Content-Based Image Retrieval using Local Features Descriptors and Bag-of-Visual Words

Mohammed Alkhawlani

Ibb University
Ibb, Yemen

Mohammed Elmogy

Dept. of Information Technology,
Faculty of Computers and
Information, Mansoura University
Mansoura, Egypt

Hazem Elbakry

Dept. of Information Systems,
Faculty of Computers and
Information, Mansoura University
Mansoura, Egypt

Abstract—**Image retrieval is still an active research topic in the computer vision field. There are existing several techniques to retrieve visual data from large databases. Bag-of-Visual Word (BoVW) is a visual feature descriptor that can be used successfully in Content-based Image Retrieval (CBIR) applications. In this paper, we present an image retrieval system that uses local feature descriptors and BoVW model to retrieve efficiently and accurately similar images from standard databases. The proposed system uses SIFT and SURF techniques as local descriptors to produce image signatures that are invariant to rotation and scale. As well as, it uses K-Means as a clustering algorithm to build visual vocabulary for the features descriptors that obtained of local descriptors techniques. To efficiently retrieve much more images relevant to the query, SVM algorithm is used. The performance of the proposed system is evaluated by calculating both precision and recall. The experimental results reveal that this system performs well on two different standard datasets.**

Keywords—Content-based Image Retrieval (CBIR); Scale Invariant Feature Transform (SIFT); Speeded Up Robust Features (SURF); K-Means Algorithm; Support Vector Machine (SVM); Bag-of-Visual Word (BoVW)

I. INTRODUCTION

Image retrieval is the field of the study that concerned with looking, browsing, and recovering digital images from an extensive database. CBIR is viewed as a dynamic and quick advancing research area in image retrieval field. It is a technique for retrieving images from a collection by similarity. The retrieval based on the features extracted automatically from the images themselves. Many of CBIR systems, which is based on features descriptors, are built and developed.

A feature is defined as capturing a certain visual property of an image. A descriptor encodes an image in a way that allows it to be compared and matched to other images. In general, image features descriptors can be either global or local. The global feature descriptors describe the visual content of the entire image, whereas local feature describes describe a patch within an image (i.e. a small group of pixels) of the image content. The superiority of the global descriptor extraction is the increased speed for both feature extraction and computing similarity. However, global features still too rigid to represent an image. Particularly, they can be oversensitive to location and consequently fail to identify important visual characteristics [1, 2].

Local feature approaches provide better retrieval effectiveness and great discriminative power in solving vision problems than global features [3]. However, the number of local features that are extracted for each image may be immense, especially in the large image dataset. Wherefore, BoVW [4, 5] is proposed as an approach to solving this problem by quantizing descriptors into "visual words."

Depending on the previous facts, the present study proposed a system for image retrieval based on local features using BoVW model. The system tries to bring more accuracy with the option to use the two main local descriptors (SIFT [6], SURF [7]).

The rest of this paper is organized as follows. Section 2 gives an overview of the BoVW model, K-Means, and SVM. Section 3 discusses two of the most commonly used local feature descriptors. Section 4 reviews some of the related work using BoVW model in image retrieval. In Section 5, the proposed architecture of our image retrieval system, which is based on local feature descriptor, is introduced. Our experimental results are manipulated in Section 6. Finally, Section 7 contains the conclusion and our future work.

II. BAG-OF-VISUAL WORD MODEL

The BoVW model is one of the most widely used ways that represents images as a collection of local features. For this reason, some researchers tend to name it as a bag of features. These local features are typically grouped of local descriptors. The total number of local descriptors that is extracted for each image may be colossal. In addition, searching nearest neighbors for each local descriptor in the image query consumes a long time. Therefore, BoVW was proposed as an approach to tackling this issue by quantizing descriptors into "visual words," which decreases the descriptors' sum drastically. Thus, BoVW makes the descriptor more robust to change. This model is very close to the traditional description of texts in information retrieval, but it is considered for images retrieval [5, 6]. BoVW is the de facto standard of image features for retrieval and recognition [7]. It consists of three main stages like the following in the sequent subsections:

A. Keypoint Detection

The first step of the BoVW model is to detect local interest points. For feature extraction of interest points, they are computed at predefined locations and scales [8]. Feature extraction is a separate process from feature representation in

BoVW approaches [9]. There are many keypoint detectors that were used in research, such as Harris-Laplace, Difference of Gaussian (DoG), Hessian Laplace, and Maximally Stable Extremal Regions (MSER) [10, 11].

B. Features Descriptors

The keypoints are described as multidimensional numerical vectors, according to their content [6]. In other words, features descriptors are used to determine how to represent the neighborhood of pixels near a localized keypoint [9]. The most efficient feature descriptors in the BoVW model are SIFT and SURF.

