of Machine Learning Research*, 3, 1107-1135.

Banerjee, M., Kundu, M. K., & Maji, P. (2009). Content-based image retrieval using visually significant point features. *Fuzzy Sets and Systems*, 160(23), 3323-3341.

Bimbo. A. del. 1999. "Visual Information Retrieval". Morgan Kaufmann Publishers, San Francisco, CA, USA.

Blei, D., Ng, A. and Jordan, M. 2003. "Latent Dirichlet allocation". In Lafferty, John. *Journal of Machine Learning Research* 3 (4–5): 993–1022.

Bolovinou, A., Pratikakis, I., Perantonis, S. Bag of spatio-visual words for context inference in scene classification. *Pattern Recognition* 46 (2013) 1039–1053.

Buch, A. Jain, S. and Pradhan, S. N. 2011. "Exploring a new direction in colour and texture based satellite image search and retrieval system". *Institute of Technology Nirma University, Ahmed Abad*, pp: 1-5.

Chaira, T. and Ray, A.K. 2005. "Fuzzy measures for color image retrieval". *Fuzzy Sets and Systems*. 150 (3) 545–560.

Castelli, V. and Bergman, L. D. 2002. "Image Databases: Search and Retrieval of Digital Imagery". *Wiley: New York*.

Chan, Shi-Kuo. and Hsu, Arding. 1992. "Image Information Systems: Where do we go from here?" *IEEE Trans. on Knowledge and Data Engineering*. 4(5), pp: 431-442.

Choy, S. K. and Tong, C. S. 2007. "Supervised texture classification using characteristic generalized Gaussian density". *J. Math. Imaging Vis.*, Vol. 29. No. 1, pp. 35-47.

Clark, S. 2013. The Importance of Images in Social Media.

<http://seanclark.com/social-media/the-importance-of-images-in-social-media/> (accessed 22 May)

Clerc, M. (2010). *Particle swarm optimization* (Vol. 93). John Wiley & Sons.

Cooper, S. The Top 20 Valuable Facebook Statistics, – Updated September 2016

<https://zephoria.com/top-15-valuable-facebook-statistics/> (accessed 20 October 2016)

Csurka, G., Dance, C., L.X. Fan, Willamowski, J. and Bray, C. 2004. "Visual categorization with bags of keypoints". *Proc. of ECCV International Workshop on Statistical Learning in Computer Vision*.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 1, pp. 886-893).

Dash, J. K., Mukhopadhyay, S., & Gupta, R. D. (2015). Content-based image retrieval using fuzzy class membership and rules based on classifier confidence. *IET Image Processing*, 9(9), 836-848.

de Almeida, Carlos W. D. De Souza, Renata M. C. R. Rodrigues, Carlos E. B. and Cavalcanti Junior. Nicomedes L. 2007. Image retrieval using the Curvature Scale Space technique and the Self Organizing Map model under Affine Transforms. *IEEE 7<sup>th</sup> International Conference on Hybrid Intelligent Systems*. Pp. 210–215.

de Souza Gazolli, K. A., & Salles, E. O. T. (2014). Exploring neighborhood and spatial information for improving scene classification. *Pattern Recognition Letters*, 46, 83-88.

de Ves, E., Benavent, X., Ruedin, A., Acevedo, D., & Seijas, L. (2010, August). Wavelet-based texture retrieval modeling the magnitudes of wavelet detail coefficients with a generalized Gamma distribution. In *Pattern Recognition (ICPR), 2010 20th International Conference on* (pp. 221-224).

Deselaers, T., Keysers, D., & Ney, H. (2008). Features for image retrieval: an experimental comparison. *Information Retrieval*, 11(2), 77-107.

Do, M. N. and Vetterli, M. 2002. "Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models". *IEEE Transactions on Multimedia*, 4 (4), pp. 517-527.

Duygulu, P., Barnard, K., de Freitas, J. F., & Forsyth, D. A. (2002). Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Computer Vision—ECCV 2002* (pp. 97-112). Springer Berlin Heidelberg.

Eitz, M., Hildebrand, K., Boubekeur, T., & Alexa, M. (2011). Sketch-based image retrieval: Benchmark and bag-of-features descriptors. *Visualization and Computer Graphics, IEEE Transactions on*, 17(11), 1624-1636.

