ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

Gaze Based Image Retrieval Using Webcam

Teena Mariya Babu [1] , Shibu K R [2]

[1] M.Tech Student, Dept. of CSE., VJCET, MG University, Kerala, India

[2] Assistant Professor, Dept. of CSE., VJCET, MG University, Kerala, India

ABSTRACT: Content based Image Retrieval (CBIR) has been a widely researched topic over the last decades and the inaccurate retrieval of relevant images with respect to the given query image is one of its main challenges. The implicit form of image retrieval has been receiving much attention in the last few years as it is useful in Human Computer Interaction (HCI). But its research is limited by the cost effective specialized devices. Thus this paper mainly focuses on a content based image retrieval which avoids the need of those kinds of devices. This is carried out by capturing user's gaze using a webcam, followed by a Bag of Words (BoW) methodology for performing the CBIR. The Gaze capturing is done by displaying images in an interface and then the gaze feature is extracted from the captured image which depends on number of times the user's gaze lies at a particular image. Thus the proposed work constitutes a less complex method for gaze tracking using a webcam. The CBIR is done by extracting low level features from image and representing it using BoW model. The WANG DB, a general purpose database which contains 1000 images is used for the testing purposes.

KEYWORDS: Content-based image retrieval (CBIR); Gaze Tracking; Human Computer Interaction; Image retrieval; Image search; Image similarity

I. INTRODUCTION

The development of mobile devices and the emergence of new companies are increasing day by day. A lot of mobile products with great specifications and cheap price were released this year among the flagship products by the leading companies and flagship killers. This competition has led to an astronomical increase of smart phones, tablets and other sophisticated mobile devices. Therefore a vast collection of images are generated mainly through social networking. As a result, an efficient method for image retrieval has gained wide importance in the current scenario. Several techniques also arrived in the past decades for image retrieval.

The extraction of low-level features from an image plays an important role in the content based image retrieval **(**CBIR). Colour, texture and shape are the major features that are generally used. Apart from that region based image retrieval and relevance feedback systems can also be used to increase the presence of relevant images in retrieval results [1]. Region based image retrieval depends on segmentation algorithms and thus each region is considered for feature extraction instead of the entire image. The selection of an effective segmentation algorithm plays the key role in those kinds of systems. In feedback based systems, images are retrieved based on either positive or negative feedback provided by the user and this can be continued till the user stops giving feedback.

Inputs to an image retrieval system can be gathered either implicitly or explicitly [1]. In explicit methods, user needs to explicitly provide details about the query image. On the other hand, in implicit methods some information about the user which is obtained from some devices to indicate the user input. The latter has been receiving much attention for the last few years. This led to the development of several gaze based CBIR systems using specialized gaze trackers. However the usage of these gaze trackers in image retrieval applications are limited by the price, non-portability and invasiveness etc.

Inaccurate retrieval of images is one of the main challenges of CBIR systems [2], [3]. Hence a solution is needed to reduce the presence of irrelevant images in the retrieval results. Furthermore, gaze tracking is a more common technique among implicit image retrieval as it is highly useful in Human Computer Interaction (HCI) and a convenient form of image retrieval. Existing implicit image retrieval techniques using gaze tracking [1], [4], [5], [6], are based on specialized cameras or sophisticated eye trackers like Tobii, SMI, Mirametrix etc. Thus the gaze based image retrieval, which tracks eye gaze has been an active research area nowadays, but it suffers from problems of non-portability and cost. Electroencephalography (EEG) [7] is yet another way to implement implicit image retrieval but it still suffers

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

from the some of the above problems. Therefore it is very clear that in the current scenario, a solution is needed to overcome these limitations of existing gaze based image retrieval techniques.

In this paper, we propose a gaze image retrieval technique using a webcam. The major contribution is the development of a gaze tracking technique using a webcam. Another contribution is the identification of a set of features to increase the precision of the existing CBIR techniques with BoW and SIFT. Thus a CBIR with BoW using SIFT, HOG and HSV color histogram is suggested to improve the precision of retrieval results.

