Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 5, May 2015, pg.689 – 699

RESEARCH ARTICLE

LARGE SCALE IMAGE RETRIEVAL USING DESCRIPTORS AND DISTANCE MEASURE

[1]Subathra Muthuraman, [2]Mrs. Swathi Baswaraju, [3]Mrs. B.Mounica

[1]Student M.Tech Software Engineering, Department of Information Science and Engineering, New Horizon College of Engineering, Karnataka, India

[2]Assistant Professor, Department of Information Science and Engineering, New Horizon College of Engineering, Karnataka, India

[3]Assistant Professor, Department of Information Science and Engineering, New Horizon College of Engineering, Karnataka, India

vmsuba@gmail.com, baswarajuswathi@gmail.com, premkumarmounica@gmail.com

Abstract— Content Based Image Retrieval is the recent method that uses the features such as color, texture and shape of the images for correct matching and retrieval of images from a large database. The proposed approach uses the local features of images and their spatial information for retrieving images from a large database which will increase the accuracy of retrieval. Spatial context information is converted into binary codes for better geometric verification. The proposed approach uses the Scale Invariant Feature Transform and Speeded up Robust Feature descriptors to describe the interest points detected from the images. Hessian affine region detector is used to detect the corners from the images. Combination of descriptors improves the retrieval process and reduces the error rate.

Keywords— Image retrieval, spatial context information, binarization, geometric verification

I. INTRODUCTION

Searching and retrieving images from a large database is a great challenge for researchers. Recent technologies such as Internet have increased the growth of image collections which has lead to the need for retrieval tools.

Content Based Image Retrieval works with images. Features are extracted from the images as the images as a whole will be very large for indexing and retrieval. Feature vector will be generated as a result of feature extraction. In the proposed method, combination of descriptors is used to describe the interest points detected from the image which results in feature vectors. Hamming distance measure is used to find the minimum distance between the query image and database images. Based on the minimum distance, the images which are similar to query images are retrieved from the database. The proposed method is fast and accurate.

This article is organized as follows. Section 2 describes related work for content-based image retrieval. In section 3, the proposed architecture is explained in detail. Experimental results are given in section 4. In section 5, conclusion and future work are presented.

II. RELATED WORK

The feature-based image retrieval model can perform automatic indexing for a large volume of images by extracting the features of their content. The main features of images used for indexing and retrieving are colors, textures and shape [1][13][22]. It may retrieve the similar images matched with features of the image presented as query image. In [4], the author describes an algorithm for retrieving images using the shape information in an image. It has also considered the 3D information of the image. The proposed linear approximation procedure captures the depth information based on the idea of shape from shading. The objects are retrieved using the similarity measure that combines both the shape and the depth information. This approach has been effective in retrieving engineering objects.

In [7], the approach of Bag of Visual Words to retrieve the relevant word images from a big database correctly is discussed. This approach is based on the principles of text retrieval system. The representation of word images are in the form of histogram of visual words. The histogram carries the information of the features in the image. Visual words are quantized to represent local features in an image. Since Bag of Visual words method does not explain the spatial relationship among visual words; the author has applied re-ranking method to the retrieved list of images in order to improve the performance. The author has validated this approach on four Indian languages and it is proved to be language independent and scalable. The author has demonstrated the utility of the proposed system across four Indian languages by using the dataset of 100K words. To demonstrate the scalability, the author has used large dataset of 1M words. The performance is measures by precision. The limitation of this process is the re-ranking step which is time consuming.

Two methods are discussed in [10] such as, Scale Invariant Feature Transform to extract the unique features from the images, where features in the images remain the same whatever may be the image scale and rotation and Speeded Up Robust Features, a scale and rotation invariant interest point detector and descriptor which makes use of integral images. The author has discussed about the process of Image registration which converts the different groups of data into a single coordinate system. Image registration is a difficult task in many applications. In Image registration, first the features are detected and matched, then a transformation function is derived according to the features in the image and finally the image is reconstructed based on the transformation function. The author has taken two images for the experiment. The features were detected in both the images by the descriptors. The author found out that SIFT descriptor detects more features when compared to SURF, but SURF was found to be fast.