ElAlami, M. E. (2011). A novel image retrieval model based on the most relevant features. *Knowledge-Based Systems*, 24(1), 23-32.

ElAlami, M. E. (2014). A new matching strategy for content based image retrieval system. *Applied Soft Computing*, 14, 407-418.

Fei-Fei Li; Perona, P. (2005). "A Bayesian Hierarchical Model for Learning Natural Scene Categories". 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 2.p. 524.

Fergus, R., Perona, P., & Zisserman, A. (2003, June). Object class recognition by unsupervised scale-invariant learning. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* (Vol. 2, pp. II-264). IEEE.

Fischler, M. A., & Elschlager, R. A. (1973). The representation and matching of pictorial structures. *IEEE Transactions on computers*, (1), 67-92.

FolkersA. and Samet.H. 2002. "Content based image retrieval using Fourier ddescriptots

o a Logo database". *Proc of the 16<sup>th</sup> Int. Conf. on Pattern Recognition*, vol. 3, pp. 521-524.

Fredembach, C., Schröder, M., & Süssstrunk, S. (2004). Eigenregions for image classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(12), 1645-1649.

Grauman, K. and Darrell, T. 2005. "The pyramid match kernel: discriminative classification with sets of image features". Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1.p. 1458.[doi:10.1109/ICCV.2005.239](https://doi.org/10.1109/ICCV.2005.239). ISBN 0-7695-2334-X.

Guo, Z.H., Zhang, L., Zhang, D. (2010): Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern Recognit.*43(3), 706–719

Gutierrez-Osuna, R. 2002. Pattern Analysis for Machine Olfaction: A Review. *IEEE Sens. J.*, 2, 189-202.

Garcia, A.Azuella, J.Riveron,E.and Pogrebnyak.E.2007. Image retrieval based on Neural Network and Wavelet transform coefficient. *Computation and Systems*, vol. 11, No. 2, pp. 145-156.

Grigorescu, S. E. Petkov, N. and Kruzinga.P. 2002. "Comparison of texture features based on Gabor filters". *IEEE Transactions on Image Processing*, 11 (10), pp. 1160-1167.

Guo, Z.H., Zhang, L., Zhang, D. (2010): Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern Recognit.*43(3), 706–719

Hadjidemetriou, E., Grossberg, E. M. and Nayar, S. K. 2004. "Multiresolution histograms and their use for recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Pp. 831–847.

Han, J. and Ma, K. 2007. "Rotation-invariant and scale-invariant Gabor features for texture image retrieval". *Image and Vision Computing*, pp. 1474–1481

Han,J.-H. Huang,D.-S.Lok,T.-M. and Lyu.M. R. 2005. A Novel Image Retrieval system Based on BPNeural Network. *Proceedings of International Joint Conference on Neural Networks, Montreal, Canada*. Pp. 1-4.

Hasegawa, M., & Tabbone, S. (2014). Amplitude-only log Radon transform for geometric invariant shape descriptor. *Pattern Recognition*, 47(2), 643-658.

Hiremath, P. S. and Pujari, J. 2007. "Enhancement of Image Retrieval by Image Attention based on Edge Features in a Multiresolution Grid Framework". *IEEE International Conference on Signal Processing and Communications*, pp. 892–895.

Huang, J. Kumar, S. Mitra, M. Zhu, W. and Zabih, R. 1997. "Image Indexing Using Color Correlograms". In *IEEE International Conference on Computer Vision and Pattern Recognition*, pp: 762–768, Puerto Rico.

Hussain, S. Hashmani, M. Moinuddin, M. Yoshida, M. and Kanjo, H. 2012. Image Retrieval based on Color and Texture Feature Using Artificial Neural Network. *Communications in Computer and Information Science*. Vol. 281, pp 501-511.

Hoffman, T. 1999. "Probabilistic Latent Semantic Analysis". Proc. of the Fifteenth Conference on Uncertainty in Artificial Intelligence.

Huang, P. W., & Dai, S. K. (2003). Image retrieval by texture similarity. *Pattern recognition*, 36(3), 665-679.

