Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 2134395, 13 pages
https://doi.org/10.1155/2018/2134395

Research Article

A Novel Technique Based on Visual Words Fusion Analysis of Sparse Features for Effective Content-Based Image Retrieval

Muhammad Yousuf,[1] **Zahid Mehmood**,[1] **Hafiz Adnan Habib**,[2] **Toqeer Mahmood**,[2] **Tanzila Saba**,[3] **Amjad Rehman**,[4] **and Muhammad Rashid**[5]

[1]*Department of Software Engineering, University of Engineering and Technology, Taxila 47050, Pakistan*
[2]*Department of Computer Science, University of Engineering and Technology, Taxila 47050, Pakistan*
[3]*College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia*
[4]*College of Computer and Information Systems, Al-Yamamah University, Riyadh 11512, Saudi Arabia*
[5]*Department of Computer Engineering, Umm Al-Qura University, Makkah 21421, Saudi Arabia*

Correspondence should be addressed to Zahid Mehmood; zahid.mehmood@uettaxila.edu.pk

Received 16 July 2017; Accepted 4 February 2018; Published 6 March 2018

Academic Editor: Marco Perez-Cisneros

Copyright © 2018 Muhammad Yousuf et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Content-based image retrieval (CBIR) is a mechanism that is used to retrieve similar images from an image collection. In this paper, an effective novel technique is introduced to improve the performance of CBIR on the basis of visual words fusion of scale-invariant feature transform (SIFT) and local intensity order pattern (LIOP) descriptors. SIFT performs better on scale changes and on invariant rotations. However, SIFT does not perform better in the case of low contrast and illumination changes within an image, while LIOP performs better in such circumstances. SIFT performs better even at large rotation and scale changes, while LIOP does not perform well in such circumstances. Moreover, SIFT features are invariant to slight distortion as compared to LIOP. The proposed technique is based on the visual words fusion of SIFT and LIOP descriptors which overcomes the aforementioned issues and significantly improves the performance of CBIR. The experimental results of the proposed technique are compared with another proposed novel features fusion technique based on SIFT-LIOP descriptors as well as with the state-of-the-art CBIR techniques. The qualitative and quantitative analysis carried out on three image collections, namely, Corel-A, Corel-B, and Caltech-256, demonstrate the robustness of the proposed technique based on visual words fusion as compared to features fusion and the state-of-the-art CBIR techniques.

1. Introduction

Image retrieval on the basis of image contents has been a vigorous area of research in the last three decades [1]. Many approaches have been introduced regarding image retrieval on the basis of image contents [2, 3]. A text-based image retrieval system has two issues. Firstly, the annotation task takes a longer time, which makes it unfeasible for huge databases. Secondly, assigning keywords for image annotation is subjective. These two drawbacks led to the development of a new system, which is CBIR [2]. CBIR aims to develop techniques which can be used for extracting similar images from image archives. Current CBIR methods are further categorized as global and local features [1, 4, 5]. Low-level features such as color, texture, shape, and spatial layout form the basis of CBIR [3, 6–10]. The main problem with CBIR is the issue of the semantic gap [3, 11] prevailing among high-level image concepts and low-level image features. The bag-of-visual-words (BoVW) model is a standard way to scramble local features into a vector of fixed length. It is one of the most widely used image feature representation methods [12]. The BoVW framework was suggested for the first time in the text retrieval domain for the analysis of text documents. It has subsequently been used in applications of computer vision [12–17]. In this model, feature vectors are quantized into visual words to formulate a dictionary or codebook. Visual words are formulated by clustering the local features [18].

Figure 1: Images of two different semantic categories with close visual appearance and semantic layout.

Human eyes discriminate images based on their visual contents. When we apply a feature extraction technique to the images that have a similar visual appearance, it may produce close feature vectors values that reduce the performance of the CBIR. The images shown in Figure 1 belong to two different semantic categories. These images are visually as well as semantically similar to each other. When a machine learning technique like support vector machine (SVM) classifies such type of images, it is possible that some images may be wrongly classified due to their similar semantic or visual appearance, which reduces the performance of the CBIR system.

SIFT performs better in the case of scale changes and on invariant rotations. However, SIFT does not perform better when there are low contrast and illumination changes within an image [19]. LIOP performs better in cases of low contrast and illumination changes within an image [20]. SIFT even performs better when there is large rotation and scale changes, while LIOP does not perform well in such cases [20].

In this article, we propose a novel technique based on visual words fusion as well as features fusion of the SIFT and LIOP feature descriptors based on the bag-of-visual-words (BoVW) methodology in order to deal with the aforementioned issues. For each image collection, the images are categorized into training and test sets, and SIFT and LIOP features are extracted separately from each image in the sets. After that, k-means clustering algorithm [21] is applied to the extracted features that represent image features in the form of clusters. Each cluster is specified as a visual word, and the combination of visual words constitutes a dictionary. For the proposed technique based on visual words fusion of SIFT and LIOP descriptors, clustering is applied individually to the extracted SIFT and LIOP features that have produced two dictionaries. After that, both dictionaries are fused or integrated together which results in the fusion of SIFT and LIOP visual words. For the proposed technique based on features fusion of SIFT and LIOP descriptors, both extracted features are fused together. Subsequently, clustering is applied to the fused features that constitute a single dictionary. These visual words are used to formulate a histogram from each image in the training set. Following this, these histograms are used to train the SVM classifier. At the end, images are retrieved from an image collection by applying the similarity measure technique based on the Euclidean distance between the query image and the images stored in an image collection.

The main contributions of this research article are as follows:

(1) A novel image representation in the form of the visual words fusion of SIFT and LIOP feature descriptors based on the BoVW methodology

(2) A novel image representation in the form of the features fusion of SIFT and LIOP feature descriptors based on the BoVW methodology

(3) Reduction of the semantic gap between low-level features of an image and high-level semantic concepts

The remaining sections of this article are organized as follows: the relevant state-of-the-art CBIR techniques are briefly described in Section 2 entitled as "Related Work." The detailed methodology of the proposed technique is discussed in Section 3 entitled as "Proposed Methodology." Section 4 presents the details of the experiments and performance analysis on three image collections. Section 5 concludes the proposed technique.

2. Related Work

CBIR has been an active research area for the last three decades due to its wide range of applications in image retrieval techniques [22]. The term "content-based" refers to the fact that the search technique evaluates the actual contents of an image rather than using traditional image annotation techniques for image retrieval. The term "content" in this framework refers to texture, color, shape, or any other information that can be derived from the image itself. There are various types of image retrieval techniques which are based on texture, shape, color, and spatial layout [23, 24]. Different interest points based detectors and descriptors have been proposed for feature extraction in image retrieval techniques [25–30].