C. Building Vocabulary

In the previous stage, the total extracted feature descriptors are large. To solve this problem, the feature descriptors are clustered by applying the clustering algorithm, such as K-Means technique [12] to generate a visual vocabulary. Each cluster is treated as a distinct visual word in the vocabulary, which is represented by their respective cluster centers. The size of the vocabulary is determined using the clustering algorithm. In addition, it depends on the size and the types of the dataset [7].

The BOW model can be formulated as follows. First, BoVW is usually defining the training dataset as S including images represented by $S = s_1, s_2, \ldots, s_n$, where s is the extracted visual features. After that, used clustering algorithm like K-Means, which is based on a fixed number to visual words W represented by $W = w_1, w_2, \ldots, w_v$, where v is the cluster number. Then, the data is summarized in a V×N occurrence table of counts $N_{ij} = n(w_i, s_j)$, where $n(w_i, s_j)$ denotes how often the word w_i is occurred in an image s_j [6].

On the other hand, K-Means is one of the most unsupervised learning algorithms that take care of the well-known clustering issue. It defines the size of K clusters based on the features extracted from the images themselves [13]. It is used to calculate the nearest neighbors of the points and the cluster center. It is usually utilizing the method of computation by approximating the nearest neighbor method. This method can be scaled to similarly large vocabulary sizes by the use of approximate nearest neighbor methods [12].

SVM is supervised machine learning technique [14]. It shows the image database as two sets of vectors in a high or infinite-dimensional space. It relies on a fundamental principle, which is called a maximum margin classifier. A maximum margin classifier is a hyperplane, which separates two 'clouds' of points at equal distance. The margin between the hyperplane and the clouds is maximal. SVM built a hyperplane or set of hyperplanes that increases the margin among the images that are relevant and not relevant to the query [15]. The goal of SVM classification technique is to find an ideal hyperplane to separate the irrelevant and relevant vectors using maximizing the size of the margin between both classes [16].

An image classification is a machine learning technique. It is a step used to accelerate image retrieval in big-scale databases and is used to increase retrieval precision. Similarly, in the absence of labeled data, unsupervised clustering is found to be helpful to increase the retrieval velocity and to improve retrieval precision. Image clustering based on a similarity measure, while the image classification has been performed using different techniques that does not require the use of similarity measures [15, 17].

III. LOCAL FEATURE DESCRIPTORS

In computer vision, local feature technique contains two parts [18]: feature detector and feature descriptor. Feature detector determines regions of an image that have unique content, like corners. Feature detection is used to find interest points (keypoints) in the image that remain locally invariant. Therefore, it can detect them even in the presence of scale change or rotation. Whereas, feature descriptor involves computing a local descriptor, which is usually done on regions centered on detected interest points. Local descriptors depend on image processing to transform a local pixel neighborhood into a compact vector representation [19].

On the other hand, the local descriptors are broadly used in many of computer vision research, such as robust matching, image retrieval, and object detection and classification. In addition, using local descriptors enables computer vision algorithms to deal strongly with rotation, occlusion, and scale changes.

Local feature algorithms depend on the idea of determining some interest points in the image and implementing a local analysis on them, rather than looking at the image as a whole. There are numerous algorithms for describing local image regions, such as SIFT and SURF. The SIFT and SURF descriptors depend on local gradient computations. The following subsections will discuss the SIFT and SURF algorithms briefly.

A. Scale Invariant Feature Transform (SIFT)

Lowe [3] developed SIFT as a continuation of his previous work on invariant feature detection. It has four computational phases: (a) extrema detection, (b) keypoint localization, (c) orientation assignment, and (d) keypoint description.

The first phase examines the image under different octaves and scales to isolate points of the image that are different from their surroundings. These points, which are called extrema, are potential candidates for image features. In keypoint localization phase, it selects some of extrema points to be keypoints. Candidate keypoints are refined by reject extrema points that are caused by edges and by low contrast points. In the orientation assignment phase, it represents every keypoint and neighbors as a set of vectors using the magnitude and the direction. In the last phase, it takes a collection of vectors in the neighborhood of every keypoint and combines this information with a set of eight vectors called the descriptor. The neighborhood is divided into 4×4 regions, in each region the vectors are histogrammed in eight bins. SIFT provides a 128 element of the keypoint descriptor.

B. Speeded Up Robust Features (SURF)

Bay *et al.* [4] introduced the SURF algorithm as a scale- and rotation-invariant interest point detector and the descriptor. SURF algorithm is a mixing of crudely localized information and the distribution of related gradient. SURF algorithm is similar to SIFT algorithm, but it is much more simplified and faster in computation and matching.

SURF algorithm depends on the Hessian Matrix to detect keypoints. It uses a distribution of Haar wavelet responses at the keypoint's neighborhood. The final descriptor is obtained by concatenating the feature vectors of all the sub-regions and represented with 64 elements.

