Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

Content-Based Image Retrieval Based on Late Fusion of Binary and Local Descriptors

Nouman Ali [1,2,*], Danish Ali Mazhar [2], Zeshan Iqbal [2], Rehan Ashraf [1], Jawad Ahmed [3] and Farrukh Zeeshan Khan [2]

[1] Department of Computer Science, National Textile University, Faisalabad, Pakistan; rehan@ntu.edu.pk

[2] Faculty of Telecommunication & Information Engineering, University of Engineering & Technology, Taxila, Pakistan; danishalimazhar@gmail.com (D.A.M.); zeshan.iqbal@gmail.com (Z.I.); farrukh.zeeshan@uettaxila.edu.pk (F.Z.K.)

[3] Department of Basic Sciences, University of Engineering & Technology, Taxila, Pakistan; jawad.ahmed@uettaxila.edu.pk

* Correspondence: noumanali@ntu.edu.pk

Abstract

One of the challenges in Content-Based Image Retrieval (CBIR) is to reduce the semantic gaps between low-level features and high-level semantic concepts. In CBIR, the images are represented in the feature space and the performance of CBIR depends on the type of selected feature representation. Late fusion also known as visual words integration is applied to enhance the performance of image retrieval. The recent advances in image retrieval diverted the focus of research towards the use of binary descriptors as they are reported computationally efficient. In this paper, we aim to investigate the late fusion of Fast Retina Keypoint (FREAK) and Scale Invariant Feature Transform (SIFT). The late fusion of binary and local descriptor is selected because among binary descriptors, FREAK has shown good results in classification-based problems while SIFT is robust to translation, scaling, rotation and small distortions. The late fusion of FREAK and SIFT integrates the performance of both feature descriptors for an effective image retrieval. Experimental results and comparisons show that the proposed late fusion enhances the performances of image retrieval.

Keywords: CBIR; Late Fusion; SVM; BOVW

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

I. Introduction

In image classification and retrieval-based problems, the extraction of a meaningful image descriptor is an open research problem [1]. Due to an increase in the number of image archives, it is necessary to design an effective system for image search [1], [2]. The conventional annotations-based approach for image retrieval relies on text and keywords-based image search [1], [3], [4]. Human manual efforts and difference in the visual perception make the conventional approach less effective [1], [2]. The modern image search systems retrieve the images on the basis of image visual contents and this is referred as CBIR [1]. Image retrieval has vast applications in many domains like image analysis, search of image over internet, medical image retrieval, remote sensing and video surveillance [1]. The recent research is carried out to enhance the performance of image retrieval by retrieving the images that are similar to the query image [1], [5], [6]. In recent few years, different types of image representations are proposed that are associated with different application domains [1], [2], [7], [8], [9], [10]. In this paper, we are interested to explore a novel image representation based on the late fusion of binary and local features. The proposed framework can describe the visual contents in a meaningful way and the main focus of this research is image retrieval.

In CBIR, the commonly used image representations are mainly based on global and local features [1], [2], [11]. The global feature representations are based on color, texture and shape [1]. Color is one of the fundamental image feature and it is not dependent on size, direction and angle [2]. Color features lack spatial distribution and perceptual meaning [1], [2]. Texture features are divided into two main classes and texture is extracted to capture the spatial attributes in the group of pixels [2], [5]. The main limitation of spatial texture techniques is their sensitivity to noise and distortion [2], [5]. The spectral texture techniques are dependant on shape and have shown good performance in-case of square regions [5]. The sensitivity to noise and computational complexity limits the use of texture features in many application domains [2], [5]. According to Zhang et al. [12], region-based and contour-based are the two main approaches that are commonly applied to extract shape features. The sensitivity to scaling, translation and rotation are the main limitations of shape-based image representations [2], [5].

The local feature descriptors like Scale Invariant Feature Transform (SIFT) [13], Speeded-Up Robust Features (SURF) [14] and Histogram of Oriented Gradients (HOG) [15] have shown

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

good performance in various applications of image retrieval [7], [16]. Local feature descriptors are robust to translations, scaling, rotation, and small distortions and are applied in various image matching applications [7]. According to the recent literature [7], the classification and retrieval performance of local features is better than global features.