Liu et al. [7] propose a novel descriptor known as microstructure descriptor (MSD). MSD is determined by

underlying colors and edge orientation which perfectly depicts the image features. To retrieve the images effectively, the method assimilates color, texture, shape, and spatial layout information. However, this approach is inadequate for global properties of the image and is unable to exploit relations among positions of dissimilar entities in the proposed design. Mansoori et al. [2] also propose a CBIR technique based on a SIFT descriptor, a hue descriptor, and soft assignment. The SIFT is used for extracting keypoints, while local patches around them are described by applying SIFT and hue descriptors. The distinct vocabulary is created for each descriptor which is then quantized by applying a k-means clustering algorithm. In this model, the soft assignment is used instead of a hard assignment in order to overcome the forfeiture in quantization that can reduce retrieval performance. The proposed technique reveals enhanced performance in comparison with other comparable CBIR techniques. Chang et al. [6] present a novel framework for content-based image retrieval by investigating the particle swarm optimization algorithm (PSOA). The proposed technique extracts three kinds of features from each image, namely, color, texture, and shape features, to find the similarities between the query image and images from the catalog. It employs appropriate distance measure for each kind of feature utilized. The PSOA is incorporated to elevate the proposed technique via finding out close prime combinations among features and their corresponding similarity measurements. Shen and Wu [4] develop an innovative method for CBIR by merging color, spatial, and texture features of the image. A feature vector is formed by utilizing all three of these features. The CENsus transform hISTogram (CENTRIST) feature is used for spatial structure and a principle component analysis (PCA) is applied on CENTRIST for dimension reduction. This algorithm incorporates diverse density (DD) and multiple instance learning (MIL) to achieve objective occurrences. This technique produces better results when compared to the state-of-the-art CBIR techniques. However, a few limitations of this method have been found, leading to the conclusion that more research is needed in some aspects. Talib et al. [5] introduce a framework for CBIR by constructing a weighted dominant color (DC) descriptor. In order to extract semantic features, the descriptor assigns weights to each DC in the image. This technique overcomes the shortcomings of dominant color descriptor (DCD) and diminishes the consequence of image background during the image matching decision. The technique tends to increase the performance. Pedronette et al. [31] exploit the reranking technique for retrieving images based on their visual contents. The proposed technique improves the effectiveness of CBIR. The reranking method does not entail distance information among complete ranked lists or images of a given collection. The proposed technique counts on the ranked list that was generated by efficient indexing structures and it is considered appropriate for large image collections as it scales up very well.

Zheng et al. [32] embed multiple binary features at the indexing level for large scale image retrieval. The multi-IDF scheme models correlation between features. The Hamming embedding method is used as a matching verification method. In order to lessen the effect of incorrect detection and boost the accuracy of visual matching, SIFT visual words are integrated with binary features. Karakasis et al. [33] propose a CBIR technique that uses an affine moment in order to describe the invariants lying in the local areas of the image for the sake of image retrieval. The produced moments are incorporated into the BoVW model in order to produce detailed feature vectors. A setup of three different design elements is used. Firstly, affine moments are computed. Secondly, invariants are calculated over the results of the real image. In the last phase, the process of normalization is executed in order to increase the range of invariants. The second phase intends to improve the first phase, while the third phase improves the results of the second phase. Rahimi and Moghaddam [34] introduce a CBIR technique based on intraclass and interclass features. Intraclass features are called the distribution of color tone, whereas singular value decomposition (SVD) and complex wavelet transform produce interclass features. A self-organizing map (SOM) is given by these features based on the artificial neural network (ANN) in order to improve the performance of the CBIR. Rashno et al. [35] introduce a novel CBIR technique in which feature extraction is done through wavelet transform and color feature selection. In this scheme, each image in the image collection is represented using a feature vector which is comprised of texture features from wavelet transform and color features from RGB and HSV domains. For texture features based on wavelet transform, images are decomposed into four subbands and then a low-frequency subband is used as texture features. For color features, DCD is used for the quantization of the image, while color statistics and histogram features are calculated. The ant colony optimization technique is used for selecting relevant and unique features from the entire feature set which contains both color and texture features. Mehmood et al. [36] present a CBIR technique that utilizes local and global histograms of visual words from the image. Both histograms contain the information regarding the semantics of an image. The global histogram is constructed by utilizing the visual information of the whole image, whereas the local histogram is constructed by extracting visual information from a local rectangular region of the image. The local histogram contains the spatial information of the salient objects within the image. The proposed technique has significantly improved the performance of the CBIR.

Zhao et al. [38] propose a CBIR technique which integrates three image descriptors for identifying visual contents of the image. These features are based on color, texture, and shape. The association in the distribution of color range in an image is taken by color distribution entropy. The color level cooccurrence algorithm makes use of the texture level matrix in order to seize the recurrence of textures as descriptors. The shape, rotation, and rescaling are done by the use of invariant moments. Euclidean distance is used to compute the similarity measure. de Ves et al. [39] put forward a subjective methodology in order to reduce the semantic gap while incorporating concerned users' interests and their relative responses. The main intention is to achieve the objective of reducing the semantic gap using the PCA and regression

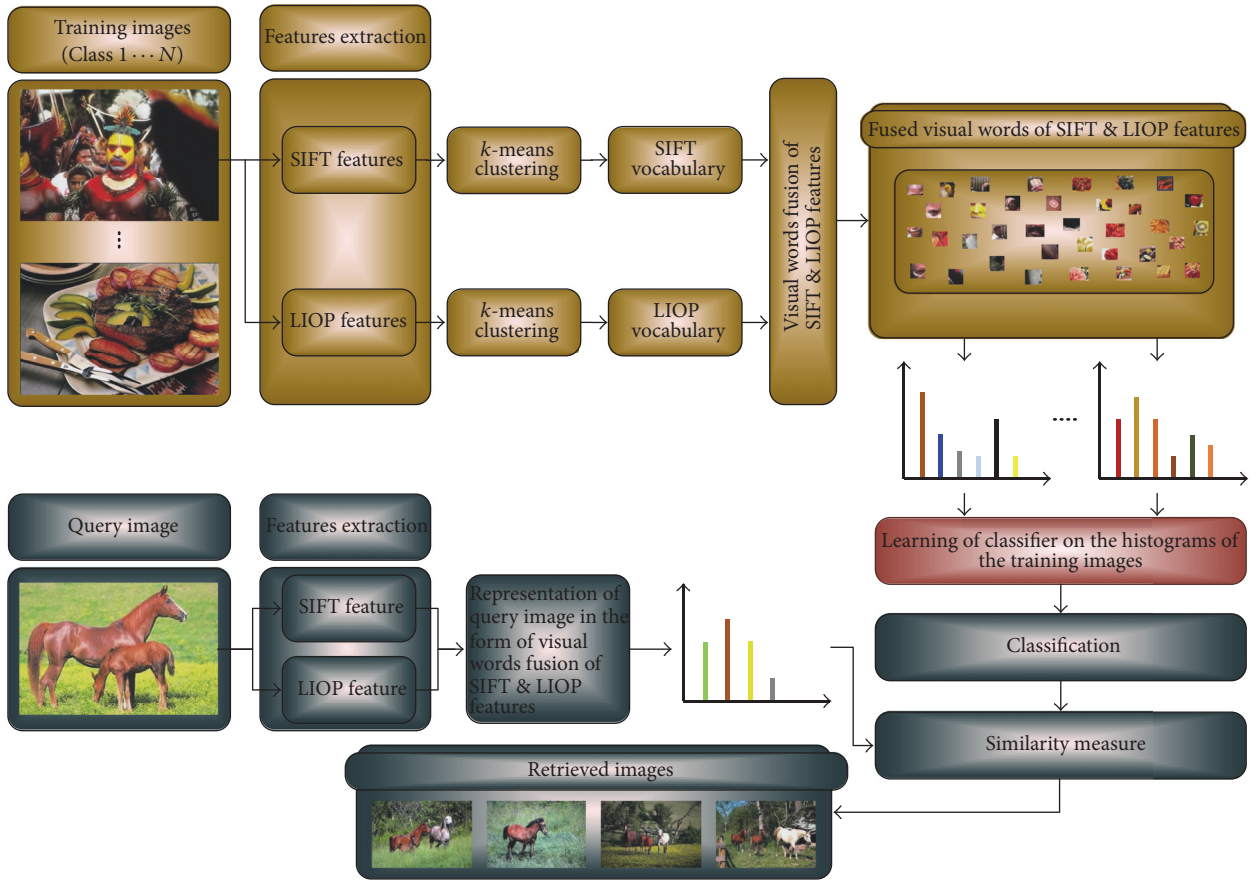


Figure 2: Block diagram of the proposed technique based on visual words fusion of SIFT and LIOP descriptors.

model. The former approach is responsible for rescaling the feature vectors, thereby reducing their dimensions, whereas the latter one is adjusted by the use of groups of nonoverlapped principal components. The local and dynamic nature of the proposed algorithm helps to achieve the intended results semantically. Xia et al. [40] present a CBIR technique to preserve the privacy of images in the cloud. While the cloud has solved the problem of low storage, at the same time, the privacy of users is highly concerned while outsourcing the images. The proposed technique exploits KNN in order to encode the visual features. These features are then utilized to compute the relevance, which in turn is utilized in the reranking procedure. In order to prevent the illegal copying and dissemination of retrieved images, the water-based protocol is exploited. Significant improvement has been observed in image search. The drawback of this technique lies in the lack of strength of the watermarking method.