The SIFT and SURF algorithms are nowadays the most widely used feature-based techniques in the computer vision community. These algorithms have proven their efficiency and robustness in the invariant feature localization (invariant to image rotation, scaling, and changes in illumination) [4, 20].

IV. RELATED WORK

The subject of image retrieval is discussed intensively in the literature. The success of using BoVW model had also contributed to increasing the number of researchers and studies. For example, Cakir *et al.* [21] studied CBIR using BoVW model. They discussed how BoVW considers an image as a document, going through using K-Means as vector quantization uses.

Zhang *et al.* [22] proposed a bag of images for CBIR schemes. They supposed that the image collection composed of image bags rather than independent individual images. They contain some relevant images that have same perceptual meaning. The image bags were built before image retrieval. In addition, a user's query is an image bag, named query image bag. In this condition, all image bags in the image collection are sorted according to their similarities to the query image bag. It hypothetically represented that the new idea can enhance the image retrieval process. However, this work needs to develop more efficient ways to measure the dissimilarity between two image bags.

Ponitz *et al.* [23] attempted to solve the problem of detecting images limitations in huge scale image databases. They decide to enhance the methodology of BoVW by improving the distance measure between image signatures to avoid the occurrence of vague features. They utilized SIFT algorithm for local visual features acquisition. Only 60% of all images were randomly chosen, and their features utilized for clustering. These features were then quantized. 100 random images are selected as input images. The images were changed with mounting distortion to test the robustness of the application. It needs more discrimination force of the actual image description.

Liu [8] reviewed BoVW model in image retrieval system. He provided details about BoVW model and explained different building strategies based on this model. First, he presented several procedures that can be taken in BoVW model. Then, he explained some popular keypoint detectors and descriptors. Finally, he looked at strategies and libraries to generating vocabulary and do the search.

Alfanindya *et al.* [24] presented a method for CBIR by using SURF with BoVW. First, they used SURF to computed interest points and descriptors. Then, they created a visual dictionary for each group in the COREL database. They concluded from their experiments that their method outperforms some other methods in terms of accuracy. The major challenge in their work was that the proposed method is highly supervised. It means that they n need to determine the number of groups before they perform classification.

The primary aim of this paper is to design a system for image retrieval based on local feature descriptors using BoVW model. Most of the previous image retrieval using BoVW systems used only one local descriptor. Whereas, our proposed system uses both SIFT and SURF descriptors. It provides a comparison of the actual performance of those local descriptors with BoVW in image retrieval field.

V. SYSTEM ARCHITECTURE

We propose a system for image retrieval based on extracting local features using BoVW model. The system uses SIFT or SURF techniques to extract keypoints and compute the descriptor for those keypoints. K-Means algorithm is used to obtain the visual vocabulary. As shown in Figure 1, the proposed system consists of two stages: a training stage and a testing stage. During the training stage, the proposed system is given below:

1) *For each image in the dataset:*
- Convert image to a grayscale.

- Resizing the image to (300,300 pixels) to get uniformed results.

- Image features are extracted and associated these characteristics to local descriptors.

- Cluster the set of these local descriptors for the amount of bags using a K-Means algorithm to construct a vocabulary of K clusters.

2) *For each feature descriptor in the image:*
- Find the nearest visual word from the vocabulary for each feature vector with L2 distance based matching.

- Compute the Bag-of-words image descriptor as is a normalized histogram of vocabulary words encountered in the image.

- Save the Bag-of-words descriptors for all image.

At the test stage, the proposed system is given below, for each input image:

- The input image is pre-processed for keypoints extraction.

- Local descriptors are computed from the pre-processed input image.

- Compute the Bag-of-Words vector with the algorithm defined above.