II. RELATED WORK

A survey was conducted on existing CBIR, existing implicit image retrieval techniques with a main focus towards the gaze tracking techniques. Thus the inaccurate retrieval of images poses one of the greatest challenges of CBIR [2], [3]. The basic idea of proposed technique is based on [1], particularly the general idea of gaze tracking. It is a gaze based image retrieval for region based image retrieval using a single camera. An image retrieval system that integrates eye tracking and BoW architecture was proposed in [4]. It uses an SMI eye tracker and fixation duration as the gaze feature. In [1], [4], the CBIR is performed using SIFT based BoWs. Kozma *et al*. [5], introduced a GaZIR, a dynamic interface for browsing images using eye tracking. GaZIR uses a Tobii 1750 eye tracker and fixations and saccades are considered as gaze features. A method for implicit image annotation and retrieval by implicitly monitoring user attention through eye-tracking is suggested in [6]. It used a binocular set of 60-Hz cameras with infra-red filters and the eye-tracking technology is the faceLAB 5.0 software package. The image retrieval [7] is yet another example of implicit image retrieval using the power of brain state decoding and visual content analysis.

Several other techniques focussed on region based image retrieval and relevance feedback to improve the precision of retrieval results. In [8], an effective region-based image retrieval based on explicit relevance feedback from the user is given. An explicit region-based image retrieval using the concept of region codes can be seen in [9]. A region based image retrieval based on an automatic relevance feedback without the need of a real user is proposed in [10].

III. SYSTEM ARCHITECTURE

The overview of the system is shown in Fig 1. The main features include a gaze tracking framework and an image retrieval system using BoW [15]. Images are displayed to the user using an image interface. The user's gaze is captured using a webcam placed on the top of the screen. The query image is determined by the occurrences of fixation. Thus image having maximum fixation occurrences is the query image. Image features are extracted from the query image. They include SIFT descriptor [11], HOG descriptor [12] and HSV histogram [13], [17]. Code words are generated for those features and compared with that of database. Finally, the Top-N images are retrieved as CBIR results

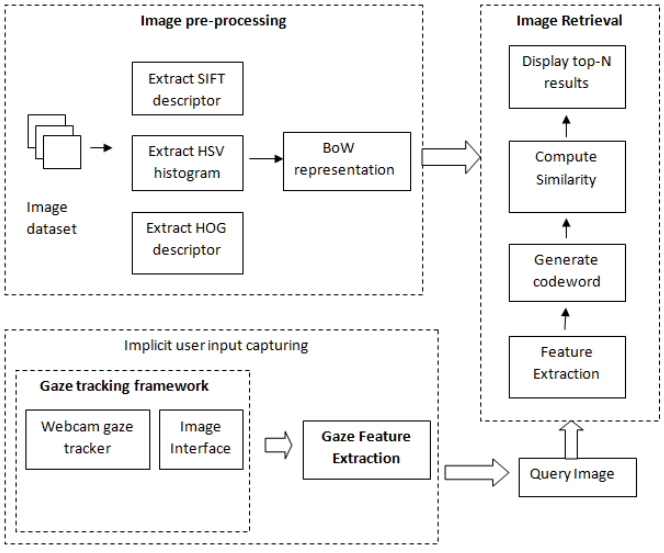

Fig.1. System Architecture.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

A. *Gaze Tracker:*

The working of the proposed gaze-tracking framework is described in this section. The basic idea of gaze tracking is based on the method proposed in [1]. Every incoming frame is transformed to the gray scale color space and followed by a contrast stretching and enhancement. Initially the user needs to manually mark the centre of two eyes in an off-line step. Then a set of points surrounding this two eye centre are automatically extracted. These set of points are called tracking points that are used to track the position of eye centre in the next frame. Then for every subsequent frame, the corresponding location of eye centre is estimated according to the average displacement of the set of tracking points. The corresponding set of tracking points in every subsequent frame is determined by computing the Pyramidal Lucas-Kanade optical flow [14] on two adjacent frames. Thus the centre of two eyes obtained from a frame is used to extract the eye area of both eyes.

A hard thresholding technique is subsequently applied to these extracted eye area to segment the region of iris from the background. This is followed by a hole filling operation to fill the holes in iris region, which are created due to the reflections caused by the background light. Then by applying binary erosion on this extracted iris region will remove all small black areas other than the iris. This is to done to clearly get the iris region in the eye area. The results after applying these operations on the extracted eye area are shown in Fig. 2.

(a) (b) (c) (d)

Fig 2. Extraction of region of iris: (a) Extracted eye area. (b) After thresholding. (c) After filling holes. (d) After erosion.

The gaze tracking framework includes a USB webcam and an image interface. In order to track user's gaze, initially user needs to mark eye centers for the first frame. Based on this a set of tracking points around eye centre are automatically extracted using two bounding boxes. A bounding box 1 of width 27 pixels and height 15 pixels from eye centre is used to represent eye area. Another bounding box 2 of width 65 pixels and height 90 pixels from eye centre is used to represent area around eyes to select tracking points. Only the tracking points within a displacement of 8 pixels coming under bounding box 2 are chosen. Also exclude the tracking points coming under bounding box 1.