3. Proposed Methodology

This section describes the detailed procedure of the proposed technique based on visual words fusion as well as features fusion of SIFT and LIOP descriptors based on the BoVW methodology for an effective CBIR. The block diagram of the proposed technique based on visual words fusion of SIFT and LIOP descriptors is shown in Figure 2.

The detailed procedure of the proposed technique is given as follows:

(1) For each image in the training and test sets, SIFT and LIOP features are computed.

(2) The SIFT features [48] are computed from each image over dense grid by applying the following mathematical equations:

$$h(t,i,j) = \left(k_i k_j * \overline{J}_t\right)\left(T + m\sigma\left[\frac{x_i}{y_i}\right]\right),$$
$$\overline{J}_t(x) = \omega_{\text{ang}} < \left(j(x) - \theta_t \left| j(x)\right|\right), \tag{1}$$

where σ is scale, θ is orientation, T is the center of the detected keypoint of the SIFT descriptor, m is descriptor magnification factor, J is gradient, h is the histogram of descriptors, ω_{ang} represents the angular velocity, and (x_i, y_j) represent the coordinate points of the (ith, jth) position. The kernels k_i and k_j are defined for a sample coordinate point (x, y) by the following mathematical equations:

$$k_i(x) = \frac{1}{\sqrt{2\pi}\sigma_{\text{win}}}\exp\left(\frac{-1(x - x_i)^2}{2\sigma_{\text{win}}^2}\right)\omega\left(\frac{x}{m\sigma}\right),$$
$$k_j(y) = \frac{1}{\sqrt{2\pi}\sigma_{\text{win}}}\exp\left(\frac{-1(y - y_i)^2}{2\sigma_{\text{win}}^2}\right)\omega\left(\frac{y}{m\sigma}\right), \tag{2}$$

where the side of the flat window is represented by σ_{win}.

(3) The LIOP features [20] are also computed from each image by applying the following mathematical equation:

$$\text{LIOP descriptor} = \left(\text{des}_1, \text{des}_2, \ldots, \text{des}_l\right),$$
$$\text{des}_l = \sum_{x \in \text{bin}_l} w(x)\,\text{LIOP}(x),$$
$$\text{where LIOP}(x) = \varphi\left(\gamma\left(P(x)\right)\right), \quad (3)$$
$$P(x) = \left(I(x_1), I(x_2), \ldots, I(x_n)\right) \in P^n,$$
$$w(x) = \sum_{i,j} \text{sgn}\left(\left|I(x_i) - I\left(x_j\right)\right| - T_{lp}\right) + 1.$$

In the above equation, for a sample point x_n, $I(x_n)$ represents the intensity of the nth neighboring sample, $P(x)$ is the N-dimensional feature vector of the intensities which represents the N neighboring sample points of a point x in the local patch, the mapping γ sorts the elements of the N-dimensional feature vector, preset threshold is represented by T_{lp}, sign function is represented by sgn, $w(x)$ represents the weighted function of the LIOP descriptor, the feature mapping function is represented by φ, and i, j represent the coordinate position of the nth sample point x_n.

(4) For the proposed technique based on visual words fusion of SIFT and LIOP descriptors, k-means [21] clustering technique is applied to the extracted features of SIFT and LIOP descriptors that produced two dictionaries. The resultant SIFT-based dictionary contains visual words of SIFT-based features, while LIOP-based dictionary contains visual words of LIOP-based features. Both dictionaries are fused together in order to perform visual words fusion of SIFT and LIOP features. The dictionary of each descriptor is formulated by applying the following mathematical equation on the extracted features of each descriptor:

$$R = \sum_{i=1}^{k} \sum_{x_l \in s_i} \left(x_l - u_i\right)^2, \quad (4)$$

where R represents the dictionary, u_i is the mean of all the points in the cluster s_i, and x_l represents the lth cluster or visual word.

After applying the clustering technique to extracted features of SIFT and LIOP descriptors, it produces two dictionaries that are represented by the following mathematical equations:

$$D_{\text{SIFT}} = \left\{v_{s1}, v_{s2}, v_{s3}, \ldots, v_{sn}\right\}$$
$$D_{\text{LIOP}} = \left\{v_{l1}, v_{l2}, v_{l3}, \ldots, v_{ln}\right\}, \quad (5)$$

where D_{SIFT} and D_{LIOP} are the resultant dictionaries that contain n visual words (i.e., $\{v_{s1}, v_{s2}, v_{s3}, \ldots, v_{sn}\}$ and $\{v_{l1}, v_{l2}, v_{l3}, \ldots, v_{ln}\}$) of SIFT and LIOP-based features, respectively.

After computing dictionaries for SIFT and LIOP feature descriptors, both dictionaries are concatenated which results in visual words fusion of both descriptors, represented mathematically as follows:

$$D_R = \left\{D_{\text{SIFT}}; D_{\text{LIOP}}\right\}, \quad (6)$$

where D_R is the resultant dictionary that contains SIFT and LIOP features in the form of fused visual words for more compact representation of image visual contents.

(5) For the proposed technique based on features fusion of SIFT and LIOP descriptors, SIFT and LIOP features are computed from each image, fused or integrated together, and at the end, k-means clustering technique [21] is applied to the fused features which produces a single dictionary.

The proposed technique based on visual words fusion of SIFT and LIOP descriptors results in better performance compared to the proposed technique based on features fusion of the SIFT and LIOP descriptors and the state-of-the-art CBIR techniques because the size of the dictionary representing visual contents of the images is twice as large compared to features fusion technique, which represents visual contents of the images by formulating a single dictionary.

(6) After applying the k-means [21] clustering technique, the visual contents of each image are now in the form of visual words. These visual words are used to build a histogram for each image.

(7) For image classification, the SVM classifier is selected along with Hellinger kernel [49] instead of the linear kernel. The learning of the SVM classifier is performed using histograms that are formulated from each image in the training set. The Hellinger kernel function is used with the SVM classifier because it explicitly computes the features map instead of computing the kernel values, while the classifier still remains linear. The mathematical representation of the Hellinger kernel function of the SVM on the normalized histograms is as follows:

$$K\left(n, n'\right) = \sum_i \sqrt{n(j)\,n'(j)}, \quad (7)$$

where n and n' represent the normalized histograms of each image.

(8) After training the proposed CBIR model, the testing of the proposed technique is performed by taking an image from the test set and applying the same aforementioned process to compute the histogram from the test image. The images are retrieved by measuring the similarity between the test image representation and training images stored in an image collection by applying the Euclidean distance formula.

4. Evaluation Metrics, Experimental Results, and Discussions

This section presents the performance measurements of the proposed technique. The performance is evaluated using precision, recall, and precision-recall (PR) curve parameters on Corel-A/1000 [50, 51], Corel-B/1500 [30], and Caltech-256 [52] image collections and the results are compared with the state-of-the-art CBIR techniques. All the results of the experiments are reported by performing each experiment 10 times. The dictionary size and features percentages per image are two important parameters that affect the performance of the proposed technique. Increasing the size of the dictionary at some certain level for compact representation of the visual contents of the images increases the performance

of the image retrieval, while larger sizes of the dictionary result in overfitting problem of CBIR. Similarly, in order to reduce the computational cost of the proposed technique that is slightly increased due to visual words fusion as well as the features fusion of SIFT and LIOP feature descriptors, performance analysis is carried out using different features percentages per image as reported in the subsequent sections.