- In the matching step, grab the best results via SVM Classification.
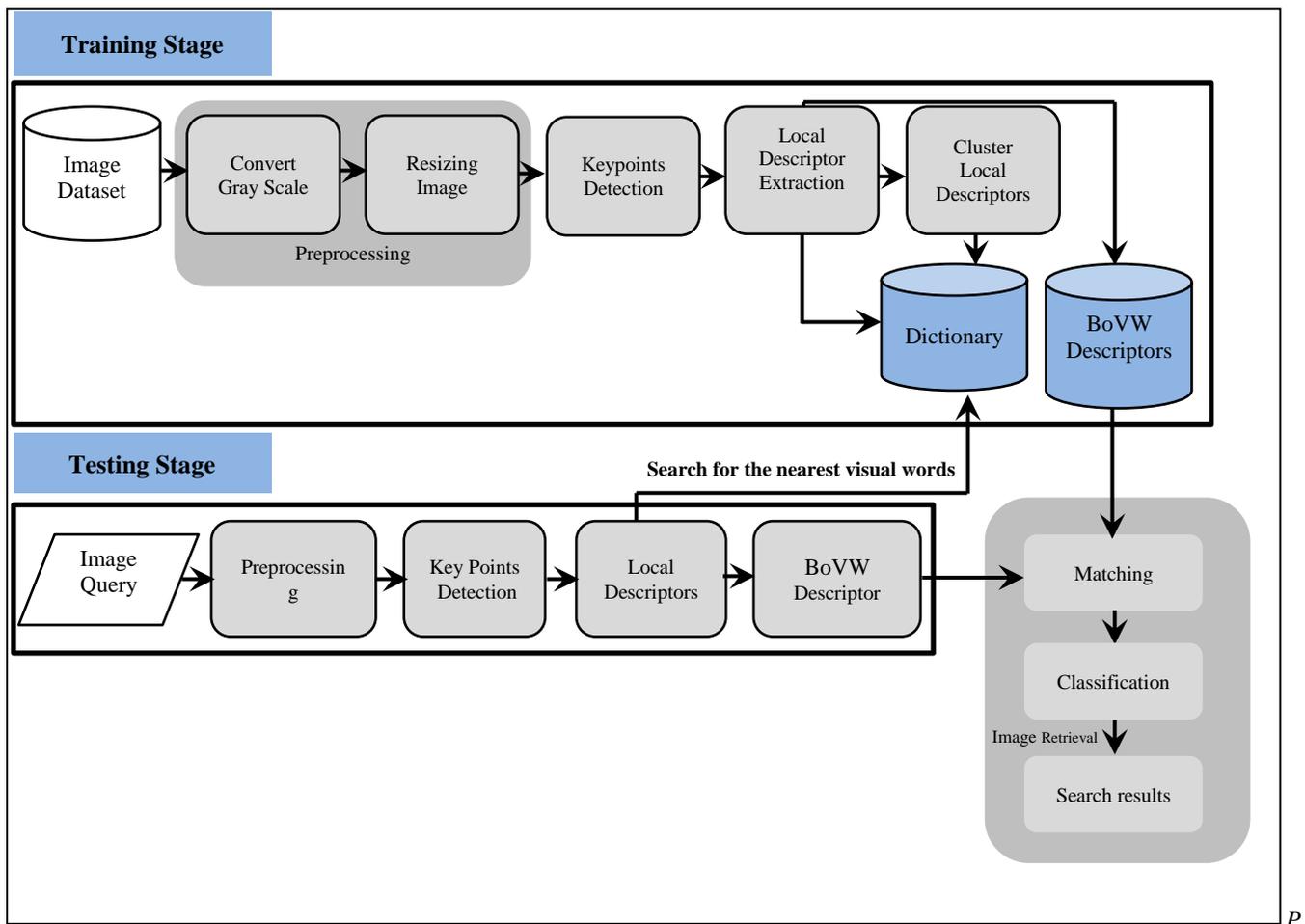
Fig. 1.   The architecture of the proposed system

A. Reprocessing

The preprocessing step consists of converting the image to grayscale and resizing process. Due to the local descriptors algorithms that deal only with density information, the images are converted to the grayscale. After that, the images are resized to 300x300 pixels to normalize the results.

B. Keypoint Detection and Description

The most important step in the proposed system is to extract the local descriptors from the processed image. There are many keypoint description techniques, such as Harris, SIFT, and SURF. In this paper, SIFT and SURF description were chosen in order to test the performance of the proposed system. Once keypoints are extracted from the image, the system computes the local description of each keypoint, as shown in Figure 2.

C. BoVW Descriptor

In this step, the BoVW model is used to create the vocabulary. First, we compute the centroid of the vocabulary that is closest to the feature vector using Brute Force matcher method. Then, we calculate the difference between the centroid and the feature vector. Finally, we compute the bag-of-words image descriptor as a normalized histogram of vocabulary words.
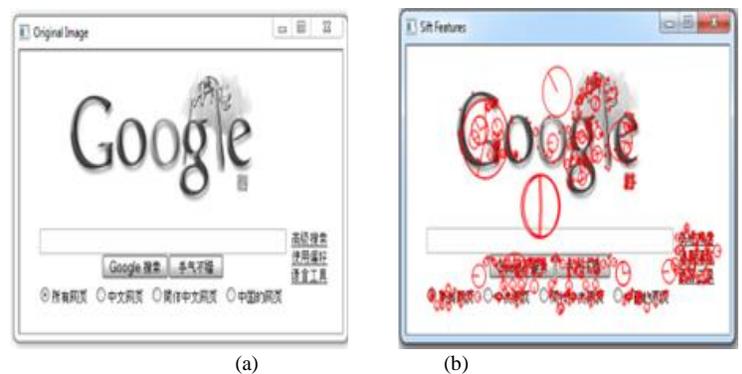

(a)   (b)

Fig. 2.   The local feature extration for one of the used images, (a) The gayscale image, (b) The extracted local features

D. Matching and Classification

At this stage, the descriptor query is used to match the BoVW descriptors in the database. The nearest neighbor approach was used to retrieve similar images.