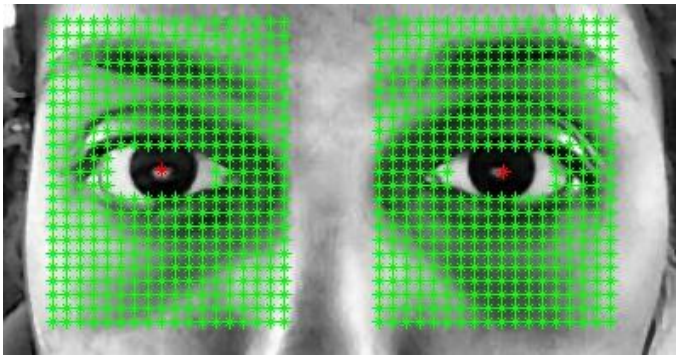

Fig.3. Tracking points around eye centre.

Fig. 3 shows extraction of tracking points around eye centres. Here red point and green points indicate eye centres and tracking points respectively of two eyes. Then for each incoming frame, repeat this procedure. Then the tracking points of subsequent frames are determined using Pyramidal Lucas Kanade optical flow and mean of their displacement is used to determine eye centres for current frame. Also exclude tracking points that are incorrectly tracked using the above mentioned bounding boxes. Then the eye area is extracted using bounding box 1 around eye centres of current frame. This is followed by the application of three operations to the extracted eye area .i.e, binary thresholding, hole filling, binary erosion.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

B. *Image Interface:*

The image interface consists of six blocks. Each image is placed on a particular block. Thus six images are shown to the user in the interface. Images are arranged in two rows of three images each. Resolution of 1200 x 600 is used for the interface image. The dimension of each image depicted in the interface constitutes 400 x 300. Every time the images are displayed to user randomly from the database. The diagram of image interface is shown in Fig. 4.

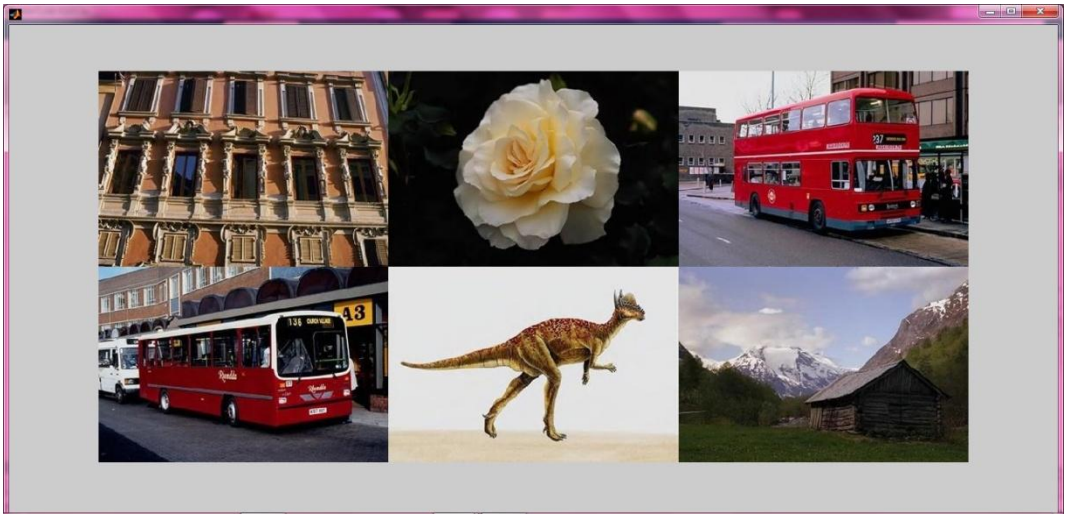

Fig.4. Image Interface

C. *Gaze Feature Extraction*

The gaze feature extraction is based on the concept of fixation [1], [4], [5]. A fixation occurs on an image when user's gaze lies on a particular block. User's gaze may lie on several images shown in the blocks. Thus each block has its own occurrences of fixation. Thus the goal here is to determine the block having maximum fixation count. For that, divide the eroded eye area into six blocks, i.e., two rows and three columns. Then compute the percentage of iris in each block and the height of iris. The total iris percentage across each vertical partition determines the column in which the user is gazing. The row is determined by considering a threshold for height of iris. A threshold of 18 pixels was used for experiments. This may vary from individual to individual. Thus a fixation on that particular block is obtained by determining the column and row. All these procedures are applied to each incoming frames. Finally the block having maximum fixation count is considered as the Block of Interest (BOI). Then query image is the image representing that particular BOI.