The precision measures the specificity or accuracy while recall measures the sensitivity or robustness of the CBIR techniques. Both are mathematically represented by the following equations:

$$P = \frac{I_r}{I_t}, \qquad R = \frac{I_r}{I_s}, \tag{8}$$

where I_r represents the number of correctly retrieved images, I_t represents the total number of retrieved images, and I_s represents the total number of the images in a particular semantic category.

4.1. Analysis of the Evaluation Metrics on the Corel-A Image Collection. The Corel-A image collection is a subset of the WANG image collection. It contains 1000 images that are categorized into 10 semantic categories and the resolution of each image in this image collection is either 256 × 384 or 384 × 256. Each semantic category in this image collection contains 100 images. For a performance analysis of this image collection, images are divided into two sets known as training (70% images) and test (30% images) sets. The images in the training set are used to train the proposed model, while images in the test set are used to test the performance of the proposed model. In order to find the best performance of the proposed technique based on visual words fusion of SIFT and LIOP feature descriptors, different sizes of the dictionary (i.e., 20, 50, 100, 200, 400, 600, 800, 1000, and 1200) using different features percentages (i.e., 10%, 25%, 50%, 75%, and 100%) per image are formulated. The reason for selecting different features percentages per image is to reduce the computational cost that is slightly increased due to the visual words fusion as well as features fusions of SIFT and LIOP feature descriptors without affecting the performance of the proposed techniques.

The performance analysis in terms of the mean average precision (MAP) versus different sizes of the dictionary of the proposed technique based on features fusion of SIFT and LIOP descriptor that is compared with the MAP performance of the standalone SIFT and standalone LIOP techniques based on the BoVW methodology is presented in Figure 3. According to the experimental details shown in Figure 3, the best MAP performance of 82.90% is achieved on a dictionary size of 800 visual words using 75% feature per image. The proposed technique based on features fusion of SIFT and LIOP descriptors outperform in terms of the MAP performance as compared to the MAP performance of the standalone SIFT and standalone LIOP techniques on all the reported dictionary sizes.

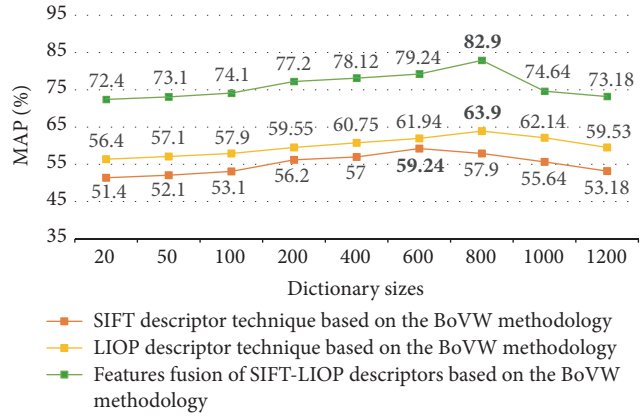


Figure 3: Performance comparison in terms of MAP performance between the proposed techniques based on features fusion, standalone SIFT, and standalone LIOP features on different sizes of the dictionary on the Corel-A image collection.

Table 1 presents the experimental details of the proposed technique based on visual words fusion of SIFT and LIOP descriptors on different reported sizes of the dictionary using different features percentages per image. The best MAP performance of 87.30% is achieved with a dictionary size of 800 visual words using 50% features per image. In order to verify the statistical significance of the experimental results of the proposed technique based on visual words fusion, the results of the statistical analysis are also reported in Table 1. The statistical results of the nonparametric Wilcoxon matched-pairs signed-rank test are also reported by comparing obtained MAP performance on dictionary size of 800 visual words with other reported dictionary sizes (20, 50, 100, 200, 400, 600, 800, 1000, and 1200) as well as with [36] using standard 95% confidence interval value. According to the statistical results of the nonparametric Wilcoxon matched-pairs signed-rank test, the proposed technique based on visual words fusion is statistically more effective because the value of P is less than the level of the significance (i.e., $\propto \leq 0.05$) for all the reported dictionary sizes.

In order to demonstrate the robustness of the proposed technique based on visual words fusion of SIFT and LIOP descriptors, its MAP performance is also compared with the MAP performance of the proposed technique based on features fusion as well as with the state-of-the-art CBIR techniques [36, 41–44], whose experimental details are shown in Figure 4 and Table 2. According to the experimental details, the proposed technique based on visual words fusion significantly outperforms in terms of the performance analysis as compared to its competitor CBIR techniques. The performance analysis in terms of the precision-recall (PR) curve as shown in Figure 5 is also carried with the state-of-the-art CBIR techniques [36, 37] which also demonstrate the robustness of the proposed technique based on visual words fusion of SIFT and LIOP descriptors on the Corel-A image collection.

The image retrieval results of the proposed technique based on visual words fusion of SIFT and LIOP descriptors for the semantic category "Beach" of the Corel-A image

Table 1: Statistical analysis and MAP performance of the proposed technique based on visual words fusion on different dictionary sizes and features percentages per image (bold values indicate the best performance).

Features percentages per image	Performance analysis in terms of the MAP performance (in %) on the different sizes of the dictionary								
	20	50	100	200	400	600	800	1000	1200
10%	74.30	74.60	76.10	78.50	79.30	80.70	84.50	76.60	75.30
25%	75.30	75.60	76.20	79.20	79.60	81.60	84.60	77.40	75.60
50%	75.60	76.00	76.50	80.60	81.70	82.30	**87.30**	77.50	76.30
75%	75.80	76.30	77.50	81.30	82.10	83.01	85.60	78.20	76.40
100%	76.00	77.60	79.10	81.70	82.30	83.60	85.70	78.50	77.30
MAP	75.40	76.10	77.10	80.20	81.00	82.24	**85.90**	77.64	76.18
Std. error	0.29	0.48	0.56	0.61	0.64	0.51	0.50	0.33	0.34
Std. deviation	0.66	1.09	1.25	1.36	1.43	1.14	1.12	0.74	0.77
Conf. interval	74.50–76.20	74.60–77.30	75.50–78.60	78.50–81.90	79.20–82.70	80.80–83.60	84.10–86.90	76.70–78.50	75.20–77.10
Statistical analysis using nonparametric Wilcoxon matched-pairs signed-rank test									
P value	0.043	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Z-value	2.023	2.03	2.02	2.02	2.02	2.03	2.02	2.02	2.02

Table 2: Performance analysis of the proposed technique based on visual words fusion on the Corel-A image collection which is reported using dictionary size of 800 visual words and features percentage of 50% per image (bold values indicate the best performance).

Semantic category	Proposed technique based on the visual words fusion	SIFT-LBP [41]	LGH-BoVW [36]	Color SIFT-EODH [42]	Poursistani et al. [43]	Yildizer et al. [44]
Africa	73.20	57.0	73.03	**74.60**	70.24	50.00
Beach	**75.00**	58.0	74.58	37.80	44.44	70.00
Buildings	**80.30**	43.0	80.24	53.90	70.80	20.00
Buses	95.50	93.0	95.84	**96.70**	76.30	80.00
Dinosaurs	**100**	98.0	97.95	99.00	**100**	90.60
Elephants	87.40	58.0	**87.64**	66.00	63.80	60.00
Flowers	98.30	83.0	85.13	92.00	92.40	**100.00**
Horses	**97.10**	68.0	86.29	87.00	94.70	80.00
Mountains	**83.80**	46.0	82.43	58.50	56.20	50.00
Food	**82.40**	53.0	78.96	62.20	74.50	20.00
MAP	**87.30**	65.7	84.21	72.77	74.34	62.00

collection and semantic categories "Sunset" and "Postcards" of the Corel-B image collection are shown in Figures 6, 9, and 10, respectively. The numeric value shown at the top of each image is the score of the respective image. The image shown at the top of each figure is the query image, while the rest of the images are the retrieved images that are obtained by applying the Euclidean distance formula between a score of the query image and scores of the retrieved images. The images whose numeric values are more close to the score of the query image are more identical to the query image which shows reduction of the semantic gap between low-level features of the image and high-level image semantic concepts and vice versa.