Finally, SVM classification was used to grab the best results, which has the most similarity with the image query.

VI. THE PERFORMANCE EVALUATION AND RESULTS

A. Dataset

The system was evaluated by using two different standard datasets: the Flickr Logos 27 dataset [25] and Amsterdam Library of Object Images (ALOI) dataset [26]. The Flickr Logos 27 dataset is an annotated logo dataset downloaded from Flickr, and it consists of three image collections/sets. The training set contains 810 annotated images, corresponding to 27 logo classes/brands (30 images for each class). Figure 3 shows some image samples of the training set. The query set consists of 270 images. There are five images for each of the 27 annotated classes, summing up to 135 images that contain logos. Some image samples from the queries set are presented in Figure 4.

Fig. 3. Some sample images from the Flickr Logos dataset for the training

Fig. 4. Some sample images from the Flickr Logos dataset for the queries

ALOI is a color image collection of one-thousand small objects, which is recorded for scientific purposes under various imaging circumstances (viewing angle, illumination angle, and illumination color). Over a hundred images of each object were recorded, yielding a total of 110,250 images. A large variety of object shapes, transparencies, and surface covers are considered. It makes this database quite interesting to evaluate object-based image retrieval approaches [27]. Some image samples of the training set and queries set are presented in Figures 5 and 6.
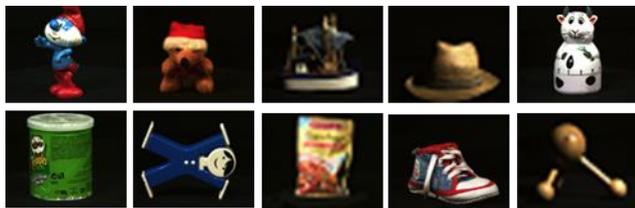

Fig. 5. Some sample images from ALOI dataset for the training

Fig. 6. Some sample images from ALOI dataset for the queries

B. Experimental Results

The performance of our system was measured using precision and recall measures. Recall measures the ability of the system to retrieve all the images that are relevant while precision measures the ability of the system to retrieve only the images that are relevant.

Eq. (1) is used to calculate the precision of the retrieval performance:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (1)$$

True Positives is the number of the images that are correctly retrieved from the image datasets. While, False Positives is the number of images that are incorrectly retrieved from the image datasets. In addition, the recall of the retrieval performance was calculated by Eq. (2):

$$Recall = \frac{TruePositives}{TruePositives + Missed} \quad (2)$$

The missed parameter is the number of relevant images that is not retrieved. Additionally, Precision-Recall graphs were used to measure the accuracy of our image retrieval system. They are used to evaluate the performance of any search engine.

All tests were performed on an HP-ElitBook-2740p laptop with Intel Core i5, 2.40 GHz processor, 4GB RAM, and Windows 7 Ultimate 64-bit as an operating system. The system was implemented in Microsoft Visual Studio 2013 using OpenCV version 2.4.9 for the graphical processing functions and C Sharp for the GUI design with EmguCV as a wrapper.

In the Flickr Logos 27 dataset, ten classes randomly selected (Google, FedEx, Porsche, Red Bull, Starbucks, Intel, Sprite, DHL, Vodafone, NBC) for training stage. The total number for training stage is of 300 images and 50 images in the testing stage. In the test stage, each image has been queried twice, once using SURF and other using SIFT. Precision and recall values appear directly below the images retrieved, as shown in Figures 7, 8. Table 1, Figure 9 (Precision-Recall graphs) show the values of the average of the precision and recall of all images in the test set with 10 class (5 images for each class).
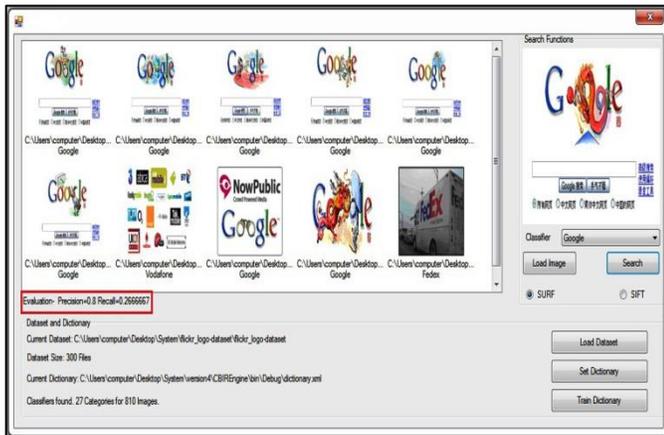
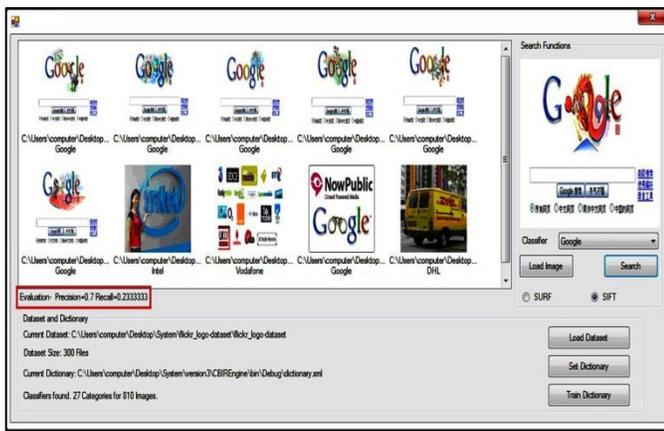Fig. 7. A snapshot of our proposed system in Flickr dataset using SURF technique