The recent advances in computer vision turned the attention of research to the use of binary descriptors [7], [17]. Binary Robust Invariant Scalable Keypoints (BRISK) [18], Fast Retina Keypoint (FREAK) [19], Binary Robust Independent Elementary Features (BRIEF)[20] and Features from Accelerated Segment Test (FAST) [21] are the examples of binary descriptors. Local and binary descriptors are used to describe the image information in a high-dimensional feature space by using keypoints [7]. The late fusion (also known as visual word integration) based on proper features selection improves the effectiveness of image retrieval. [5], [22], [23]. The performance of FREAK is reported efficient for classification-based problems [7], [17] while SIFT is robust to translation, scaling, rotation, and small distortions [24].

Due to these facts, in this paper, we investigated the late fusion/visual word integration of FREAK and SIFT. Binary and local features are extracted from training images. A quanization algorithm such as k-means is applied to cluster two codebooks (visual vocabularies), that are based on descriptors of FREAK and SIFT, respectively. The codebooks of binary and local features are combined to represent an image in a dimension that is twice the size of constructed vocabulary. The main contributions of the proposed research is the CBIR based on the late fusion of FREAK (binary descriptor) and SIFT (local descriptor).

The remaining paper is organized as follow: Section II describes the related work. Section III provides an overview about the proposed research methodology. Section IV presents evaluation measures and results obtained on three image benchmarks while Section V concludes our work and points towards the future directions in the research.

II. Related Work

CBIR is used to search the images that are visually similar to the query image [1], [10], [25]. Wang et al. [26] proposed Spatial Weighting BOF (SWBOF) by extracting the spatial information from different sub-blocks of the images. Adjacent block distance, local entropy and variance are used for the extraction of spatial information. The spatial weighting is achieved by weighting the correspondence visual words with the help of texture information. Yuan et al.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

[23] proposed image retrieval by using integration SIFT and Local Binary Pattern (LBP). The combination of SIFT-LBP is selected to enhance the performance of image retrieval in-case of noisy background and ambiguous objects. Yu et al. [22] proposed two new image representations based on early features fusion for an effective image retrieval. The clusters are constructed by applying a weighted scheme to maintain a balance between two features. The two new feature integrations are image-based SIFT-LBP and patch-based HOG-LBP. The image-based integration of SIFT-LBP outperforms the state-of-the-art approaches [22]. Zhang et al. [27] proposed a rotation invariant image matching system by using combination of SIFT and LBP. The regions of LBP descriptors are calculated by using SIFT detector. Kabbai et al. [28] proposed a new approach to extract invariant features from the regions of interest. The uniform pattern is applied to the LBP and Center Symmetric Local Binary Pattern (CSLBP) for a robust image matching application.

According to [6], the spatial information can be extracted by dividing an image into Level 1 and Level 2 triangles. Dense SIFT is used for feature extraction and three different classifiers are applied to determine the best retrieval performance. Zeng et al. [29] proposed an image representation that is based on generalized histogram of quantized colors. Gaussian Mixture Models (GMMs) is applied for quantization and Expectation-Maximization (EM) algorithm is used for training. Bayesian Information Criterion (BIC) is applied for determination of quantized color bins. Images are retrieved on the basis of similarity between the respective spatiograms. Walia et al. [30] proposed a fusion framework for color-based image retrieval. The color and texture are extracted by applying Color Difference Histogram (CDH) and Angular Radial Transform (ART) and modification in CDH algorithm is proposed to make it more effective. Dubey et al. [31] proposed a rotation and scale invariant hybrid image descriptor for an efficient image retrieval. The color features are extracted by quantizing RGB color space while texture is extracted by structuring the patterns that are generated from locally structured elements. Color and textural are integrated to construct the inherently Rotation and Scale-invariant Hybrid image Descriptor (RSHD). Montazer et al. [32] proposed a new learning method for RBFNNs (Radial Basis Function Neural Networks). Particle Swarm Optimization (PSO) is applied to initialize the radial basis function units in more accurate way. The spatial information of the data and non-linearity of the function are approximated to determine the widths of RBFNNs. According to Wan et al. [33], the modern machine learning techniques based on Convolutional Neural Networks (CNN)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

can reduce the semantic gaps. CNN model pre-trained on large dataset can be used for feature extraction and the optimized feature extraction techniques based on deep learning outperforms conventional features. In our previous work [34], we proposed the visual words integration (late fusion) of two local features. Different Weighed Averages (WA) of local features are also calculated to sort out the second best performance for image retrieval. The research presented in this paper is different from our previous work [34] as in this paper, we replaced SURF with FREAK to evaluate the late fusion (visual words integration) of SIFT (local descriptor) with FREAK (binary descriptor). The experimental results demonstrates the dominant effect of FREAK when used with the late fusion SIFT descriptor.