III.PROPOSED ARCHITECTURE

The high level architecture of the proposed image retrieval system is shown in the **figure 3.1.** The development process in the proposed image retrieval system consists of the following steps: User input (Query Image), Smoothening of query image, Feature extraction, binarization, similarity computation and retrieval of images.

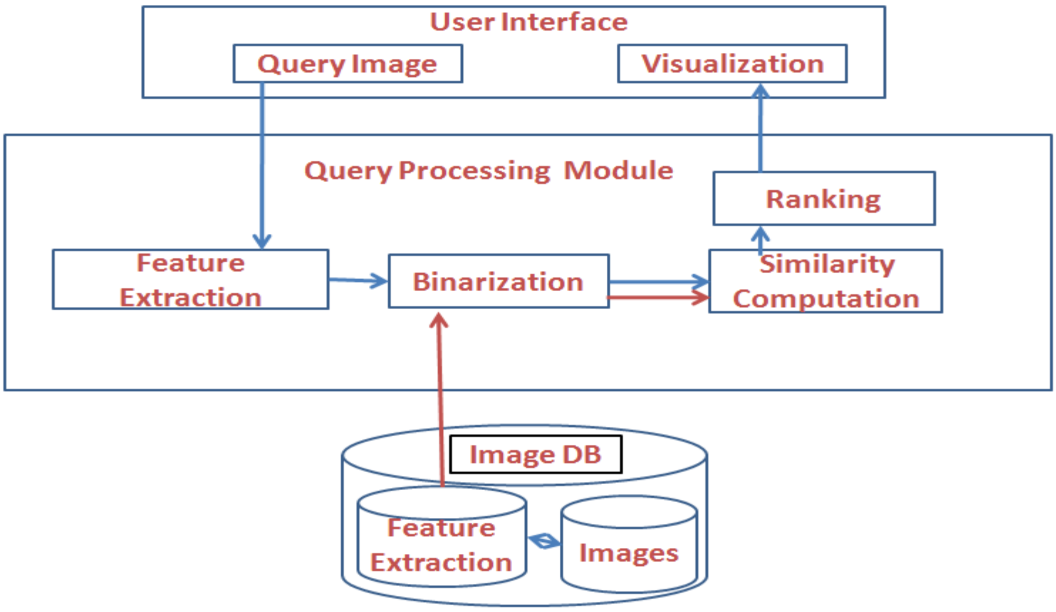


Figure.3.1 Architecture of the CBIR system

The query image is given by the user through the interface and the user can view the relevant images being retrieved through the same interface. Once the query image is given by the user, it is smoothened by using the filters in order to remove the noise from the image. The detail of the image is represented in numerical form after it is processed. Hessian affine region detector is used to detect the interest points from the images that are the corners. The given input image is smoothened by using the Gaussian filter in order to remove the noise from the image. Once it is smoothened, the derivatives are computed in order to find the variation in the intensity values. The second order derivative is also computed by the detector and it searches for the point where the hessian determinant becomes maximal. Hessian matrix is a square matrix as follows:

$$H = \begin{pmatrix} Ixx & Ixy \\ Ixy & Iyy \end{pmatrix}$$

The determinant of the hessian matrix is computed as follows:

$$Det (H) = IxxIyy – I^2xy$$

Equation 3.1: Determinant of Hessian Matrix

Searching is carried out by computing the resulting image containing the determinant value and then applying the non maximum suppression using a 3 * 3 window. This 3 * 3 window is swept over the full image. Now the comparison is done among the pixels. The pixels that has the greater value compared to its immediate neighbours is alone kept. Hessian affine detector gives the locations having a value above a pre defined threshold value. The resulting detector results are placed on corners.

Scale Invariant Feature Transform detects and describes the interest point of the images and generates the descriptors which are feature vectors. The given input image is smoothened using the guassian filter. The first and second order derivatives are calculated in order to find Laplacian of Guassian. The Laplacian of Guassian is calculated as

$$\Delta 2G = Ixx + Iyy$$

Equation 3.2: LOG calculation

The values are then arranged in a matrix format. The descriptor follows the following steps such as the detection of scale space extrema,localization of key points, orientation assignment ,and key point descriptor generation inorder to extract the key points from the images.

- **Scale Space Extrema detection**

Images are represented at different scales and this representation is parameterised by the size of the smoothening kernel.