Fig. 8. A snapshot of our proposed system in Flickr dataset using SIFT technique

TABLE I.   THE AVERAGE OF THE PRECISION AND RECALL OF EACH CLASS (FLICKR LOGOS DATASET)

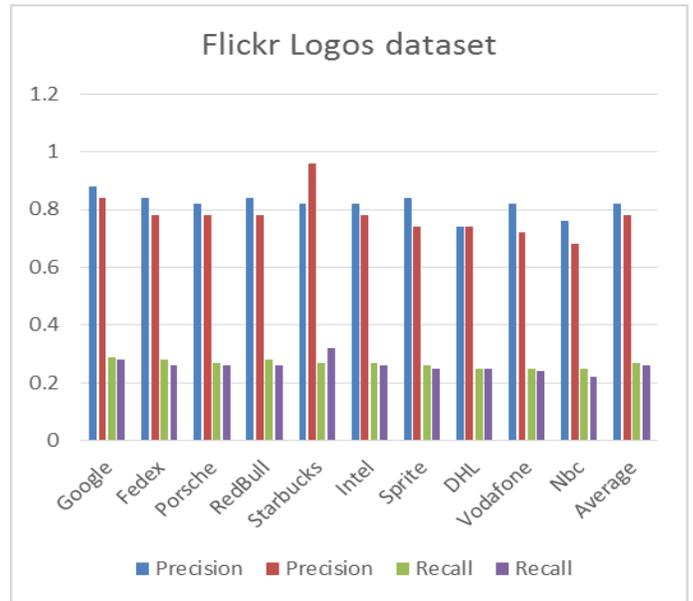| Class	Precision		Recall	
	SURF	SIFT	SURF	SIFT
Google	0.88	0.84	0.29	0.28
Fedex	0.84	0.78	0.28	0.26
Porsche	0.82	0.78	0.27	0.26
RedBull	0.84	0.78	0.28	0.26
Starbucks	0.82	0.96	0.27	0.32
Intel	0.82	0.78	0.27	0.26
Sprite	0.84	0.74	0.26	0.25
DHL	0.74	0.74	0.25	0.25
Vodafone	0.82	0.72	0.25	0.24
Nbc	0.76	0.68	0.25	0.22
Average	**0.82**	**0.78**	**0.27**	**0.26**

Fig. 9. The graph of the precision and Rcall of each class in Flickr Logos dataset

In ALOI dataset, the similar procedures that conducted for Flickr Logos 27 dataset were used. Accordingly, ten object images randomly selected from ALOI dataset (Big Smurf, Blue girls shoe, Boat, Christmas bear, cow kitchen clock, Green Pringles box, head, pasta and sugo, toy keys, Wooden massage) for training stage. Therefore, the total number of 300 object images for training stage and 50 object images for the testing stage. In the test set, each object has been queried twice, once using SURF and other using SIFT. Precision and recall values appear directly as shown in Figures 10, 11.
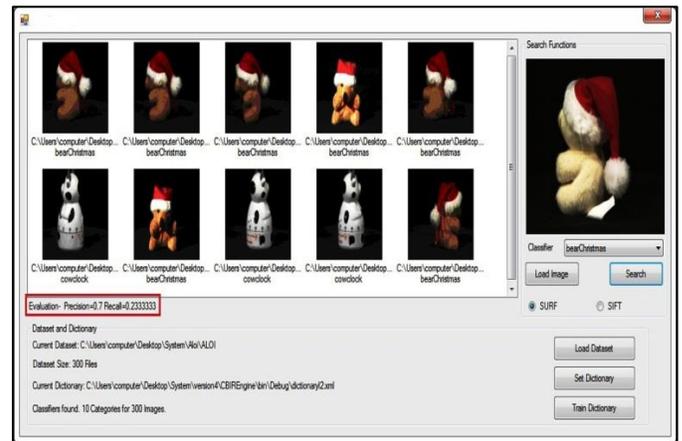