III. Bag of Features Representation

Fig. 1 represents the block diagram of the proposed image representation that is based on late fusion of binary and local features. The proposed research is based on BoVW image representation model [35]. The detail about features extraction using FREAK, SIFT and late fusion of FREAK and SIFT is mentioned in the following sub-sections.

A. Fast Retina Keypoint (FREAK)

FREAK is a binary descriptor that is computed on the basis of brightness comparison tests around the keypoints and is inspired from the human visual system [19]. In the first step, a circular sampling grid is applied to generate retinal sampling pattern with higher density of points near the center . One-bit Difference of Gaussian (DoG) is applied to compute a binary descriptor. The feature are calculated by applying a saccadic search. In the last step, the sum of local gradients over selected pairs are used to compute the rotation of keypoints [19].

B. Scale Invariant Feature Transform (SIFT)

There are four main steps that are involved to compute SIFT descriptors [13]. The first step involves the computation of interest points, for this purpose the Difference of Gaussian (DoG) is used to produce several Gaussian blurred images. The neighborhood image are compared for the calculation of DoG. The second step involves the calculation of extrema to find out the stable keypoints and low contrast and pints along the edges are removed by apply Taylor series. Determinant and trace of Hessian metric are used to remove the outliers. In the third step, to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

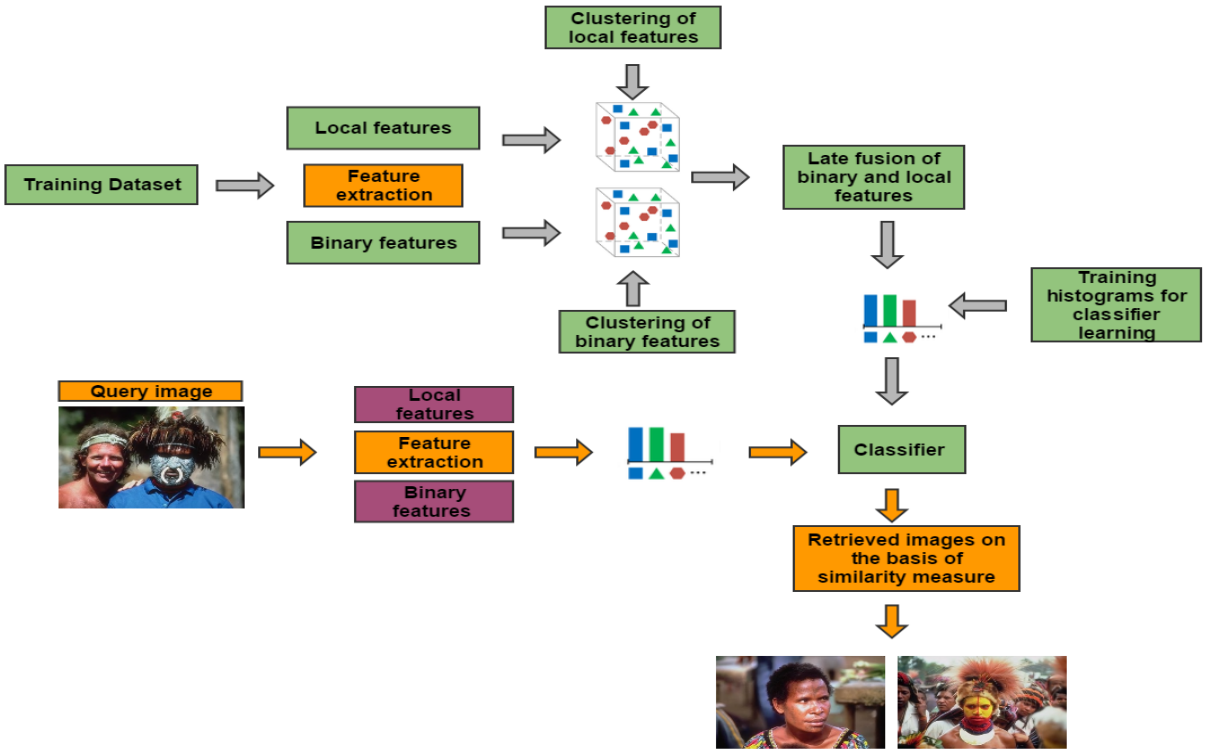

Fig. 1: Proposed framework based on late fusion of FREAK and SIFT.

achieve rotation invariance, the principal orientation is calculated for the keypoints. The last step involves the computation of SIFT descriptor. For each keypoint, a set of orientations histograms are created on 4x4 pixel neighborhoods, with 8 orientation bins [13].

C. Proposed Late Fusion Based on Binary and Local Descriptors

1) In BoVW model, a raw image I is represented as:

$$I = (a_{r,s}) \tag{1}$$

where *r,s* are the pixels of image.

2) Binary and local features (FREAK and SIFT) are extracted from a set of training images and an image *I* is represented as:

$$I = \{d_1, d_2, ..., d_T\} \tag{2}$$

Where d_1 to d_T are the image descriptors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

3) A quantization algorithm such as *k*-means is applied to construct the codebook (visual vocabulary) consisting of Z words, represented as *CB:*

$$CB = \{w_1, w_2, ..., w_Z\} \tag{3}$$

where *CB* is the codebook consisting of w_Z visual words, separate codebooks are constructed for FREAK and SIFT by extracting the respective features.