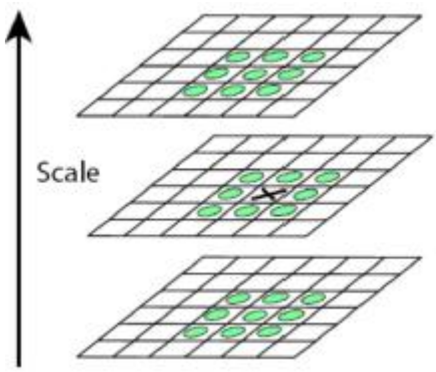


Figure 3.2: Scale Space Extrema Detection [23]

The selected pixel X in the figure has to be compared with the remaining 26 pixels in current and adjacent scales. By this comparison, the pixel that is larger or smaller than all 26 pixels is choosen.

- **Keypoint localization**

Points from the extrema detection is taken. Since there are lot no of points, the locations of keypoints may not be accurate. The outliers are rejected, that is the weak points are removed by setting up a threshold value.

- **Orientation Assignment**

An orientation is assigned to each keypoint. The keypoint descriptor can be represented relative to their orientation. It can achieve invariance to image rotation. The derivatives, gradient magnitude, and direction of the smoothened input image at the scale of keypoint is computed.

$$M(x,y) = sqrt (Ix^2 + Iy^2)$$
$$\theta(x,y) = \tan^{-1} Ix/Iy$$

Equation 3.3: Magnitude and Direction Computation

A wieghted direction histogram is created in a neighborhood of a key point. Weights are the gradient magnitudes. As shown in the figure, the peak is selected as the direction of the key point.

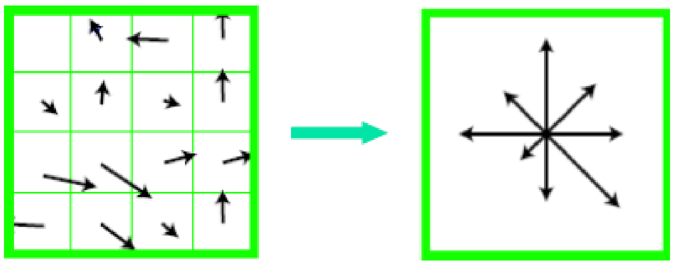


Figure 3.3: Orientation Assignment [23]

- **Keypoint Descriptor**

The relative orientation and magnitude in a 16 x 16 neighborhood at the key point is computed. Weighted histogram for 4 x 4 region is computed.The 16 histograms is concatenated in one long vector of 128 dimensions. The numbers are stored as feature vector.

8 x 8 to 2 x 2 descriptors

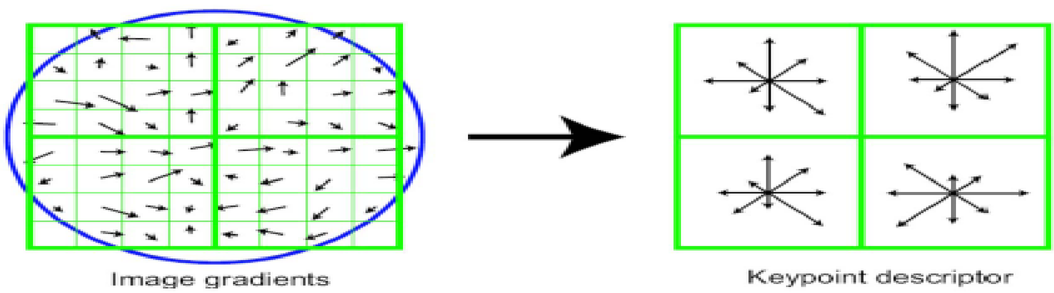


Figure 3.4: Keypoint Descriptor [23]

Speeded up Robust Feature descriptor is developed from the SIFT descriptors in order to improve the point detection methods. First, the algorithm constructs the integral image. The integral image is defined as follows:

$$I(x) = \sum_{i=0}^{i<=x} \sum_{j=0}^{j<=y} I(x,y)$$

$I_{(x,y)}$ = Input image x, y = spatial coordinate

Equation 3.4: Integral Image Construction

Then the Hessian Matrix is obtained and Scale Space is represented. To generate the orientations, the pixels are selected. It is generated based on the neighborhood of a particular interest point, so that a descriptor vector is obtained for every interest point. Thus the Key point descriptor is generated.