Irtaza, A., Jaffar, M. A., Aleisa, E., & Choi, T. S. (2014). Embedding neural networks for semantic association in content based image retrieval. *Multimedia tools and applications*, 72(2), 1911-1931.

Jain, S., Zaveri, T., & Patel, S. (2016, November). Weighted fusion of LBP and CCH features for effective content based image retrieval. In *Signal Processing and Communications (SPCOM), 2016 International Conference on* (pp. 1-5).

Janan, F., & Brady, M. (2015). Shape description and matching using integral invariants on eccentricity transformed images. *International Journal of Computer Vision*, 113(2), 92-112.

Janney, P. and Yu, Z. 2007. "Invariant features of local textures – a rotation invariant local texture descriptor", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–7.

Jhanwar, N., Chaudhuri, S., Seetharaman, G., & Zavidovique, B. (2004). Content based image retrieval using motif cooccurrence matrix. *Image and Vision Computing*, 22(14), 1211-1220.

Kang, S. and Park, S. 2009. A fusion neural network classifier for image classification. *Pattern Recognition Lett.* Pp. 789–793.

Kanimozhi, T., & Latha, K. (2015). An integrated approach to region based image retrieval

using firefly algorithm and support vector machine. *Neurocomputing*, 151, 1099-1111.

Khemchandani, R., & Saigal, P. (2015). Color image classification and retrieval through ternary decision structure based multi-category TWSVM. *Neurocomputing*, 165, 444-455.

Kiranyaz, S., Ferreira, M. and Gabbouj, M. 2008. "A generic shape/texture descriptor over multiscale edge field". *2-d walking ant histogram, IEEE Transactions on Image Processing*. pp. 377–391

Kokareh, M. Biswas, P. K. and Chatterji. B. N. 2005. "Texture image retrieval using new rotated complex wavelet filters",*IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(6), pp. 1168-1178.

Krishnan, Dr. Sheerin Banu,M.and Callins Christiyana.C. 2007. "Content Based Image Retrieval Using Dominant Color Identification Based on Foreground Objects". *International Conference on Computational Intelligence and Multimedia Applications*, vol. 3, pp: 190-194.

Kumar, K. T. M., & Stephen, E. (2016). Face Image and Information Retrieval Using LBP and Sparse Coding. *Digital Signal Processing*, 8(6), 151-153.

Lazebnik, S. Schmid, C. and Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* , volume 2, 2006, pages 2169–2178.

Lei. H. 2013. "A New Retrieval Method Based on YUV Color Information for Digital Libraries".*Third IEEE International Conference on Intelligent System Design and Engineering Applications (ISDEA)*, pp. 128-130.

Li, C., Huang, Y., & Zhu, L. (2017). Color texture image retrieval based on Gaussian copula models of Gabor wavelets. *Pattern Recognition*, 64, 118-129.

Li, Y. and Liu. G. 2011. Design Image Retrieval based on Contourlet Transform and Neural Networks. *IEEE International Conference on Management and Service Science*. pp. 1-4.

Liang, J. J., Qin, A. K., Suganthan, P. N., & Baskar, S. (2006). Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. *Evolutionary Computation*, IEEE Transactions on, 10(3), 281-295.

Liao, K., Liu, G., & Hui, Y. (2013). An improvement to the SIFT descriptor for image



representation and matching. *Pattern Recognition Letters*, 34(11), 1211-1220.

Lin, C. H., Chen, R. T., & Chan, Y. K. (2009). A smart content-based image retrieval system based on color and texture feature. *Image and Vision Computing*, 27(6), 658-665.

Liu, G. H., & Yang, J. Y. (2013). Content-based image retrieval using color difference histogram. *Pattern Recognition*, 46(1), 188-198.

Liu, K., Skibbe, H., Schmidt, T., Blein, T., Palme, K., Brox, T., & Ronneberger, O. (2014). Rotation-invariant HOG descriptors using Fourier analysis in polar and spherical coordinates. *International Journal of Computer Vision*, 106(3), 342-364.

Lowe, David G. (1999). "Object recognition from local scale-invariant features". *Proceedings of the International Conference on Computer Vision* 2. pp. 1150–1157.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.