According to the experimental results shown in Table 2, the proposed techniques based on the visual words fusion of the SIFT and LIOP descriptors outperform in terms of the MAP performance as compared to the LGH-BoVW [36] technique as well as the state-of-the-art CBIR techniques [41–44] based on the BoVW methodology. For a dictionary size of N number of visual words, the proposed technique of this article represents visual contents of the images by assigning $2 \times N$ visual words due to the feature extraction from each image by applying two feature descriptors (i.e., SIFT and LIOP that formulate two dictionaries) as well as visual words of the resultant dictionary which contains the features of the SIFT and LIOP descriptors due to visual words fusion, while in case of the LGH-BoVW [36] technique, visual contents of the images are represented by assigning N number of visual words because single feature descriptor is applied on each image as well as visual words of the resultant dictionary which also contains the feature of the single descriptor.

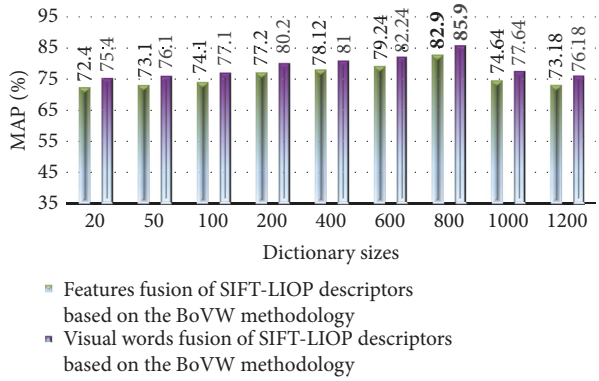


Figure 4: Performance comparison in terms of MAP performance between the proposed technique based on visual words fusion versus features fusion of SIFT and LIOP features techniques on different sizes of the dictionary on the Corel-A image collection.

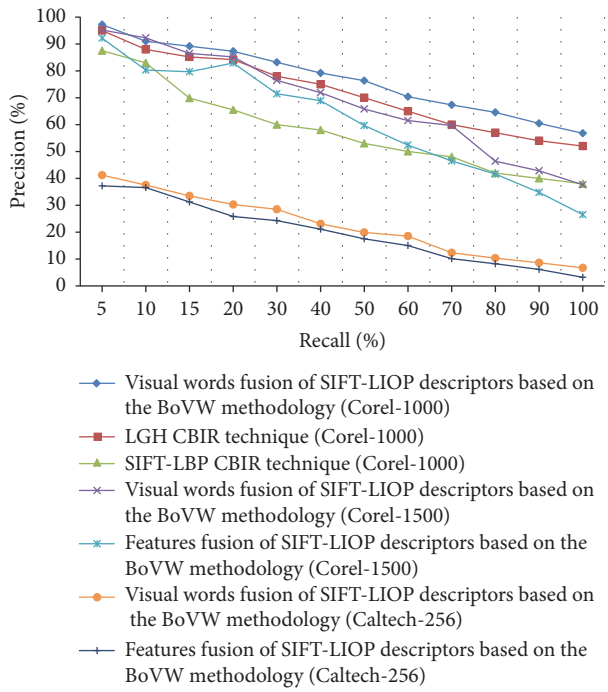


Figure 5: PR-curve comparison of the proposed technique based on visual words fusion versus features fusion of SIFT-LIOP descriptors as well as with the state-of-the-art CBIR techniques [36, 37] on the Corel-A, Corel-B, and Caltech-256 image collections.

4.2. Analysis of the Evaluation Metrics on the Corel-B Image Collection. The Corel-B image collection is a subset of the WANG image collection that contains images of different resolutions (i.e., 256 × 384, 384 × 256, 128 × 192, and 192 × 128). The total number of images in the Corel-B image collection is 1500; these are categorized into 15 semantic categories known as "Women," "Tigers," "Sunsets," "Postcards," "Caves," "Food," "Horses," "Mountains," "Flowers," "Elephants," "Dinosaurs," "Buses," "Buildings," and "Africa." The images are divided into two sets known as training (50% images) and test (50% images) sets for training and testing purposes. The performance analysis in terms of the MAP performance on different sizes of the dictionary is shown in Figures 7 and 8 and Table 3 for the proposed techniques based on visual words fusion, feature fusion, standalone SIFT, and standalone LIOP features based on the BoVW methodology. In the case of the proposed technique based on visual words fusion of SIFT and LIOP features, the best MAP performance of 85.20% is obtained with a dictionary size of 1000 visual words and using 50% features per image. The best MAP performance is achieved using the proposed technique based on features fusion of SIFT and LIOP features which is 82.96% with a dictionary size of 1000 visual words and using 75% features per image. According to the experimental details shown in Figures 7 and 8 and Table 3, the proposed technique based on visual words fusion outperforms as compared to the proposed technique based on features fusion, standalone SIFT, standalone LIOP, and the state-of-the-art CBIR techniques [3, 45] on a dictionary of all the reported sizes.

Table 3: Performance analysis of the proposed technique based on visual words fusion on the Corel-B image collection which is reported using dictionary size of 1000 visual words and features percentage of 50% per image (bold values indicate the best performance).

Performance measures	Proposed technique based on visual words fusion	GMM + mSpatiogram [45]	SQ + spatiogram [3]
MAP	**85.20**	74.10	63.95
Average recall	**17.00**	13.80	12.79

According to the experimental details shown in Figure 5 (experimental details provided earlier in Section 4.1), the performance measurement using PR-curve also demonstrates the robustness of the proposed technique based on visual words fusion that is compared with PR-curve of the proposed technique based on features fusion of SIFT and LIOP feature descriptors.

The results of image retrieval for the semantic categories "Sunset" and "Postcards" of the Corel-B image collection are shown in Figures 9 and 10.

4.3. Analysis of the Evaluation Metrics on the Caltech-256 Image Collection. We have also examined the performance analysis of the proposed technique on the Caltech-256 image collection [52]. The dimensions of each image in this collection are 300 × 200. There are 256 image semantic categories and each semantic category includes a minimum of 80 images. The total number of images in this collection is 30,607.

The performance analysis in terms of the MAP performance of the proposed technique based on features fusion, standalone SIFT, and standalone LIOP features techniques on different sizes of the dictionary is shown in Figure 11. According to the experimental details shown in Figure 11, the proposed technique based on features fusion of SIFT and LIOP descriptors performs better than the standalone SIFT and standalone LIOP features techniques based on the

Figure 6: Semantic category "Beach" of the Corel-A image collection shows a reduction of the semantic gap between retrieved images according to the query image.

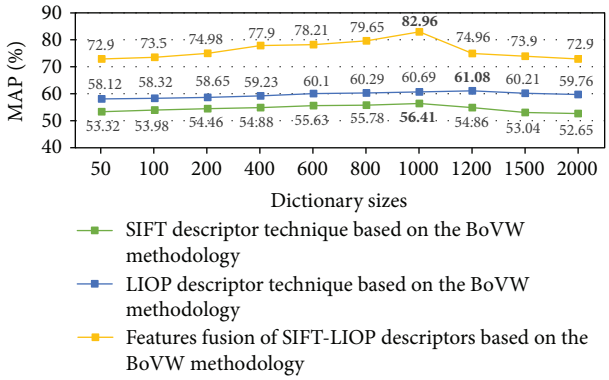


Figure 7: Performance comparison in terms of MAP performance between the proposed techniques based on features fusion, standalone SIFT, and standalone LIOP features on different sizes of the dictionary on the Corel-B image collection.