4) The features are extracted from the image and quantized in the feature space. Mapping of each visual word is done over the image areas and the nearest words are assigned to the quantized descriptors according to the following equation:

$$w(d_m) = \underset{w\ \varepsilon\ CB}{\text{argmin}}\ Dist(w, d_m) \tag{4}$$

where $w(d_m)$ is representing the visual word assigned to the m^{th} descriptor d_m while $Dist(w,d_m)$ is the distance between the descriptor d_m and visual word w. Each image is represented as a collection of patches and each patch is represented by a visual word (visual words of FREAK are mapped on the images by using the codebook of FREAK while visual words of SIFT are mapped on the images by the using the codebook of SIFT).

5) The histograms consisting of the visual words of FREAK and SIFT are combined and an image is represented in a dimension that is twice the size of constructed codebook. Consider Z as the number of visual words of the codebook. Let F_i be the set of the descriptors that are mapped to the visual word w_i then the i^{th} bin of the histogram of visual words b_i, is the cardinality of the set D_i.

$$b_i = Card(F_i) \quad and \quad F_i = \{d_m, m \in (1, ..., Z) \mid w(d_m) = w_i\} \tag{5}$$

IV. Experimental Parameters and Results

The proposed late fusion is evaluated by using three image dataset [36], [37], [38]. We used the image classification parameters as mentioned in [34]. 70% of the images from each class are selected for training and 30% for testing. Keeping in view, the unsupervised nature of clustering using *k*-means, each experiment is repeated 10 times. During every run, images are

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

randomly selected for training and testing. The set of selected images for training are used for the construction of codebook and image retrieval performance is reported by using the images from test dataset. The images are retrieved by calculating the closeness among the classifier score values within the same class.

The size of the codebook affects image retrieval performance [39], [40]. An increase in the the size of the codebook increases the retrieval precision. The larger size codebook decreases the retrieval precision due to over-fitting [39], [40]. We constructed different sizes of codebook from randomly set of selected training images to sort out the best retrieval performance. The codebook is constructed by a random selection of 25 %, 50 % and 75 % of features (both for FREAK and SIFT) per image (from the training dataset). Additional experiments are performed to evaluate the performance of FREAK and SIFT separately with the same experimental parameters to represent the dominant affect of late fusion on image retrieval performance.