The feature vectors that are obtained is combined and converted into binary codes in order to find the similarity between the images. Binary codes help in faster computation.

Hamming distance measure is used to find the similarity between the query image and trained images. Hamming distance is used to find how the binary codes of the query and trained images are similar in terms of bit by bit.

In order to find the similarity between the query image and trained images, the distance measure is used. Hamming distance is used to find how the binary codes of the query image and trained images are similar in terms of bit by bit. Search results then can be sorted based on the distance to the queried image. Retrieval module selects the number of images to present to the user as a result to user queries based on the sorted distance.

IV. EXPERIMENTS

The experiment is carried out using MATLAB R2013a (8.1.0.604). For this experiment, a dataset with 150 images is used which is in JPEG format of size 256 x 256 as shown in figure 4.1. Searching of images is based on the similarity means rather than exact matching. Each query image returns the top ten images from database. The outcome measure to evaluate the proposed system is Precision and Recall. Precision is defined as the number of relevant images retrieved devise by the total number of images retrieved using the equation 4.1, whereas Recall is defined as the number of relevant images retrieved devise by the total number of relevant images using the equation 4.2. The average precision and recall values for the proposed system are given in table 4.2.

$$\text{Precision: } \frac{\text{No of relevant images retrieved}}{\text{Total No of images retrieved}}$$

Equation4.1: Precision calculation [31]

$$\text{Recall: } \frac{\text{No of relevant images retrieved}}{\text{Total No of relevant images}}$$

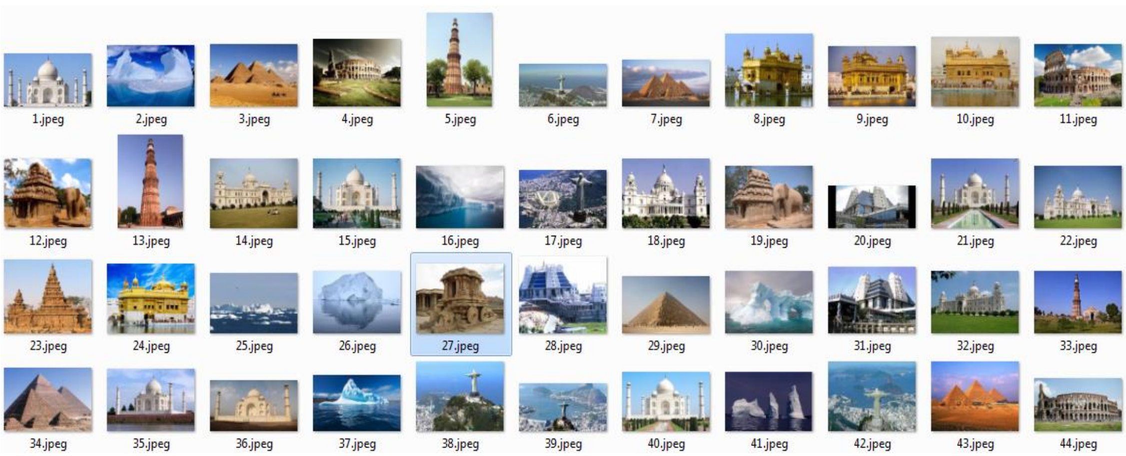
Equation4.2: Recall calculation [31]

In the proposed system, similarity comparison is done between the query image and dataset images based on the distance computed. The distance measure used here is Hamming distance.

$$\text{Hamming distance} = \sqrt{((\text{mean (query image)} - \text{mean (dataset image)})\ \text{^}2}$$

Equation4.3: Hamming Distance

The dataset used for the experiment contains 150 images which are shown below:



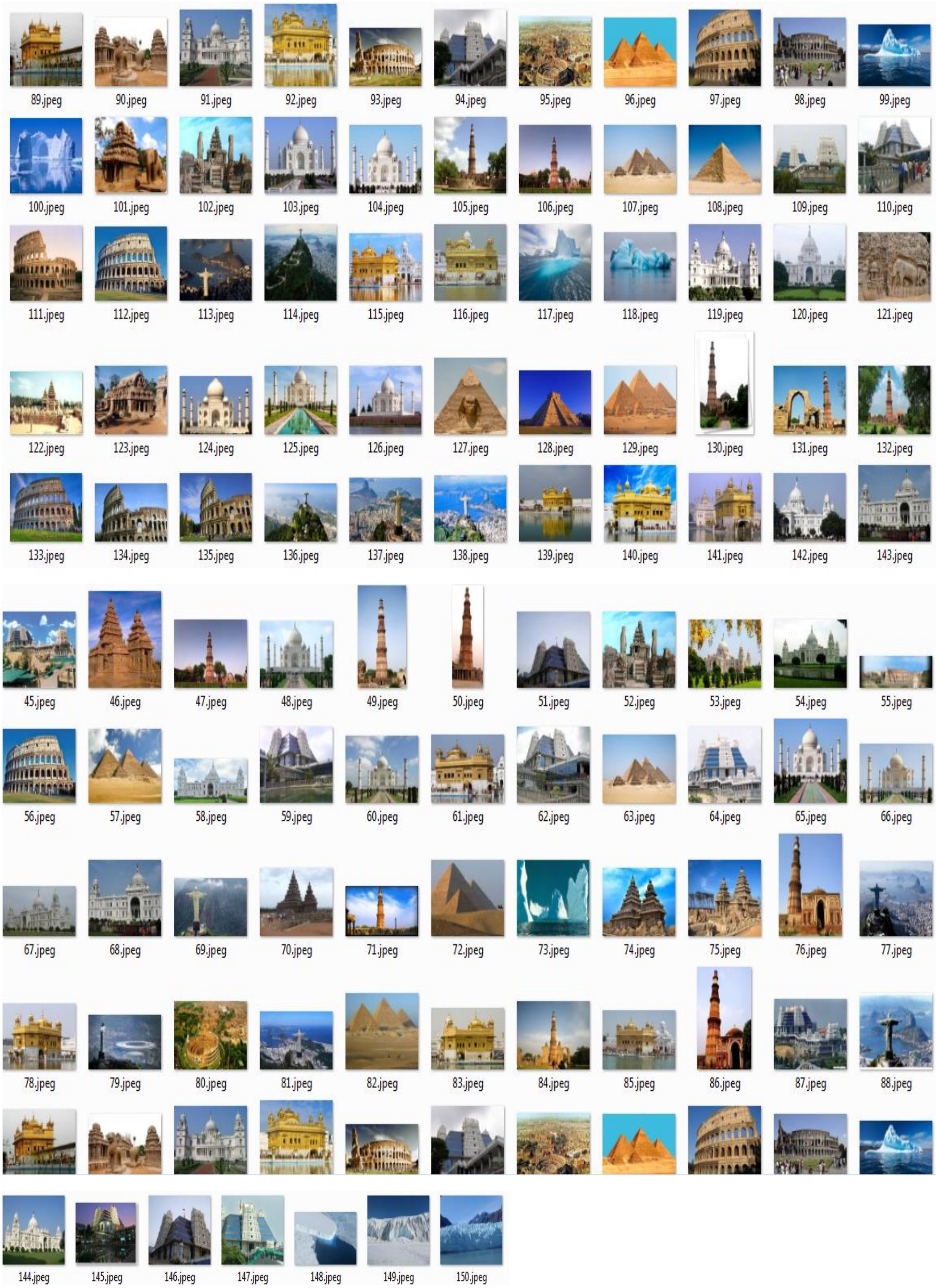


Figure 7.1 Dataset used for the proposed system

The following are the different categories of images in the dataset:

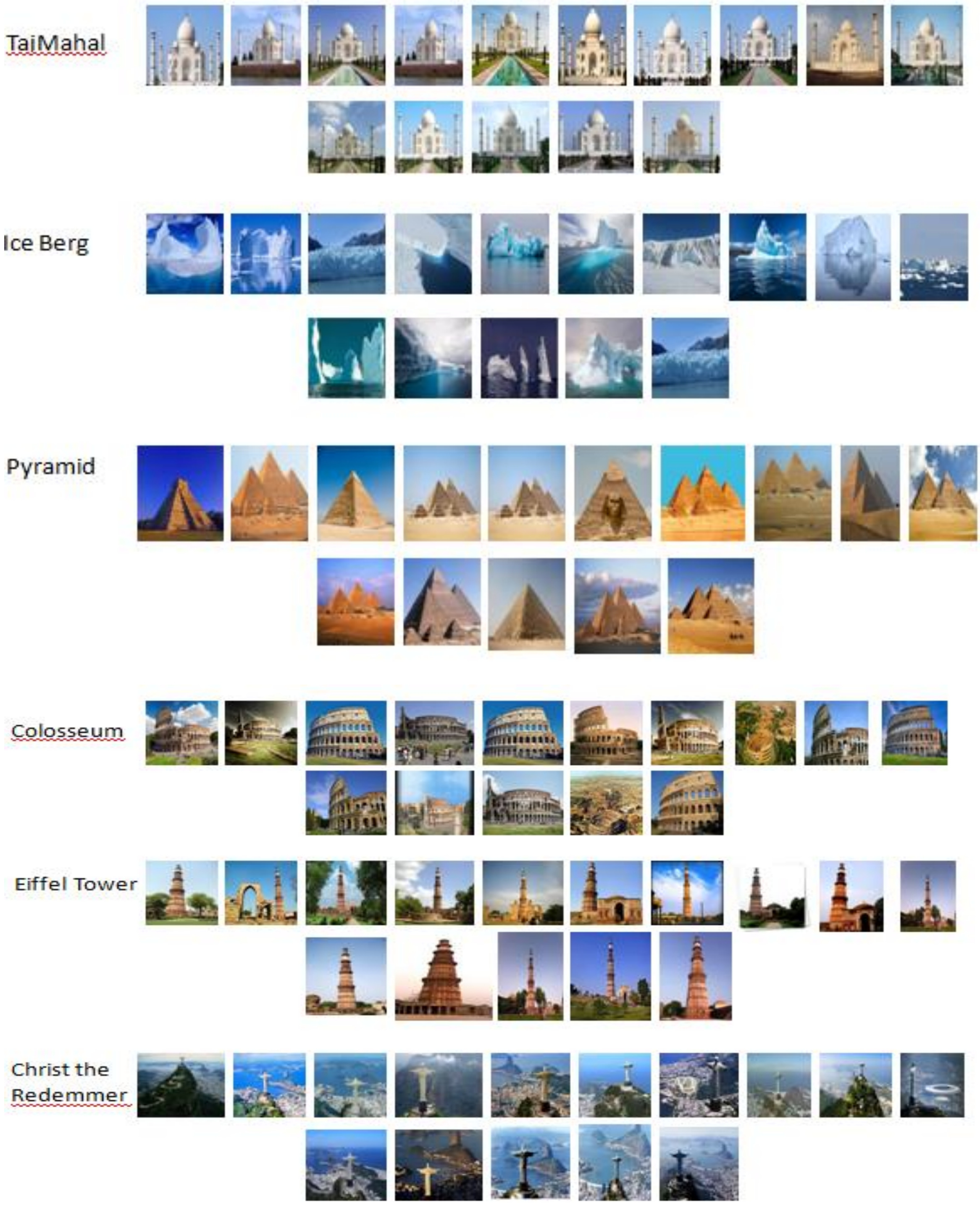


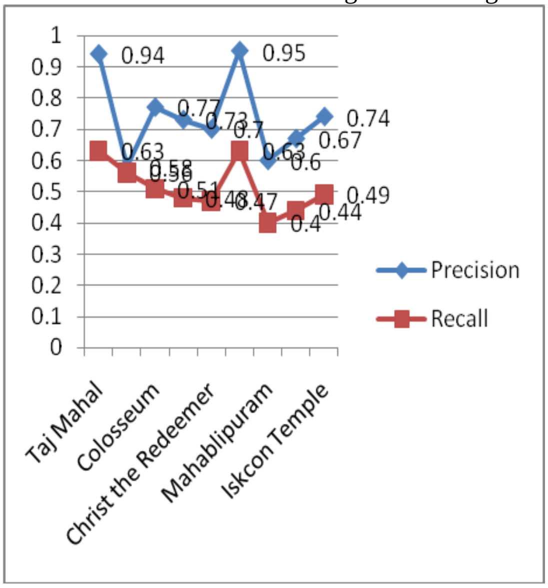
Figure 4.2 Different categories of images used for the proposed system