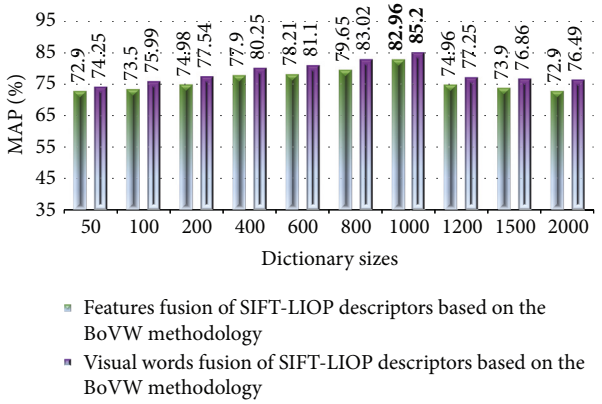


Figure 8: Performance comparison in terms of MAP performance between the proposed techniques based on visual words fusion versus features fusion of SIFT and LIOP features on different sizes of the dictionary on the Corel-B image collection.

BoVW methodology on a dictionary of all the reported sizes. According to the experimental details shown in Figure 12 and Table 4, the proposed technique based on visual words fusion of SIFT and LIOP descriptors outperforms in terms of MAP performance as compared to the features fusion technique and the state-of-the-art CBIR techniques [7, 46] on a dictionary of all the reported sizes. In the case of the proposed technique based on visual words fusion, the best MAP performance is achieved on a dictionary size of 1200 visual words that is 30.30%. The best MAP performance in case of features fusion technique is 25.82%, which is achieved on a dictionary size of 1500 visual words.

Figure 5 (experimental details provided earlier in Section 4.1) shows a comparison of performance analysis in terms of MAP performance using PR-curve between the proposed techniques based on visual words fusion versus

Table 4: Performance analysis of the proposed technique based on visual words fusion on the Caltech-256 image collection which is reported using dictionary size of 1200 visual words and features percentage of 75% per image.

Performance measures	Proposed technique based on visual words fusion	MN-ARM [7]	DCT [46]
MAP	30.30	28.21	23.91
Average recall	06.06	05.64	04.78

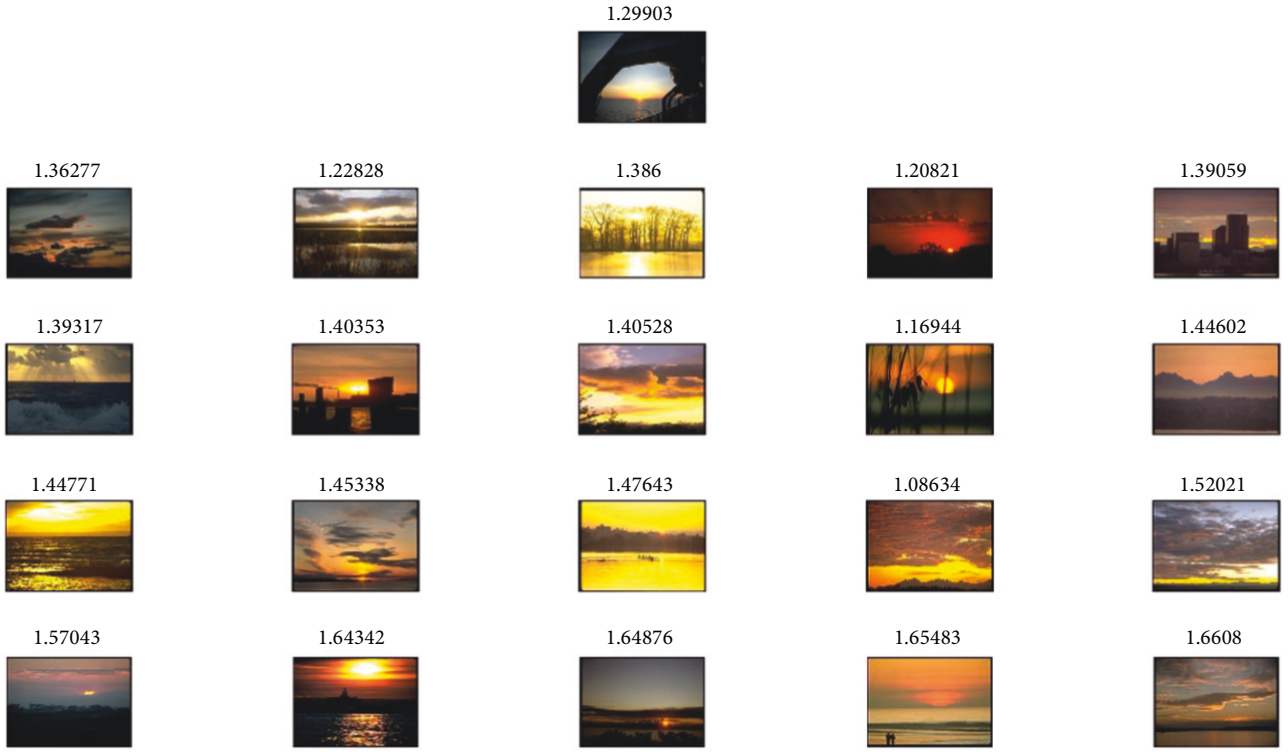


Figure 9: Semantic category "Sunset" of the Corel-B image collection shows a reduction of the semantic gap between retrieved images according to the query image.

Table 5: Performance analysis in terms of the computational complexity of complete framework.

Retrieved images	Proposed technique based on the visual words fusion of SIFT-LIOP	LGH technique [36]	FBWN technique [47]
Foremost-20	**0.7761**	0.7837	0.87

features fusion. According to the experimental details shown in Figure 5, the performance analysis using PR-curve also demonstrates the robustness of the proposed technique based on visual words fusion as compared to the proposed technique based on features fusion of SIFT and LIOP descriptors.

4.4. Performance Analysis in Terms of the Computational Complexity. All the experiments are performed on a Dell laptop with the following specifications: Intel (R) Pentium CPU B950 @ 2.10 GHz, 2.00 GB RAM, external SSD hard drive with a capacity of 120 GB, and Windows 7 64 bit operating system. The proposed technique is implemented in MATLAB R2015b and the dictionary is formulated offline by taking all the images of a training set. The performance is tested at runtime by taking a sample image from the test set using Corel-A image collection. The computational complexity (in seconds) of the complete framework from features computation to retrieved images is shown in Table 5 which is a proof of the robustness of the proposed technique in terms of the computational complexity as compared to the state-of-the-art CBIR techniques [36, 47].

5. Conclusions

The semantic gap between the low-level features of an image and high-level semantic concepts is an important issue that affects the performance of the CBIR. Increasing the size of the dictionary to represent visual contents of the images at some certain level increases the performance of the image retrieval, while larger sizes of dictionary tend to overfit. In this article, the proposed technique based on visual words fusion of SIFT and LIOP feature descriptors significantly improves the performance of the image retrieval by reducing the semantic gap issue of CBIR and assigning more visual words per image. The performance of the proposed technique based on visual words fusion is significantly improved as compared to the features fusion technique and the state-of-the-art CBIR techniques because the size of the dictionary to represent visual contents of the images is twice as large compared to the feature fusion technique. Additionally, the resultant dictionary contains features of the SIFT and LIOP descriptors in the form of visual words as compared to the state-of-the-art CBIR techniques. In order to reduce the computational cost of the proposed technique, which is slightly increased due to the fusion of SIFT and LIOP feature descriptors, different feature percentages per image are suggested without affecting the performance of the proposed technique.

Figure 10: Semantic category "Postcards" of the Corel-B image collection shows a reduction of the semantic gap between retrieved images according to the query image.