A. *Evaluation Measures*

We selected precision and recall to determine the performance of our proposed late fusion [2]. Precision determines the number of correctly retrieved images.

$$Precision = \frac{K_r}{Xr} \tag{6}$$

where K_r represents the number of relevant images similar to the query and X_r indicates the number of images retrieved by the system in response to the query.

$$Recall = \frac{K_r}{Xc} \tag{7}$$

where X_c is total number of images of that class in the database.

B. *Performance Evaluation using Corel-1000*

There are 10 classes in Corel-1000 image dataset and each class contains 100 images. Fig. 2 represents a sample of randomly selected images from each class of the Corel-1000. We selected Corel-1000 for the evaluation of proposed framework as it has been recently used to evaluate the performance CBIR research [22], [26], [34], [32], [41]. We varied the codebook size to sort out the best retrieval performance of proposed work. Table I presents the numerical values of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

Fig. 2: Randomly selected images from each class of Corel-1000 image dataset [36].

Mean Average Precision (MAP) obtained from the proposed framework for top 20 retrievals. Fig. 3 presents a comparison of MAP as a function of codebook size using different evaluation parameters.

TABLE I: MAP as a function of codebook size (late fusion of FREAK and SIFT)

Codebook size and features % used	50	100	200	300	400	600	800
25 %	68.85	70.45	74.36	74.52	74.68	74.58	73.99
50 %	69.31	70.56	73.86	74.91	74.16	74.96	74.01
75 %	70.45	71.13	73.99	73.69	74.55	74.87	74.25
Mean	69.54	70.71	74.07	74.37	**74.46**	**74.80**	74.08

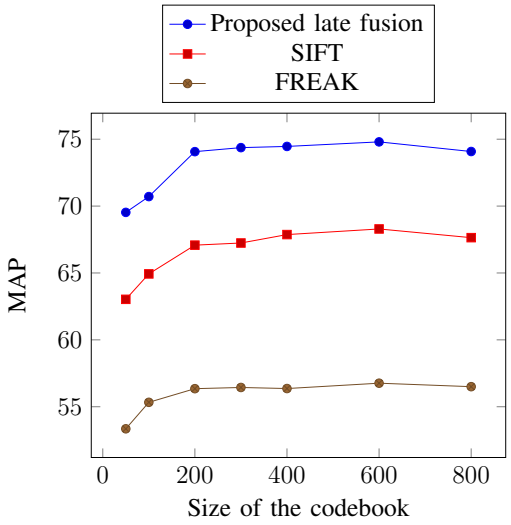

Fig. 3: MAP as a function of codebook size (Corel-100).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

The results and comparisons presented in Table I and Fig. 3 indicate that the best MAP is obtained from the proposed last fusion on a codebook of size 600 with a value of 74.80 %. The best MAP obtained by using SIFT and FREAK on a codebook with a size of 600 is 68.29% and 56.76%, respectively. The MAP on codebook of each size by using the proposed late fusion is higher than that of SIFT and FREAK. To present a sustainable performance, the best MAP obtained from the proposed late fusion is compared with existing research [22], [26], [34], [32], [41]. Table II and Table III presents the comparison of precision and recall for top 20 retrievals (as the research selected for comparison is also reported for top 20 retrievals).

TABLE II: Comparison of retrieval precision.

Name of class and method	Proposed frame-work	WA SIFT - SURF [34]	RBF-NN [32]	Color SIFT [41]	Spatial SIFT [26]	SIFT-LBP [22]
Africa	63.64	**69.75**	58.73	**74.60**	64	57
Beach	**60.99**	54.25	48.94	37.80	54	**58**
Buildings	**68.21**	**63.95**	53.74	53.90	53	43
Buses	92.75	89.65	**95.81**	**96.70**	94	93
Dinosaurs	**100.00**	98.7	98.36	**99**	98	98
Elephants	**72.64**	48.8	64.14	65.90	**78**	58
Flowers	**91.54**	**92.3**	85.64	91.20	71	83
Horses	80.06	**89.45**	80.31	**86.90**	93	68
Mountains	**59.67**	47.3	54.27	**58.50**	42	46
Food	58.56	**70.9**	**63.14**	62.20	50	53
Mean	**74.80**	**70.58**	70.31	72.67	69.70	65.70

The best values of precision and recall are mentioned as bold in Table II and Table III. The MAP of proposed late fusion outperforms state-of-the-art research [22], [26], [34], [32], [41]. Fig. 4 and Fig. 5 represents the image retrieval results obtained on the basis of similarity measures for the semantic class ”Flowers” and ”Buses”, respectively. The classifier output label determines the class of the image while the images are retrieved on the basis of closeness among classifier score values.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

TABLE III: Comparison of recall.