Fig. 10. A snapshot of our proposed system in ALOI dataset using SURF technique
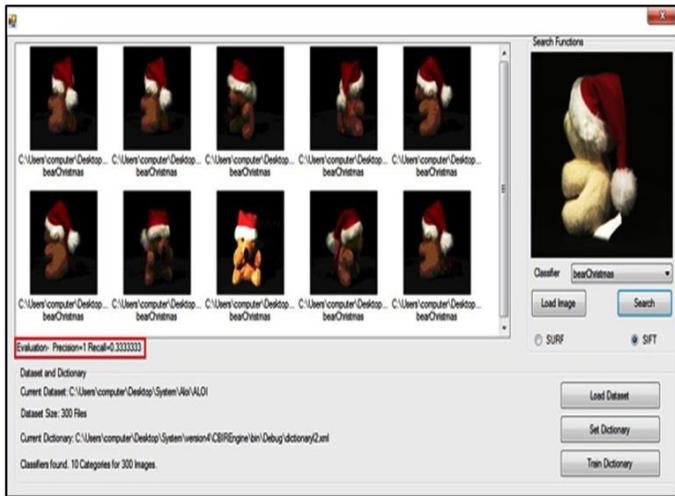
Fig. 11. A snapshot of our proposed system in ALOI dataset using SIFT technique

Table 2 and Figure 12 (Precision-Recall graphs) showing the values of the average of the precision and recall of all images in the test set with ten objects.

TABLE II. THE AVERAGE OF THE PRECISION AND RECALL OF EACH OBJECT (ALOI DATASET)

Object Name	Precision		Recall	
	SURF	SIFT	SURF	SIFT
Boat	0.94	0.94	0.31	0.31
Toy keys	0.84	0.76	0.28	0.25
Cow kitchen clock	0.82	0.84	0.27	0.28
Wooden massage	0.84	0.84	0.28	0.28
Big smurf	0.76	0.82	0.25	0.27
Pringles box	0.94	0.94	0.31	0.31
Pasta and sugo	0.96	0.94	0.32	0.31
Head	0.84	0.94	0.27	0.31
Blue girls shoe	0.94	0.96	0.31	0.32
Christmas bear	0.72	0.82	0.24	0.27
Average	**0.86**	**0.88**	**0.28**	**0.29**

As shown in the results of Fiker dataset, SURF algorithm was the batter than SIFT algorithm. The reason is that the SURF has good matching rate compared with SIFT. However, the results of SIFT with ALOI dataset was the better than SURF.

The reason may be due to the SIFT is more suitable for objects because it extracts more features. Also, maybe SURF are not robust enough in various imaging circumstances. However, seems both SIFT and SURF more suitable according of the type of the dataset.
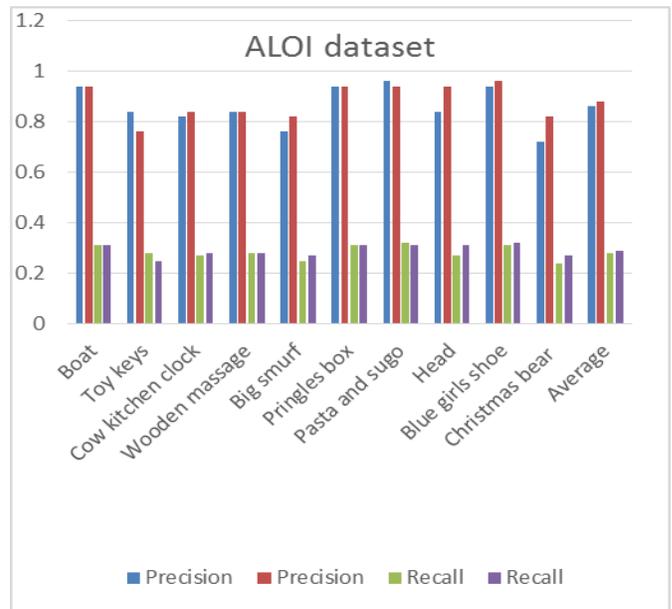

Fig. 12. The graph of the precision and Rcall of each object in ALOI dataset

VII. CONCLUSION

With advances in the multimedia technologies and the social networks, CBIR is considered an active research topic. Recent CBIR systems rely on the use of the BoVW model for being enables efficient indexing for local image features. This paper presented a system for CBIR, which uses local feature descriptors to produce image signatures that are invariant to rotation and scale. The system combines the robust techniques, such as SIFT, SURF, and BoVW, to enhance the retrieval process. In the system, we used a k-means algorithm to cluster the feature descriptors in order build a visual vocabulary. As well as, SVM is used as a classifier model to retrieve much more images relevant to the query efficiently in the features space.

We compared two different features descriptors techniques with BoVW model. Based on the experimental results, it is found that both SIFT and SURF are appropriate depending on the type of used dataset. The performance of the proposed system is evaluated by calculating the precision and recall on two different standard datasets. The experiments demonstrated the efficiency, scalability, and effectiveness of the proposed system.