D. *Image Pre-processing and Retrieval:*

Image pre-processing and retrieval is based on BoW methodology [15]. The low level features such as SIFT descriptor, block based HOG descriptor and block based HSV color histogram are used for feature extraction. The size of codebook taken is 100, 8 and 8 for SIFT, HOG, and HSV histogram respectively. For block based HOG and block based HSV histogram, image is divided into 4 x 4 blocks. In block based HOG, 16 orientation bins are considered from each block. In block based HSV histogram, each block represents a 32 bin descriptor with 8 bins for hue, 2 bins for saturation and 2 bins for value. For image retrieval, input query is the image with maximum fixation count. The normalized manhattan distance [16] is taken as the similarity metric. In proposed work (BoW-SIFT+HOG+HSV), apply the following weights to this similarity scores: 80% for SIFT score, 10 % for both HOG and HSV. Finally, the three similarity scores are combined. Thus images are ranked based on the similarity score which is calculated using eq. (2) [18] and finally display top-N results are shown to the user.

$$MD = \sum_{c=1}^{C} | Xc - Yc | \qquad \text{eq. (1)}$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

where *MD* is the manhattan distance, C indicates the size of the codeword for a particular image feature and X, Y indicate two codewords.

$$NMD = \frac{MD}{C} \qquad \text{eq. (2)}$$

where *NMD* is the normalized manhattan distance

IV. EXPERIMENTAL RESULTS

The proposed method is implemented using MATLAB R2013a on a 64-bit Windows7-PC with a 2.2 GHz processor and 2GB RAM memory. The database used to evaluate proposed work is the WANG DB which is obtained from [19]. The WANG database [20], [21], [22] is a subset of 1,000 images of the Corel stock photo database which form 10 classes of 100 images each. The images are in JPEG format of size 384x256 and 256x386.

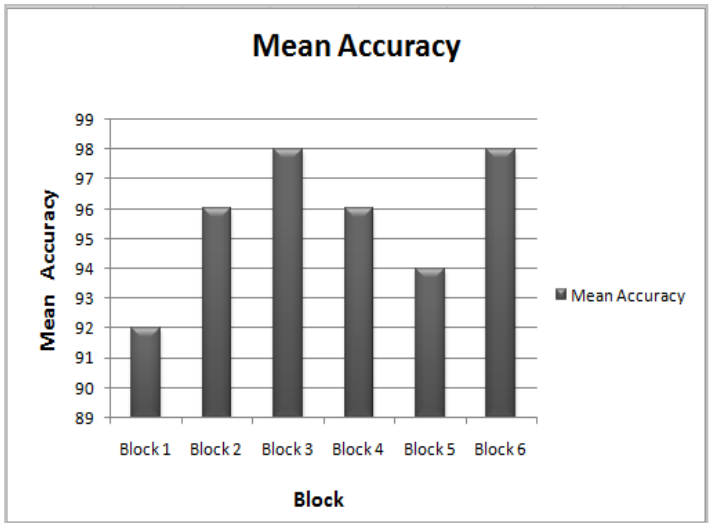

Fig.5. Gaze tracking results

For gaze tracking, mean accuracy of each block is evaluated using 5 users. 10 frames of each user's gaze are captured for each block. Each user is asked to gaze at one block at a time. While testing accuracy of proposed gaze tracker, each user was instructed to follow the instructions to be followed during gaze tracking such as place the webcam at the top and middle of the screen, adequate lightening conditions, no head movement, be close to the camera, adjust the eye level below the webcam. Hence based on the mean accuracy graph shown in Fig. 5, it is observed that an accuracy of 90 percent can be easily obtained if the user is strictly following the instructions to be followed during gaze tracking.

The performance of proposed image retrieval is evaluated using WANG DB. The image retrieval results obtained by both existing and proposed work are shown in Fig. 6 and Fig. 7 respectively for the query image of a dynosaur.