Name of class and method	Proposed frame-work	WA SIFT - SURF [34]	RBF-NN [32]	Color SIFT [41]	Spatial SIFT [26]	SIFT-LBP [22]
Africa	12.73	**13.95**	11.75	**14.92**	12.80	11.40
Beach	**12.20**	10.85	9.79	7.56	10.80	**11.60**
Buildings	**13.64**	**12.79**	10.75	10.78	10.60	8.60
Buses	18.55	17.93	**19.16**	**19.34**	18.80	18.60
Dinosaurs	**20.00**	19.74	19.67	**19.80**	19.60	19.60
Elephants	**14.53**	9.76	12.83	13.18	**15.60**	11.60
Flowers	**18.31**	**18.46**	17.13	18.24	14.20	16.60
Horses	16.01	**17.89**	16.06	**17.38**	18.60	13.60
Mountains	**11.93**	9.46	10.85	**11.70**	8.40	9.20
Food	11.71	**14.18**	**12.63**	12.44	10.00	10.60
Mean	**14.96**	**14.12**	14.06	14.53	13.94	13.14

Fig. 4: Retrieval result on the basis of similarity measures for the semantic class ”Flowers”

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

Fig. 5: Retrieval result on the basis of similarity measures for the semantic class ”Buses”

Fig. 6: Randomly selected images from each class of Corel-1500 image dataset [37].

C. Performance Evaluation Using Corel-1500

There are 15 classes in the Corel-1500 [37] image benchmark and each class contains 100 images [37]. Fig. 6 represents a sample of randomly selected images from each class of the Corel-1500. We evaluated the proposed framework on Corel-1500 and compared the results with state-of-the-art CBIR methods [29], [34]. Fig. 3 presents a comparison of MAP as a function of codebook size using different evaluation parameters.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

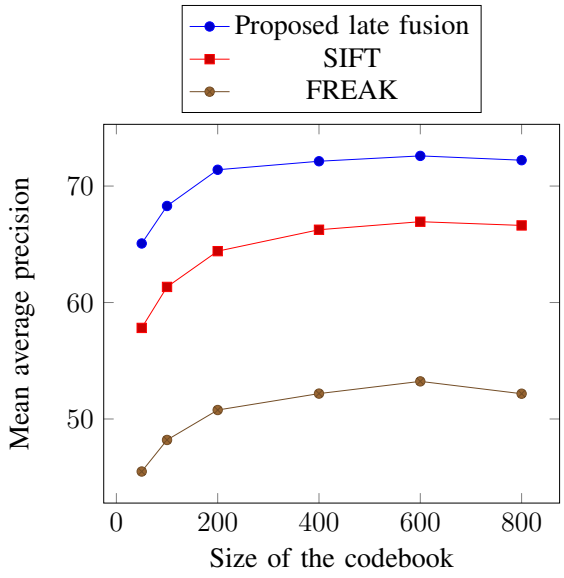

Fig. 7: MAP as a function of codebook size (Corel-1500).

The comparison results results presented in Fig. 3 indicate that the best MAP is obtained from the proposed last fusion on a codebook of size 600 with a value of 72.59 %. The best MAP obtained by SIFT and FREAK is 66.94 % and 53.23 %, respectively. Table IV presents a comparison of precision and recall values.

TABLE IV: Comparison of precision and recall.

Image retrieval performance	Proposed late fusion	WA SIFT-SURF [34]	SQ + Spatiogram [29]
Precision	**72.60**	68.05	63.95
Recall	**14.52**	13.61	12.79

The proposed framework based on late fusion of FREAK and SIFT provides a better retrieval performance with higher values of precision and recall than the existing research [29], [34].

D. Performance Evaluation using Oliva and Torralba (OT-Scene)

There are 08 classes in OT-Scene image dataset that contains a total of 2688 images. Fig. 8 represents a sample of randomly selected images from each class of the OT-Scene image dataset.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

Fig. 8: Randomly selected images from each class of OT-Scene image dataset [38].