The Precision and Recall values for different categories of images are:

Category	Precision	Recall
Taj Mahal	0.94	0.63
Ice Berg	0.58	0.56
Colosseum	0.77	0.51
Eiffel Tower	0.73	0.48
Christ the Redeemer	0.70	0.47
Harmandir Sahib	0.95	0.63
Mahablipuram	0.60	0.40
Victoria Palace	0.67	0.44
Iskcon Temple	0.74	0.49

Table 4.1: Precion and Recall for each category

The graph showing the precision and recall values for different categories of images are:



Graph 4.1: Precision and Recall for each category

The list of images similar to the query image are :

Retrieved relevant images based on the minimum similarity distance

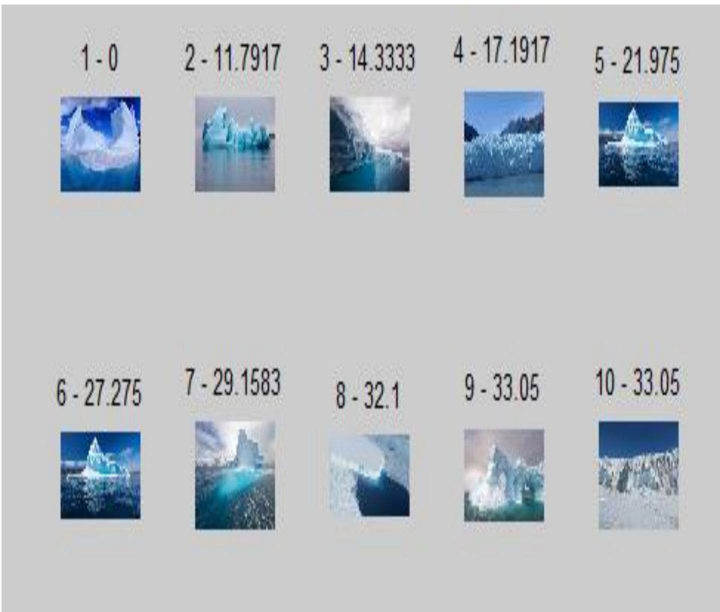


Figure 4.3: Retrieved list of images

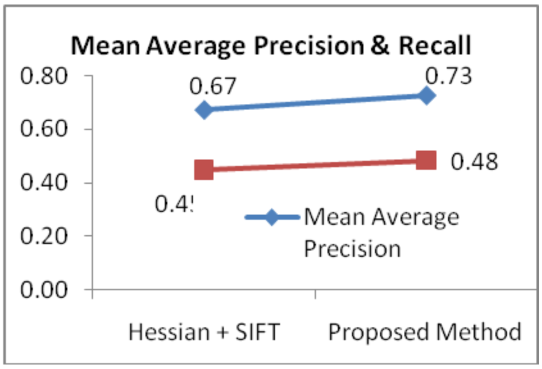
The Precision and the Recall measure for the proposed method is as follows:

Method	Mean Average Precision	Recall
Hessian + SIFT	0.67	0.45
Proposed Method	0.73	0.48

Table 4.1: Mean Average Precion and Recall

Comparison Graph

The proposed algorithm was tested for different type of images and found working effectively.Mean Average Precision increases when more descriptors are used to describe the interest points and spatial context.



Graph 4.2: Mean Average Precision and Recall

V. CONCLUSIONS

The proposed image retrieval system takes in to account the features, their spatial information for better matching and retrieval. The features spatial context information is represented in binary code. The experiment carried out on the dataset verifies the efficiency of the algorithm. In future, work will be carried out in reducing the number of features in images, with the same retrieval performance. The experiment will be tested on the complicated dataset images.

ACKNOWLEDGEMENT

I would like to express my thanks to **Mrs. Swathi Baswaraju,** Assistant Professor, Dept. of ISE, and **Mrs. B.Mounica**, Assistant Professor, Dept.of ISE, New Horizon College of Engineering, Bengaluru who have always guided me in detailed technical aspects throughout this work. I would also like to express my thanks to **Mr. Lokesh M.R,** Senior Assistant Professor, Dept of ISE, New Horizon College of Engineering, Bengaluru for his valuable suggestions throughout this work.