Lu, T. and Chang, C. 2007. “ Color image retrieval technique based on color features and image bitmap”. *Information Processing and Management* 43 (2), pp. 461–472.

Lu, Z., Peng, Y., & Ip, H. H. (2010). Image categorization via robust pLSA. *Pattern Recognition Letters*, 31(1), 36-43.

Manipoonchelvi, P and Muneeswaran, K., 2011. “Significant region based image retrieval using curvelet transform”. *International Conference on Recent Advancements in Electrical, Electronics and Control Engineering*. Pp. 291 – 294.

Manjunath, B. S. Ohm, J. R. Vasudevan, V. V. and Yamada. A. 2001. "Color and Texture Descriptors". *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), PP: 703–715.

Marton, Z. C., Seidel, F., Balint-Benczedi, F., & Beetz, M. (2013). Ensembles of strong learners for multi-cue classification. *Pattern Recognition Letters*, 34(7), 754-761.

Mathiassen, J. R. Skavhaug, A. and Bo. K. 2002. "Texture Similarity Measure Using Kullback-Leibler Divergence between Gamma Distributions", *In: 7th. ECCV*, Denmark, pp. 133–147.

Martí, J., Freixenet, J., Batlle, J., & Casals, A. (2001). A new approach to outdoor scene description based on learning and top-down segmentation. *Image and Vision Computing*, 19(14), 1041-1055.

Meng, F. Guo, B. and Fang, Y. 2010. "Novel image retrieval method based on interest points". *3<sup>rd</sup> International Congress on Image and Signal Processing*. Pp. 1582-1585.

Meng, X., Wang, Z., & Wu, L. (2012). Building global image features for scene recognition. *Pattern Recognition*, 45(1), 373-380.

Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10), 1615-1630.

Moghaddam, H. A., Khajoie, T.T., Rouhi, A.H. and Tarzjan, M.S. 2005. " Wavelet correlogram: a new approach for image indexing and retrieval", *Pattern Recognition*. Pp. 2506–2518.

Mojsilovic, A., Gomes, J., & Rogowitz, B. E. (2002). Isee: Perceptual features for image library navigation. In *Electronic Imaging 2002* (pp. 266-277). International Society for Optics and Photonics.

Montazer, G. A., & Giveki, D. (2015). An improved radial basis function neural network for object image retrieval. *Neurocomputing*, 168, 221-233.

Montazer, G. A., Sabzevari, R., & Khatir, H. G. (2007). Improvement of learning algorithms for RBF neural networks in a helicopter sound identification system. *Neurocomputing*, 71(1), 167-173.

Nallaperumal, K., Banu, M. S., & Christiyana, C. C. (2007). Content based image indexing and retrieval using color descriptor in wavelet domain. In *Conference on Computational Intelligence and Multimedia Applications. International Conference on* (Vol. 3, pp. 185-189).

Nevatia. R. 1982. Machine Perception. Prentice-Hall.

Oja, T. Pietikainen, M. and Maenpaa, T. 2002. "Multi-resolution grayscale and rotation invariant texture classification with local binary patterns". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (7), pp. 971-987.

Ojala, T., Pietikäinen, M., Mäenpää, T. (2002): Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(7), 971–987

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic

representation of the spatial envelope. *International journal of computer vision*, 42(3), 145-175.

Oliva, A., & Torralba, A. (2002). Scene-centered description from spatial envelope properties. In *Biologically motivated computer vision* (pp. 263-272). Springer Berlin Heidelberg.

Olkiewicz K. A. and Kaczmar.U. M. 2010. Emotion-based Image Retrieval—an Artificial Neural Network Approach. *Proceedings of the International Multi conference on Computer Science and Information Technology* pp. 89–96.

Park, S. B. Lee, J. W. and Kim.S. K. 2004. Content-based image classification using a neural network. *Pattern Recognition Letters*. Pp. 287–300.

Park, S. Seo, K.and Jang. D.2005. Expert system based on artificial neural networks for content based image retrieval. *Expert systems with Applications*. Pp. 589–597.

Pass, G. Zabih, R.and Miller. J.1996, "Comparing Images Using Color Coherence Vectors. In *Proceedings of the fourth ACM international conference on Multimedia*, pp: 65–73.