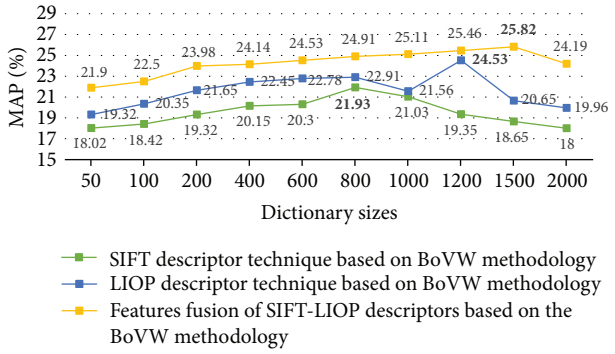


Figure 11: Performance comparison in terms of MAP performance between the proposed techniques based on features fusion, standalone SIFT, and standalone LIOP features on different sizes of the dictionary on the Caltech-256 image collection.

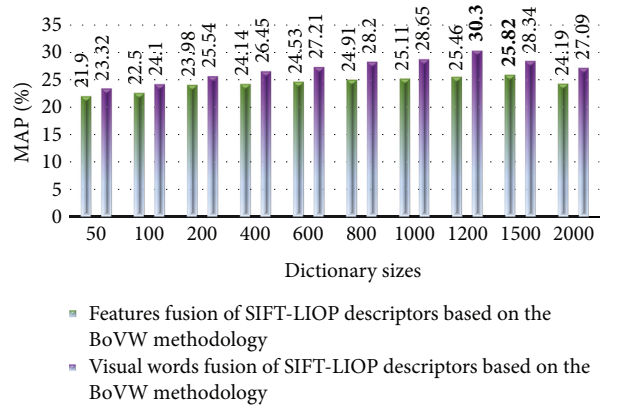


Figure 12: Performance comparison in terms of MAP performance between the proposed techniques based on visual words fusion versus features fusion of SIFT and LIOP features on different sizes of the dictionary on the Caltech-256 image collection.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All the authors contributed equally to this work.

Acknowledgments

This work was partially supported by the Machine Learning Research Group, Prince Sultan University Riyadh, Saudi Arabia [RG-CCIS-2017-06-02]. The authors are grateful for this financial support.

References

[1] N. Shrivastava and V. Tyagi, "Content based image retrieval based on relative locations of multiple regions of interest using selective regions matching," *Information Sciences*, vol. 259, pp. 212–224, 2014.

[2] N. S. Mansoori, M. Nejati, P. Razzaghi, and S. Samavi, "Bag of visual words approach for image retrieval using color information," in *Proceedings of the 2013 21st Iranian Conference on Electrical Engineering, ICEE 2013*, Mashhad, Iran, May 2013.

[3] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognition*, vol. 46, no. 1, pp. 188–198, 2013.

[4] G.-L. Shen and X.-J. Wu, "Content based image retrieval by combining color texture and CENTRIST," in *Proceedings of the Constantinides International Workshop on Signal Processing (CIWSP '13)*, pp. 1–4, January 2013.

[5] A. Talib, M. Mahmuddin, H. Husni, and L. E. George, "A weighted dominant color descriptor for content-based image retrieval," *Journal of Visual Communication and Image Representation*, vol. 24, no. 3, pp. 345–360, 2013.

[6] B.-M. Chang, H.-H. Tsai, and W.-L. Chou, "Using visual features to design a content-based image retrieval method optimized by particle swarm optimization algorithm," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 10, pp. 2372–2382, 2013.

[7] G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, "Image retrieval based on micro-structure descriptor," *Pattern Recognition*, vol. 44, no. 9, pp. 2123–2133, 2011.

[8] M. E. Elalami, "A new matching strategy for content based image retrieval system," *Applied Soft Computing*, vol. 14, pp. 407–418, 2014.

[9] G. W. Jiji and P. J. Durairaj, "Content-based image retrieval techniques for the analysis of dermatological lesions using particle swarm optimization technique," *Applied Soft Computing*, vol. 30, pp. 650–662, 2015.

[10] X.-Y. Wang, Y.-J. Yu, and H.-Y. Yang, "An effective image retrieval scheme using color, texture and shape features," *Computer Standards & Interfaces*, vol. 33, no. 1, pp. 59–68, 2011.

[11] X.-Y. Wang, Y.-W. Li, H.-Y. Yang, and J.-W. Chen, "An image retrieval scheme with relevance feedback using feature reconstruction and SVM reclassification," *Neurocomputing*, vol. 127, pp. 214–230, 2014.

[12] C. Tsai, "Bag-of-words representation in image annotation: A Review," *ISRN Artificial Intelligence*, vol. 2012, pp. 1–19, 2012.

[13] Z. Mehmood, T. Mahmood, and M. A. Javid, "Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine," *Applied Intelligence*, vol. 48, no. 1, pp. 166–181, 2017.

[14] Z. Mehmood, S. M. Anwar, and M. Altaf, "A novel image retrieval based on rectangular spatial histograms of visual words," *Kuwait Journal of Science*, vol. 45, no. 1, pp. 54–69, 2018.

[15] N. Ali, K. B. Bajwa, R. Sablatnig, and Z. Mehmood, "Image retrieval by addition of spatial information based on histograms of triangular regions," *Computers & Electrical Engineering*, pp. 539–550, 2016.

[16] T. Mahmood, A. Irtaza, Z. Mehmood, and M. Tariq Mahmood, "Copy–move forgery detection through stationary wavelets and local binary pattern variance for forensic analysis in digital images," *Forensic Science International*, vol. 279, pp. 8–21, 2017.

[17] T. Mahmood, Z. Mehmood, M. Shah, and Z. Khan, "An efficient forensic technique for exposing region duplication forgery in digital images," *Applied Intelligence*, pp. 1–11, 2017.

[18] Z. Liu, H. Li, L. Zhang, W. Zhou, and Q. Tian, "Cross-indexing of binary SIFT codes for large-scale image search," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2047–2057, 2014.

[19] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 1150–1157, IEEE, Kerkyra, Greece, September 1999.

[20] Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 603–610, Barcelona, Spain, November 2011.

[21] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society*, vol. 28, no. 1, pp. 100–108, 1979, Series C (Applied Statistics).

[22] J.-M. Guo, H. Prasetyo, and J.-H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 466–481, 2015.

[23] T. Mahmood, T. Nawaz, R. Ashraf et al., "A survey on block based copy move image forgery detection techniques," in *Proceedings of the International Conference on Emerging Technologies (ICET '15)*, pp. 1–6, Peshawar, Pakistan, December 2015.

[24] T. Mahmood, T. Nawaz, Z. Mehmood, Z. Khan, M. Shah, and R. Ashraf, "Forensic analysis of copy-move forgery in digital images using the stationary wavelets," in *Proceedings of the 6th International Conference on Innovative Computing Technology, INTECH 2016*, pp. 578–583, Dublin, Ireland, August 2016.

[25] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[26] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.

[27] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 886–893, IEEE, San Diego, Calif, USA, June 2005.

[28] H. Bay, T. Tuytelaars, and L. van Gool, "SURF: speeded up robust features," in *European conference on computer vision*, Lecture Notes in Computer Science, pp. 404–417, Springer, Berlin, Germany, 2006.

[29] D. J. Mankowitz and S. Ramamoorthy, "BRISK-based visual feature extraction for resource constrained robots," in *RoboCup 2013: Robot World Cup XVII*, S. Behnke, M. Veloso, A. Visser, and R. Xiong, Eds., vol. 8371 of *Lecture Notes in Computer Science*, pp. 195–206, Springer, Berlin, Germany, 2014.

[30] Z. Mehmood, F. Abbas, T. Mahmood, M. A. Javid, A. Rehman, and T. Nawaz, "Content-based image retrieval based on visual words fusion versus features fusion of local and global features," *Arabian Journal for Science and Engineering*, pp. 1–20, 2018.