We evaluated the proposed framework using OT-Scene benchmark and compared the results with state-of-the-art CBIR methods [30], [42]. Fig. 9 presents a comparison of mean precision as a function of codebook size using different evaluation parameters. Table V represents the comparison of MAP values.

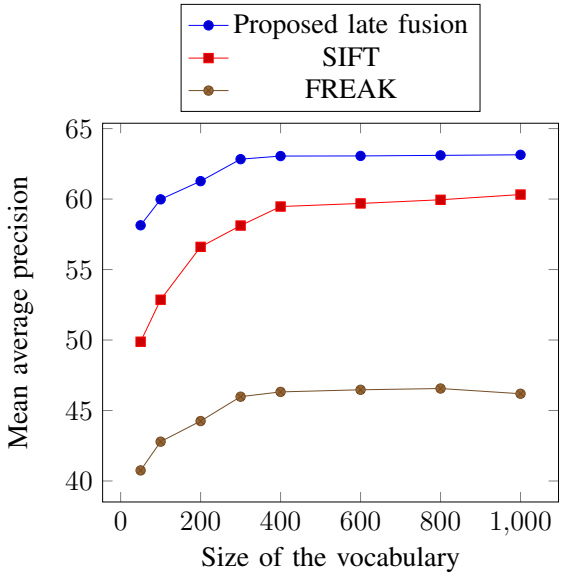

Fig. 9: MAP as a function of codebook size (OT-Scene).

TABLE V: Mean precision comparison.

Image retrieval performance	Proposed late fusion	Morphology based IR [42]	Min Max Fusion [30]
MAP	**63.14**	60.7	51.04

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

The results and comparisons presented in Fig. 9 and Table V indicate that the proposed late fusion of FREAK and SIFT provides a better retrieval performance than existing research [30], [42].

V. Conclusion and Future Directions

The research scope of this paper is based on the late fusion/visual word integration of binary and local descriptor for an effective image retrieval. Keeping in view, the good recognition ability and efficient computation time, we selected FREAK as the binary descriptor. SIFT is selected as the local feature as it is robust to change in translation, scaling, rotation, and small distortions. The image is represented in the form of late fusion of FREAK and SIFT. The proposed research is based on the BoVW framework and classification is performed by using SVM. Testing is performed by varying the size of codebook, to sort out the best image retrieval precision. The results obtained from the proposed research are compared with state-of-the-art CBIR research. The late fusion of FREAK and SIFT outperforms several CBIR methods. In future, we will evaluate our proposed work on real world retrieval problem by using a pre-trained CNN model.

Acknowledgments

The authors would like to thank Professor James Z. Wang, Pennsylvania State University, USA for sharing Corel image benchmark that is used for the evaluation of the proposed research.

References

[1] T. Dharani and I. L. Aroquiaraj, "A survey on content based image retrieval," in *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on*. IEEE, 2013, pp. 485–490.

[2] A. Alzubi, A. Amira, and N. Ramzan, "Semantic content-based image retrieval: A comprehensive study," *Journal of Visual Communication and Image Representation*, vol. 32, pp. 20–54, 2015.

[3] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data," *Journal of digital imaging*, vol. 26, no. 6, pp. 1025–1039, 2013.

[4] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "A simultaneous feature adaptation and feature selection method for content-based image retrieval systems," *Knowledge-Based Systems*, vol. 39, pp. 85–94, 2013.

[5] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346–362, 2012.

[6] 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*, vol. 54, pp. 539–550, 2016.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

[7] D. Mukherjee, Q. J. Wu, and G. Wang, "A comparative experimental study of image feature detectors and descriptors," *Machine Vision and Applications*, vol. 26, no. 4, pp. 443–466, 2015.

[8] R. Ashraf, T. Mahmood, A. Irtaza, and K. Bajwa, "A novel approach for the gender classification through trained neural networks," *J. Basic Appl. Sci. Res*, vol. 4, pp. 136–144, 2014.