Patheja, P. S. Wao, A. A. and Prakash Maurya.J. 2012. CBIR On Learning of Neural Network With Feedback Relevance. *International Journal of Research in Engineering & Applied Sciences*. Vol. 2. Pp. 1360-1370.

Penatti, O. A., Silva, F. B., Valle, E., Gouet-Brunet, V., & Torres, R. D. S. (2014). Visual word spatial arrangement for image retrieval and classification. *Pattern Recognition*, 47(2), 705-720.

Penatti, Otavio. A. B. Valle,Eduardo.andTorres. Ricardo.da. S. 2012. "Comparative Study of global color and texture descriptors for web image retrieval".*Int. J. Via.Commun. Image R*. pp. 359-380.

Patil, J. K., & Kumar, R. (2016). Analysis of content based image retrieval for plant leaf diseases using color, shape and texture features. *Engineering in Agriculture, Environment and Food*.

Pham, M. T., Mercier, G., Bombrun, L., & Michel, J. (2016). Texture and Color-based Image Retrieval Using the Local Extrema Features and Riemannian Distance. *arXiv preprint arXiv:1611.02102*.

Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. (2007, June). Object retrieval

with large vocabularies and fast spatial matching. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on* (pp. 1-8).

Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle swarm optimization. *Swarm intelligence*, 1(1), 33-57.

Prechelt, L. (1994). Proben1: A set of neural network benchmark problems and benchmarking rules.

Puzicha, J. Hofmann, T. and Buhmann. J. M. 1997. "Non-parametric similarity measures for unsupervised texture segmentation and image retrieval". *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.pp. 267-272.

Qi, X., & Han, Y. (2005). A novel fusion approach to content-based image retrieval. *Pattern Recognition*, 38(12), 2449-2465.

Qin, J., & Yung, N. H. (2010). Scene categorization via contextual visual words. *Pattern Recognition*, 43(5), 1874-1888.

Qin, J., & Yung, N. H. (2012). Feature fusion within local region using localized maximum-margin learning for scene categorization. *Pattern Recognition*, 45(4), 1671-1683.

Rahimi, M., & Moghaddam, M. E. (2015). A content-based image retrieval system based on Color Ton Distribution descriptors. *Signal, Image and Video Processing*, 9(3), 691-704.

Ramana, T. V., Rao, K. V., & Prasad, G. S. C. (2016). Literature Survey to Improve Image Retrieval Efficiency by Visual Attention Model. In *Information Systems Design and Intelligent Applications* (pp. 121-127). Springer India.

Randen, T. and Husoy, J. -H. 1999. "Filtering for texture classification: a comparative study". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21 (4), pp. 291-310.

Richard, J. Coeurjolly, D. and Baskurt. A. 2005. "Generalization of angular radial transform for 2D and 3D shape retrieval". *Pattern Recognition Letters*, vol. 26(14).

Safar, M. Shahabi, C. and Sun. X. 2000. "Image Retrieval by Shape: A Comparative Study". In *IEEE International Conference on Multimedia and Expo(I)*, pp:141-144.

Sai, N. S. T. and Patil, R. C. 2009. "New Feature Vector for Image Retrieval: Sum of Value of Histogram Bins".*International Conference on Advances in Computing, Control, & Telecommunication Technologies*, pp: 550-554.

Sadek, S., Al-Hamadi, A., Michaelis, B., & Sayed, U. (2009). Image Retrieval Using Cubic Splines Neural Networks. *International Journal of Video & Image Processing and Network Security (IJIPNS)*, 9(10), 17-22.

Serre, T., Wolf, L. and Poggio, T. 2005."Object Recognition with Features Inspired by Visual Cortex".2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 2.p. 994.[doi:10.1109/CVPR.2005.254](https://doi.org/10.1109/CVPR.2005.254). ISBN 0-7695-2372-2.

Shanmugavadivu, P., Sumathy, P., & Vadivel, A. (2016). FOSIR: Fuzzy-Object-Shape for Image Retrieval applications. *Neurocomputing*, 171, 719-735.