REFERENCES

[1] Charles E. Jacobs, Adam Finkelstein, and David H.Salesin, "Fast multiresolution image querying," Proceedings of SIGGRAPH95, Los Angeles, California, August 1995.
[2]. Jerome Landr, Frederic Truchetet, "Fast Image Retrieval using Hierarchical Binary Signatures", ISSPA on Signal Processing and its Applications, 2007.
[3]. Amit Jain, Ramanathan Muthuganapathy, and Karthik Ramani, "Content-Based Image Retrieval Using Shape and Depth from an Engineering Database", G. Bebis et al. (Eds.): *ISVC* 2007, Part II, LNCS 4842, pp. 255–264, 2007.
[4]. P M Panchal, S R Panchal, S K Shah," A Comparison of SIFT and SURF", *IJIRCCE*, Vol.1.Issue 2, April 2003.
[5]. Ravi Shekhar and C.V. Jawahar, "Word Image Retrieval using Bag of Visual Words", in Proc.IEEE, 2012.
[6]. Subathra Muthuraman,Mrs.Swathi Baswaraju,Mrs.B.Mounica, "Survey on Contextual Hashing for Searching and Retrieval of Images using Descriptors and Distance Measure",*IJERT*,Vol.4,Issue No.1,Jan 2015.

[7]. Zhen Liu, Houqiang Li, Wengang Zhou, Ruizhen Zhao, and Qi Tian,"Contextual Hashing for Large-Scale Image Search", IEEE transactions,VOL.23, No.4, April 2014."
[8]. [1].J.Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(9), 947–963, 2001.
[9].C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-maximization and Its Application to Image Querying", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(8):1026-1038, 2002.
[10].J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos", in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec.2003, pp. 1470–1477.
[11].Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang, "Spatial-bag-of features", in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3352–3359.
[12]. Marçal Rusinol, Josep Llados," Efficient Logo Retrieval through Hashing Shape Context Descriptors", in Proc. ACM, 2010.
[13]. Manimala Singha and K.Hemachandran, "Content Based Image Retrieval using
Color and Texture", *Signal & Image Processing: An International Journal (SIPIJ)* Vol.3, No.1, February 2012.
[14]. W. Zhou, H. Li, Y. Lu, and Q. Tian, "Principal visual word discovery for automatic license plate detection", *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4269–4279, Sep. 2012.
[15].P M Panchal, S R Panchal, S K Shah," A Comparison of SIFT and SURF", *IJIRCCE*, Vol.1.Issue 2, April 2003.
[16]. S. Zhang, Q. Tian, K. Lu, Q. Huang, and W. Gao, "Edge-sift: Discriminative binary descriptor for scalable partial-duplicate mobile search," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2889–2902, Jul. 2013.
[17]. S. Zhang, M. Yang, X. Wang, Y. Lin, and Q. Tian, "Semantic-aware co-indexing for image retrieval", in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Apr. 2013, pp. 1–8.
[18]. H.Kavitha, M.V.Sudhamani, "Object based Image Retrieval from Database using Combined Features", *International Journal of Computer Applications*, Vol.76, Issue No.8, August 2013.
[19]. Deepak John, S.T.Tharani, and SreeKumar.K, "Content Based Image Retrieval using HSV-Color Histogram and GLCM", *An International Journal (IJARCSMS)*, Vol.2, Issue No.1, January 2014.
[20]. Ji Wan, Dayong Wang, Steven C.H. Hoi, Pengcheng Wu,Jianke Zhu, Yongdong Zhang, Jintao Li," Deep Learning for Content-Based Image Retrieval: A Comprehensive Study", in *Proc. ACM Int. Conf. Multimedia*, 2014.
[21]. Lowe, D. "Distinctive image features from scale-invariant keypoints" International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.
[22]. http*:// en.wikipedia.org/wiki/Content-based_image_retrieval*
[23]. http://en.wikipedia.org/wiki/Scale_space
[24]. http://in.mathworks.com/products/matlab
[25]. http://en.wikipedia.org/wiki/Hessian_affine_region_detector
[26]. www.vision.ee.ethz.ch/~surf/eccv06.pdf
[27]. http://crcv.ucf.edu/people/faculty/shah.php