[9] R. Ashraf, K. Bashir, A. Irtaza, and M. T. Mahmood, "Content based image retrieval using embedded neural networks with bandletized regions," *Entropy*, vol. 17, no. 6, pp. 3552–3580, 2015.

[10] R. Ashraf, K. Bashir, and T. Mahmood, "Content-based image retrieval by exploring bandletized regions through support vector machines," *Journal of Information Science and Engineering*, vol. 32, no. 2, pp. 245–269, 2016.

[11] 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, 2016.

[12] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern recognition*, vol. 37, no. 1, pp. 1–19, 2004.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[14] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer vision–ECCV 2006*. Springer, 2006, pp. 404–417.

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[16] T. Krajnik, P. Cristóforis, M. Nitsche, K. Kusumam, and T. Duckett, "Image features and seasons revisited," in *Mobile Robots (ECMR), 2015 European Conference on*. IEEE, 2015, pp. 1–7.

[17] J. Figat, T. Kornuta, and W. Kasprzak, "Performance evaluation of binary descriptors of local features," in *Computer Vision and Graphics*. Springer, 2014, pp. 187–194.

[18] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2548–2555.

[19] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. Ieee, 2012, pp. 510–517.

[20] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," *Computer Vision–ECCV 2010*, pp. 778–792, 2010.

[21] D. G. Viswanathan, "Features from accelerated segment test (fast)," 2009.

[22] 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.

[23] X. Yuan, J. Yu, Z. Qin, and T. Wan, "A sift-lbp image retrieval model based on bag of features," in *IEEE International Conference on Image Processing*, 2011.

[24] C. Valgren and A. J. Lilienthal, "Sift, surf & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems*, vol. 58, no. 2, pp. 149–156, 2010.

[25] Z. Mehmood, S. M. Anwar, M. Altaf, and N. Ali, "A novel image retrieval based on rectangular spatial histograms of visual words," *Kuwait Journal of Science*, 2017.

[26] C. Wang, B. Zhang, Z. Qin, and J. Xiong, "Spatial weighting for bag-of-features based image retrieval," in *Integrated Uncertainty in Knowledge Modelling and Decision Making*. Springer, 2013, pp. 91–100.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2017 doi:10.20944/preprints201703.0156.v2

[27] Y.-b. Zheng, X.-s. Huang, and S.-j. Feng, "An image matching algorithm based on combination of sift and the rotation invariant lbp," *Journal of computer-aided design & computer graphics*, vol. 22, no. 2, pp. 286–292, 2010.

[28] L. Kabbai, A. Azaza, M. Abdellaoui, and A. Douik, "Image matching based on lbp and sift descriptor," in *Systems, Signals & Devices (SSD), 2015 12th International Multi-Conference on*. IEEE, 2015, pp. 1–6.

[29] 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.

[30] E. Walia and A. Pal, "Fusion framework for effective color image retrieval," *Journal of Visual Communication and Image Representation*, vol. 25, no. 6, pp. 1335–1348, 2014.

[31] S. R. Dubey, S. K. Singh, and R. K. Singh, "Rotation and scale invariant hybrid image descriptor and retrieval," *Computers & Electrical Engineering*, vol. 46, pp. 288–302, 2015.

[32] G. A. Montazer and D. Giveki, "An improved radial basis function neural network for object image retrieval," *Neurocomputing*, vol. 168, pp. 221–233, 2015.

[33] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 157–166.

[34] N. Ali, K. B. Bajwa, R. Sablatnig, S. A. Chatzichristofis, Z. Iqbal, M. Rashid, and H. A. Habib, "A novel image retrieval based on visual words integration of sift and surf," *PloS one*, vol. 11, no. 6, p. e0157428, 2016.

[35] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 1470–1477.

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

[37] J. Li and J. Z. Wang, "Real-time computerized annotation of pictures," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 6, pp. 985–1002, 2008.

[38] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001.

[39] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 490–503.

[40] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learning in computer vision, ECCV*, vol. 1, no. 1-22. Prague, 2004, pp. 1–2.

[41] 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.

[42] R. Das, S. Thepade, and S. Ghosh, "Multi technique amalgamation for enhanced information identification with content based image data," *SpringerPlus*, vol. 4, no. 1, pp. 1–26, 2015.

© 2017 by the authors. Licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).