Shrivastava, N., & Tyagi, V. (2015). Region Based Image Retrieval Using Integrated Color, Texture and Shape Features. In *Information Systems Design and Intelligent Applications* (pp. 309-316). Springer India.

Shrivastava, N., & Tyagi, V. (2015). An integrated approach for image retrieval using local binary pattern. *Multimedia Tools and Applications*, 1-15.

Sim, D., Kim, H. and Park, R. 2001. "Fast texture description and retrieval of DCT-based compressed images", *Electronics Letters*. pp. 18–19.

Singh C. and Pooja. 2012. "An effective image retrieval using the fusion of global and local transforms based features". *Optics & Laser Technology*, 44, pp. 2249-2259.

Singhal, A., Luo, J., & Zhu, W. (2003, June). Probabilistic spatial context models for scene content understanding. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* (Vol. 1, pp. I-235). IEEE.

Sivic, J., Russell, B.C., Efros, A.A., Zisserman, A. and Freeman, W.T. 2005."Discovering objects and their location in images".*Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*.p. 370.[doi:10.1109/ICCV.2005.77](https://doi.org/10.1109/ICCV.2005.77). ISBN 0-7695-2334-X.

Song, T., & Li, H. (2013). Wavelbp based hierarchical features for image classification. *Pattern Recognition Letters*, 34(12), 1323-1328.

Song, X., Jiao, L. C., Yang, S., Zhang, X., & Shang, F. (2013). Sparse coding and classifier ensemble based multi-instance learning for image categorization. *Signal Processing*, 93(1),

1-11.

Sudderth, E. B., Torralba, A., Freeman, W. T., & Willsky, A. S. (2005, October). Learning hierarchical models of scenes, objects, and parts. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on* (Vol. 2, pp. 1331-1338). IEEE.

Sun, J., Zhang, X., Cui, J. and Zhou, L. 2006. "Image retrieval based on color distribution entropy". *Pattern Recognition Letters*. Pp. 1122–1126.

Stehling, R. O. Nascimento, M. A. and Falcao. A. X. 2002 "A Compact and Efficient Image Retrieval Approach Based on Border/Interior Pixel Classification". In *Proceedings of the 11<sup>th</sup> ACM International Conference on Information and Knowledge Management*, pp: 102–109, McLean, Virginia, USA.

Stricker, M. A. and Orengo. M. 1995, "Similarity of Color Images". In *Storage and Retrieval for Image and Video Databases (SPIE)*, pp: 381–392.

Swain, M. and Ballard. D. 1991. "Color Indexing". *International Journal of Computer Vision*, 7(1), PP: 11–32.

Subrahmanyam, M., Maheshwari, R. P., & Balasubramanian, R. (2012). Expert system design using wavelet and color vocabulary trees for image retrieval. *Expert Systems with Applications*, 39(5), 5104-5114.

Tian, X., Jiao, L., Liu, X., & Zhang, X. (2014). Feature integration of EODH and Color-SIFT: Application to image retrieval based on codebook. *Signal Processing: Image Communication*, 29(4), 530-545.

Tirilly, P., Claveau, V., & Gros, P. (2010, March). Distances and weighting schemes for bag of visual words image retrieval. In *Proceedings of the international conference on multimedia information retrieval* (pp. 323-332). ACM.

Torralba, A. B., & Oliva, A. (1999). Semantic organization of scenes using discriminant structural templates. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* (Vol. 2, pp. 1253-1258). IEEE.

Torres, R. da. S. Silva, C. G. da. Medeiros, C. B. and Rocha. H. V. da. 2003. "Visual Structures for Image Browsing". In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, pp:167–174, New Orleans, LA, USA.

Traina A. J. M. Train, A. Faloutsos, C. and Seeger. B. 2002. "Fast indexing and

visualization of metric data sets using slim-trees". *IEEE Transactions on Knowledge and Data Engineering*, Volume: 14, Issue: 2, pp: 244 – 260.

Tsai, H. H., Chang, B. M., & Liou, S. H. (2014). Rotation-invariant texture image retrieval using particle swarm optimization and support vector regression. *Applied Soft Computing*, 17, 127-139.

Tuytelaars, T., & Mikolajczyk, K. (2008). Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3), 177-280.