In the future, we intend to study the possibility of improving the system performance using other local descriptors. We will do a comparative study between all of these descriptors according to illumination changes, scale changes, and noisy images on other types of standard datasets.

REFERENCES

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

[2] Halawani A., Teynor A., Setia L., Brunner G., and Burkhardt H. "Fundamentals and Applications of Image Retrieval: An Overview, " Datenbank-Spektrum, vol. 18, pp. 14–23. 2006.

[3] Aly M., Welinder P., Munich M., and Perona P. "Automatic discovery of image families: Global vs. local features, " Image Processing (ICIP), 16th IEEE Int. Conf. on, IEEE, pp. 777–780. 2009.

[4] Sivic, J., & Zisserman, A. "Video Google: A text retrieval approach to object matching in videos". In Computer Vision. Proceedings. Ninth IEEE International Conference on. pp. 1470-1477. 2003.

[5] Csurka G., Dance C., Fan L., Willamowski J., and Bray C. "Visual categorization with bags of keypoints, " Workshop on Statistical learning in compute. Vis., ECCV. Vol. 1. No. pp.1-22. 2004.

[6] Lowe D.G., "Distinctive image features from scale-invariant keypoints, " Int. journal of compute. Vis., vol. 60, pp. 91–110. 2004.

[7] Bay H., Tuytelaars T., and Van Gool L. "Surf: Speeded up robust features" Compute. Vision–ECCV, Springer, pp. 404–417. 2006.

[8] Liu, J. "Image retrieval based on bag-of-words model ". Cornell University Library:1304.5168,2013.

[9] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F,& Van Gool, L. "A comparison of affine region detectors". International journal of computer vision, 65(1-2), pp. 43-72.,2005.

[10] Mukherjee, J., Mukhopadhyay, J., & Mitra, P. "A survey on image retrieval performance of different bag of visual words indexing techniques" . In Students' Technology Symposium (TechSym), pp. 99-104. IEEE,2014.

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

[12] Hartigan, J. A., & Wong, M. A. "Algorithm AS 136: A k-means clustering algorithm". Applied Statistics,pp. 100-108. 1979.

[13] Chang, R. I., Lin, S. Y., Ho, J. M., Fann, C. W., & Wang, Y. C. (2012). "A novel content based image retrieval system using k-means/knn with feature extraction". Computer Science and Information Systems, 9(4), pp.1645-1661.2012.

[14] O'Hara, S., & Draper, B. A. "Introduction to the bag of features paradigm for image classification and retrieval". Cornell University Library:1101.3354.2011.

[15] Mikolajczyk, K., Leibe, B., & Schiele, B.). "Local features for the object class recognition." In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on (Vol. 2, pp. 1792-1799). IEEE.2005.

[16] Tuytelaars, T., & Mikolajczyk, K. "Local invariant feature detectors: a survey". Foundations and Trends® in Computer Graphics and Vision, 3(3), pp.177-280.2008.

[17] Boser, B. E., Guyon, I. M., & Vapnik, V. N. "A training algorithm for optimal margin classifiers". In Proceedings of the fifth annual workshop on Computational learning theory (pp. 144-152). ACM.1992.

[18] Li, J., & Allinson, N. M. "A comprehensive review of current local features for computer vision". Neurocomputing, 71(10), pp. 1771-1787. 2008.

[19] http://www.mathworks.com/help/vision/ug/local-feature-detection-and-extraction.html (Last accessed on 8/6/2015).

[20] Guerrero, M. "A comparative study of three image matching algorithms: SIFT, SURF, and Fast." Master Thesis (Utah State University) .2011.

[21] www.cs.bilkent.edu.tr (Last accessed on 11/6/2015).

[22] Zhang, J., & Ye, L. "Image retrieval based on a bag of images". In Image Processing (ICIP), 2009 16th IEEE International Conference on (pp. 1865-1868). IEEE.2009.

[23] Pönitz, T., Stöttinger, J., Donner, R., & Hanbury, "A. Efficient and Distinct Large Scale Bags of Words".Austrian Research Promotion Agency (FFG). 2010.

[24] Alfanindya, A., Hashim, N., & Eswaran, C. "Content-Based Image Retrieval and Classification using speeded-up robust features (SURF) and grouped bag-of-visual-words (GBoVW) ". In Technology, Informatics, Management, Engineering, and Environment (TIME-E), 2013 International Conference on (pp. 77-82). IEEE. 2013.

[25] http://image.ntua.gr/iva/datasets/flickr_logos/.(Last accessed on 17/6/2015).

[26] http://aloi.science.uva.nl/. (Last accessed on 19/6/2015).

[27] Geusebroek, J. M., Burghouts, G. J., & Smeulders, A. W. "The Amsterdam library of object images." International Journal of Computer Vision,61(1), pp.103-112. 2005.