Fig.6. Existing technique results

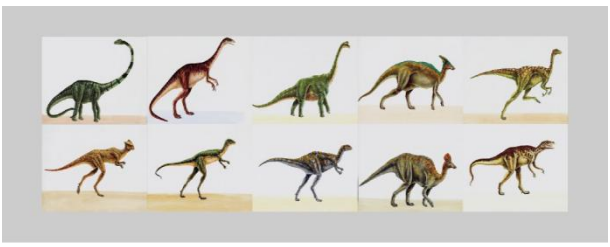

Fig 7. Proposed technique results

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

The performance of image retrieval of the system can be measured in terms of its recall and precision [23]. Precision is the number of correct images with respect to the number of all returned images. Here only the top 10 images are considered. Recall is the number of correct images divided by the total number of images in the database. Precision and recall of the given query image of dynosaur obtained using proposed work and existing work are shown in Table 1.

Table 1. Results of given query

Techique	Features	Relevant images	Precision	Recall
Proposed	BoW-SIFT+HOG+HSV	10	100	10
Existing	BoW-SIFT	7	70	7

The performances of all categories are also measured in terms of mean precision. Mean precision for a set of queries is the mean of the precision scores for each query. To compute mean precision, 50 images from each category are considered resulting into a test set of 500 images. The results are shown in the Mean-precision graph of Fig. 8 to get a more clear view. Here, x axis indicates different categories of images and y-axis indicates mean precision obtained for each category considering 50 query images per category. It is very clear that the proposed technique shows a higher mean precision for all categories when compared to existing technique. Categories like horses, dynosaur showed best improvement whereas category the beach shown the least improvement using the proposed work. Thus the object detection property of HOG is clearly visible.

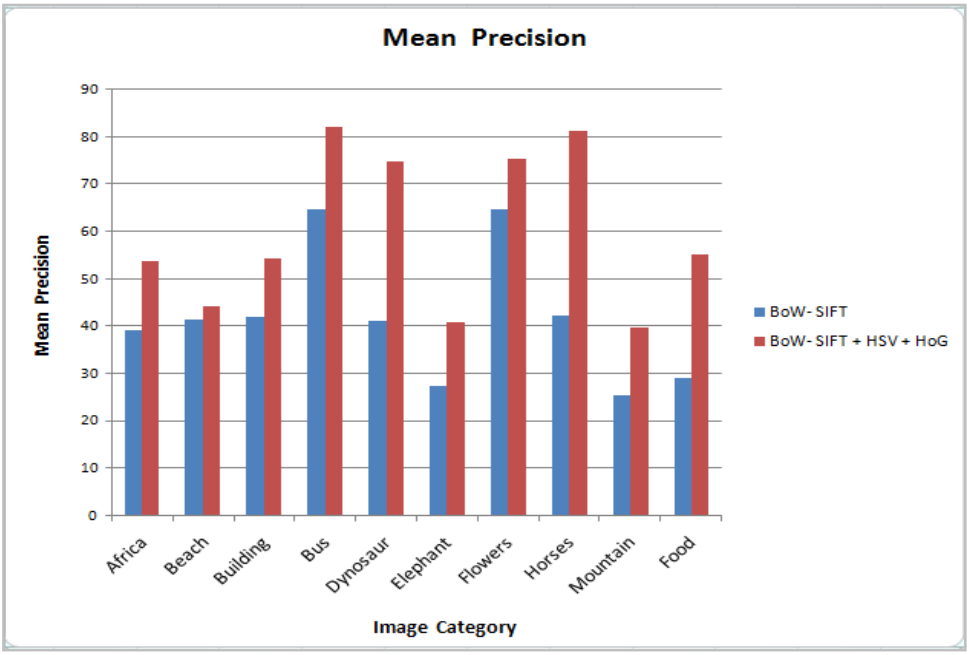

Fig 8. Image retrieval results

V. CONCLUSION AND FUTURE WORK

Content based image retrieval is used to retrieve images based on their content. Eye tracking based CBIR has been gaining much attention in the past few years with the use of several highly expensive gaze trackers. Thus a CBIR system is proposed to retrieve images using a webcam for eye tracking based image retrieval. This has less complexity when compared to the expensive commercially available eye trackers. The proposed work can act as a foundation to the future CBIR systems for cheap and portable way of eye tracking and thus to allow a more natural and user convenient way of image retrieval. Apart from that, the proposed work also suggests the use of a set of features using BoW-SIFT+HSV+HOG to improve the existing CBIR using BoW-SIFT. This new set of features enabled to increase

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

precision of image retrieval. Finally the evaluation of image retrieval and gaze tracking are also done to show the performance of the proposed system. Future works mainly include the development of methods for the automatic determination of the eye centre instead of manually marking it and thus to make the system fully automated. Furthermore, it also include the development of techniques to increase the speed of the overall system and hence to make it successful in real time applications.