Utenpattanant, A., Chitsobhuk, O. and Khawne, A. 2006. "Color descriptor for image retrieval in wavelet domain" *International Conference on Advanced Communication Technology*, vol. 1, pp. 818–821.

Van Gemert, J. C., Veenman, C. J., Smeulders, A. W., & Geusebroek, J. M. (2010). Visual word ambiguity. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(7), 1271-1283.

Verdoolaege G. and Scheunders P. 2011. "Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination". *Int. J. Comput. Vis.*, vol. 95, no. 3, pp. 265-286.

Vipparthi, S. K., & Nagar, S. K. (2015). Directional local ternary patterns for multimedia image indexing and retrieval. *International Journal of Signal and Imaging Systems Engineering*, 8(3), 137-145.

Vogel, J., & Schiele, B. (2007). Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, 72(2), 133-157.

Vogel, J., & Schiele, B. (2004). Natural scene retrieval based on a semantic modeling step. In *Image and video Retrieval* (pp. 207-215). Springer Berlin Heidelberg.

Wang, C., & Huang, K. (2015). How to use Bag-of-Words model better for image classification. *Image and Vision Computing*, 38, 65-74.

Wang, J. Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(9), 947-963.

Wang, X. Y. Yu, Y. J. and Yang H. Y. 2011. "An effective image retrieval scheme using color, texture and shape features". *Journal of Computer Standards and Interfaces*. Vol. 33,

Issue.1. Pp. 59-68.

Wang, Y., & Gong, S. (2007, September). Conditional random field for natural scene image classification. In *Proceedings of the British Machine Vision Conference, Warwick*.

Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., & Li, J. (2014, November). Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 157-166). ACM.

Walia, E., & Pal, A. (2014). Fusion framework for effective color image retrieval. *Journal of Visual Communication and Image Representation*, 25(6), 1335-1348.

Wikipedia, Internet. 2013

<http://en.wikipedia.org/wiki/Internet>(دسترسی در ۱۳۹۲/۲/۱)

Wong, W.-T. and Hsu. S.-H .2006. Application of SVM and ANN for image retrieval. *European Journal of Operational Research*. Pp. 938–950.

Wong, W.-T. and Shih.F. Y. 2007. Shape based image retrieval using support vector machines, Fourier descriptors and self-organizing maps. *Information Sciences*. Pp. 1878-1891.

Wu, P., Manjunath, B.S., Newsam, S. and Shin, B.S. 2001 “A texture descriptor for browsing and similarity retrieval”. *Signal Processing: Image Communication*. Pp. 33–43

Wu,Y. and Wu.Y. 2009. "Shape based Image Retrieval Using Combining Global and Local Shape Features". *2nd International Congress on Image and Signal Processing*, pp. 1-5.

Wu, L., Hoi, S. C., & Yu, N. Semantics-preserving bag-of-words models for efficient image annotation. In *Proceedings of the 1st ACM workshop on large-scale multimedia retrieval and mining*. 2009. (pp. 19–26)

Yang, M. Kpalma, K. and Ronsin. J. 2010. “A survey of shape feature extraction techniques”. *Pattern Recognition*. Pp. 1–38.

Yang, H. Y., Li, Y. W., Li, W. Y., Wang, X. Y., & Yang, F. Y. (2014). Content-based image retrieval using local visual attention feature. *Journal of Visual Communication and Image Representation*, 25(6), 1308-1323.

Yang, Y. B., Zhu, Q. H., Mao, X. J., & Pan, L. Y. (2015). Visual feature coding for image



classification integrating dictionary structure. *Pattern Recognition*, 48(10), 3067-3075.

Yildizer, E., Balci, A. M., Jarada, T. N., & Alhadj, R. (2012). Integrating wavelets with clustering and indexing for effective content-based image retrieval. *Knowledge-Based Systems*, 31, 55-66.

Yuan, H. and Zhang, X. -P. 2004. "Texture image retrieval based on a Gaussian mixture model and similarity measure using a Kullback divergence". *IEEE International Conference on Multimedia and Expo*, vol. 3, pp. 1867-1870.