[31] D. C. G. Pedronette, J. Almeida, and R. D. S. Torres, "A scalable re-ranking method for content-based image retrieval," *Information Sciences*, vol. 265, pp. 91–104, 2014.

[32] L. Zheng, S. Wang, and Q. Tian, "Coupled binary embedding for large-scale image retrieval," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3368–3380, 2014.

[33] E. G. Karakasis, A. Amanatiadis, A. Gasteratos, and S. A. Chatzichristofis, "Image moment invariants as local features for content based image retrieval using the Bag-of-Visual-Words model," *Pattern Recognition Letters*, vol. 55, pp. 22–27, 2015.

[34] M. Rahimi and M. E. Moghaddam, "A content-based image retrieval system based on color ton distribution descriptors," *Signal, Image and Video Processing*, pp. 691–704, 2015.

[35] A. Rashno, S. Sadri, and H. Sadeghiannejad, "An efficient content-based image retrieval with ant colony optimization feature selection schema based on wavelet and color features," in *Proceedings of the 2015 International Symposium on Artificial Intelligence and Signal Processing, AISP 2015*, pp. 59–64, Mashhad, Iran, March 2015.

[36] Z. Mehmood, S. M. Anwar, N. Ali, H. A. Habib, and M. Rashid, "A Novel image retrieval based on a combination of local and global histograms of visual words," *Mathematical Problems in Engineering*, vol. 2016, Article ID 8217250, 2016.

[37] X. Yuan, J. Z. Yu, Qin. Z., and T. Wan, "A SIFT-LBP image retrieval model based on bag of features," in *Proceedings of the IEEE International Conference on Image Processing*, IEEE, Brussels, Belgium, 2011.

[38] Z. Zhao, Q. Tian, H. Sun, X. Jin, and J. Guo, "Content Based Image Retrieval Scheme using Color, Texture and Shape Features," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9, no. 1, pp. 203–212, 2016.

[39] E. de Ves, X. Benavent, I. Coma, and G. Ayala, "A novel dynamic multi-model relevance feedback procedure for content-based image retrieval," *Neurocomputing*, vol. 208, pp. 99–107, 2016.

[40] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, and K. Ren, "A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2594–2608, 2016.

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

[42] X. Tian, L. Jiao, X. Liu, and X. Zhang, "Feature integration of EODH and Color-SIFT: application to image retrieval based on codebook," *Signal Processing: Image Communication*, vol. 29, no. 4, pp. 530–545, 2014.

[43] P. Poursistani, H. Nezamabadi-pour, R. Askari Moghadam, and M. Saeed, "Image indexing and retrieval in JPEG compressed domain based on vector quantization," *Mathematical and Computer Modelling*, vol. 57, no. 5-6, pp. 1005–1017, 2013.

[44] E. Yildizer, A. M. Balci, M. Hassan, and R. Alhajj, " Efficient contentbased image retrieval using multiple support vector machines ensemble," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2385–2396, 2012.

[45] S. Zeng, R. Huang, H. Wang, and Z. Kang, "Image retrieval using spatiograms of colors quantized by gaussian mixture models," *Neurocomputing*, vol. 171, pp. 673–684, 2016.

[46] D. Zhong and I. Defée, "DCT histogram optimization for image database retrieval," *Pattern Recognition Letters*, vol. 26, no. 14, pp. 2272–2281, 2005.

[47] A. ElAdel, R. Ejbali, M. Zaied, and C. B. Amar, "A hybrid approach for content-based image retrieval based on Fast Beta Wavelet network and fuzzy decision support system," *Machine Vision and Applications*, vol. 27, no. 6, pp. 781–799, 2016.

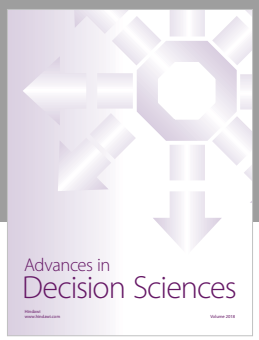
[48] A. Vedaldi and B. Fulkerson, "Vlfeat: an open and portable library of computer vision algorithms," in *Proceedings of the International Conference on Multimedia (MM '10)*, pp. 1469–1472, October 2010.

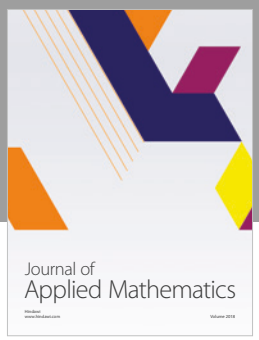
[49] A. Vedaldi and A. Zisserman, "Sparse kernel approximations for efficient classification and detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2320–2327, June 2012.

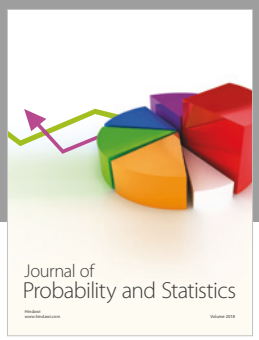
[50] J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947–963, 2001.

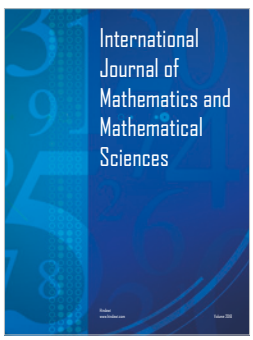
[51] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, 2003.

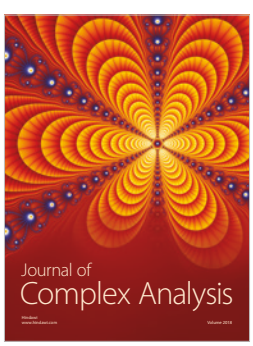
[52] G. Griffin, A. Holub, and P. Perona, *Caltech-256 Object Category Dataset*, 2007.

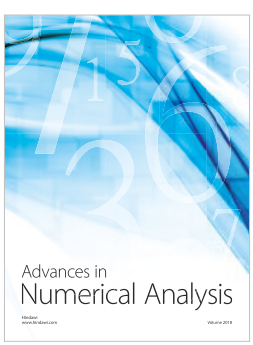


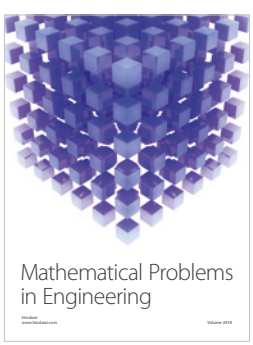


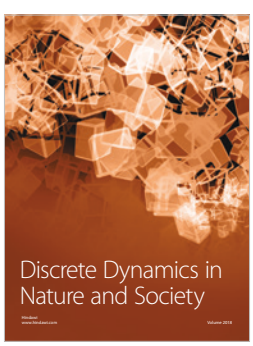


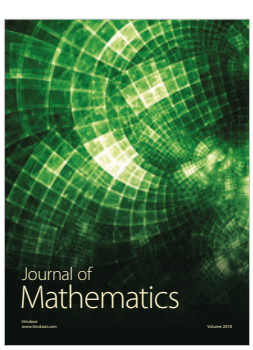


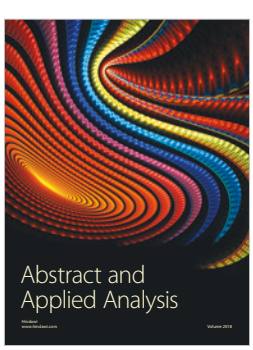


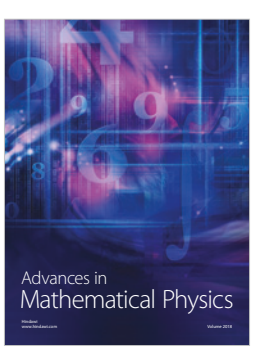












Hindawi

Submit your manuscripts at
www.hindawi.com