REFERENCES

1. G. Th. Papadopoulos, K. C. Apostolakis, and P. Daras, "Gaze-Based Relevance Feedback for Realizing Region-Based Image Retrieval," IEEE Transactions on Multimedia, Vol. 16, No. 2, Feb 2014.
2. R. datta, D. joshi, J. li, and J. z. Wang "Image Retrieval: Ideas, Influences, and Trends of the New Age" ACM Computing Surveys, Vol. 40, No. 2, Article 5, Feb 2008.
3. A. N Bhute and B. B. Meshram , "Content Based Image Indexing and Retrieval" International Journal of Graphics & Image Processing, Vol 3, Issue 4, Nov. 2013.
4. Q. Li, M. Tian, and J. Liu, "A Novel Image Retrieval System with Real-Time Eye Tracking," ICIMCS, July 2014.
5. L. Kozma, A. Klami, and S. Kaski, "GaZIR: Gaze-based Zooming Interface for Image Retrieval, "ICMI-MLMI, 2009.
6. S. N. Hajimirza, M. Proulx, and E. Izquierdo, "Reading Users Minds From Their Eyes: A Method for Implicit Image Annotation," IEEE Transactions on Multimedia, Vol. 14, No. 3, June 2012.
7. J. Wang, E. Pohlmeyer, and B. Hanna, "Brain State Decoding for Rapid Image Retrieval," MM, Oct 2009.
8. M. Fang, Y. Kuan, C. Kuo and C. Hsieh, "Effective image retrieval techniques based on novel salient region segmentation and relevance feedback," Multimedia Tools and Applications, Vol. 57, Issue 3, April 2012.
9. N. Shrivastava, and V. Tyagi, "Content based image retrieval based on relative locations of multiple regions of interest using selective regions matching," Information Sciences: an International Journal, Vol. 259, Feb 2014.
10. C. Li and C. Hsu, "Image Retrieval With Relevance Feedback Based on Graph-Theoretic Region Correspondence Estimation,"IEEE Transactions on Multimedia, Vol. 10, No. 3, April 2008.
11. D. G. Lowe, "Distinctive Image Features from Scale-Invariant Key-points," International Journal of Computer Vision, Jan 2004.
12. https://software.intel.com/en-us/node/529070.
13. S. R. Singh, Dr. S. Kohli, "Enhanced CBIR using Color Moments, HSV Histogram, Color Auto Correlogram, and Gabor Texture," International Journal of Computer Systems, May 2015.
14. J. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm," Intel Corporation, Microprocessor Research Labs, OpenCV Documents, 1999.
15. G. Csurka, C. Dance, L. Fan, J. Willamowski and C. Bray, "Visual categorization with bags of keypoints," in Proc. Workshop Statist. Learning in Comput. Vision, ECCV, 2004, Vol. 1, p. 22.
16. G. Khosla1, Dr. N. Rajpal and J. Singh3, "Evaluation of Euclidean and Manhanttan Metrics In Content Based Image Retrieval System," International Journal of Engineering Research and Applications, Sept 2014.
17. D. John, S.T. Tharani and Sreekumar K, "Content Based Image Retreival using HSV-Color Histogram and GLCM" International Journal of Advance Research in Computer Science and Management Studies, Vol 2, Issue 1, Jan. 2014.
18. http://gedas.bizhat.com/dist.htm.
19. http://wang.ist.psu.edu/docs/related/.
20. http://savvash.blogspot.in/2008/12/benchmark-databases-for-cbir.html.
21. J. Li, 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.
22. J. Z. Wang, J. Li, G. Wiederhold, ``SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 23, no.9, pp. 947-963, 2001.
23. https://en.wikipedia.org/wiki/Precision_and_recall.

BIOGRAPHY

Teena Mariya Babu is an M.Tech Student in the Computer Science and Engineering Department, Viswajyothi College of Engineering and Technology, MG University, Kerala, India. She received her Bachelor of technology in Computer Science and Engineering from Viswajyothi College of Engineering and Technology, MG University, Kerala, India in 2013. Her research interests are Image and Video Processing, HCI, Cloud Computing etc.

Shibu K R is an Assistant Professor in the Computer Science and Engineering Department, Viswajyothi College of Engineering and Technology, MG University, Kerala, India. He received his Master of Engineering in Computer Science and Engineering from Manomaniam Sundarnar University, Tirunelveli, India in 2009. His research interests are Networking, Operating System etc.