Zhang, J., Marszałek, M., Lazebnik, S. and Schmid, C. 2007. "Local Features and Kernels for Classification of Texture and Object Categories: a Comprehensive Study". *International Journal of Computer Vision* 73 (2): 213–238. doi:10.1007/s11263-006-9794-4.

Zhang, L., Zhang, L., Guo, Z., Zhang, D. (2010) : Monogenic-LBP: A new approach for rotation invariant texture classification. In: Proc. International Conference on Image Processing, pp. 2677–2680

Zhang, S., Tian, Q., Hua, G., Huang, Q., & Gao, W. (2014). ObjectPatchNet: Towards scalable and semantic image annotation and retrieval. *Computer Vision and Image Understanding*, 118, 16-29.

Zhang, C., Wang, S., Huang, Q., Liu, J., Liang, C., & Tian, Q. (2013). Image classification using spatial pyramid robust sparse coding. *Pattern Recognition Letters*, 34(9), 1046-1052.

Zhou, L., Zhou, Z., & Hu, D. (2013). Scene classification using a multi-resolution bag-of-features model. *Pattern Recognition*, 46(1), 424-433.

Zhu, Z., Chen, Q., & Zhao, Y. (2014). Ensemble dictionary learning for saliency detection. *Image and Vision Computing*, 32(3), 180-188.

Zou, J., Li, W., Chen, C., & Du, Q. (2016). Scene classification using local and global features with collaborative representation fusion. *Information Sciences*, 348, 209-226.

Zujovic, J., Pappas, T. N., & Neuhoff, D. L. (2013). Structural texture similarity metrics for image analysis and retrieval. *IEEE Transactions on Image Processing*, 22(7), 2545-2558.

## **ABSTRACT**

Advances in information technology and communications have affected all the aspects of human life. Without any doubt, finding similar data can be very effective in many fields. Nowadays, image acquisition technologies have enabled the creation of large image datasets. This has created many technical challenges for different processes. Searching digital images in large datasets is one of these challenges. It has attracted the attention of many researches in computer science.

In this thesis by having focused on Bag of Visual Words (BoVW) method, a new feature extraction method is proposed. Comparisons with other proposed methods show the superiority of our proposed method. In addition, by improving the proposed approach, its ability for semantic content based image retrieval has been evaluated. Furthermore, for further increasing the performance of our proposed approach, a new radial basis function neural network classifier is proposed. In this step, the proposed methods have been compared with the state-of-the-arts in the literature. The experimental results demonstrate the supremacy of the proposed approaches.

Finally, the proposed approaches for image retrieval have been implemented in the framework of a content based image retrieval system. The implemented system is responsive and it can be applied on the web. The experimental results on several benchmark datasets demonstrate the superiority of the proposed models in the thesis.

**Keywords.** Semantic content based image retrieval, SIFT, LBP, Bag of Visual Words, Radial basis function neural network classifier.



**Iranian Research Institute for  
Information Science and Technology (IRANDOC)  
Faculty of .....**

**Certification of Ph.D Dissertation / Master Thesis Defense**

This is to certify that on [**date month, day year**] the Jury of IRANDOC has approved the Ph.D Dissertation/ Master Thesis of Mr/Ms ..... entitled ..... . The final grade of the dissertation/ thesis is [**grade**] and it has been awarded [**rank**].

| No | the title                | first and last name | Academic Degree | University/institute | signature |
|----|--------------------------|---------------------|-----------------|----------------------|-----------|
| 1  | Supervisor               |                     |                 |                      |           |
| 2  | Adviser                  |                     |                 |                      |           |
| 3  | Co-adviser               |                     |                 |                      |           |
| 4  | Reviewer 1               |                     |                 |                      |           |
|    | Reviewer 2               |                     |                 |                      |           |
|    | Reviewer 3               |                     |                 |                      |           |
| 5  | Agent of Research Deputy |                     |                 |                      |           |



**Iranian Research Institute for  
Information Science and Technology (IRANDOC)  
Faculty of .....**

A Thesis/ Dissertation Presented for the Degree of  
Master of Science / Doctor of Philosophy in .....

**Title  
Title of Thesis/ Dissertation**

**By  
Davar Giveki**

**Supervisor  
Dr. Gholam Ali Montazer**

**Month& Year**