Image Retrieval Using Local Object Appearance and Spatial Color Layout

I.Jeena Jacob, and Dr. K.G. Srinivasagan

***Abstract*—** In the current scenario, information retrieval is one of the major research areas. The retrieval process in image databases, evaluation of similarity and their mutual relationships depends on the content representation with various features. Since 1992, the multidimensional research has been carried out to meet out the aim with color, shape, texture and spatial layout as the features. In this paper, we provide a spatial and spectral layout of the images with CPAM (Colored Pattern Appearance Model) and a descriptor for characterizing local object appearance and shape using Histogram of Oriented Gradients (HOG). In particular, to get rid of the errors in images due to occlusions, facade and lighting changes, we propose to extract HOG descriptors from the images in regular intervals. Second, fusion of HOG descriptors at different scales allows capturing important structure for image retrieval. Third, the CPAM helps to fully exploit the human vision system. Analytical comparison and experimental results show that the proposed look ahead improves the state-of-the-art in state space search methods.

***Keywords*—** Content Based Image Retrieval, HOG, CPAM.

I Introduction

THE massive increase in the applications of information technology produces lot of data which should be processed in a proper way to yield the needed information. This makes a drive for the various information retrieval systems like Content Based Image Retrieval (CBIR). The retrieval results should be effective and efficient [2]. In Content Based image retrieval (CBIR), the clue for the retrieval is produced from the image contents like color, shape, texture, spatial layout etc.

The rest of the paper is organized as follows. Section 2 gives the related works of CBIR, Section 3 elucidates the existing system where the Colored Pattern Appearance Model and Histogram of Oriented Gradients are explained. Section 4 explains the proposed work. Section 5 enumerates the implementation details and experimental result of the proposed work. Section 6 gives the conclusion followed by the references.

I.Jeena Jacob, Assistant Professor is with the SCAD College of Engineering and Technology affliated to the Anna University Chennai and she is a research scolar doing her Ph.D in Information and Communication Engineering in Anna University Chennai, Tamilnadu, India e-mail id: jeni.neha@gmail.com

Dr.K.G. Srinivasagan, Professor is with National Engineering College, Kovipatti, Tamilnadu, India. He is the Professor & Head/ PG-CSE of the same college e-mail id: kgsnec@rediffmail.com

II. Related Works

The thought of information extraction using a picture was brought into the world in 1979 in a conference on Database Techniques for Pictorial Applications held in Florence [2]. The efficient and effective management of the expanding pictorial information became a critical problem. In 1992, In the workshop on "Visual Information Management Systems" organized by the National Science Foundation [15] of the United States, some researchers identified new directions in image database management. The retrieval of images depends on the information content they are having. Color features include the conventional color histogram (CCH) [31], the fuzzy color histogram (FCH) [14], the color correlogram (CC) [32] [33] and a more recent color-shape-based feature [16]. Texture measures [4] [5] [7] look for the patterns in images and how they are defined in the images [34]. The specific textures in an image can be identified by making a model for the texture as a two dimensional gray level variation. The shape feature [3] [8] [9] gives the directional orientation of edges for the objects in the image. Often the shapes are extracted by first applying some segmentation algorithms to an image. A large number of retrieval systems have been developed by various organisations, individuals and hospitals like SIMPLIcity [1], QBIC [11] [13], Netra [12], PicToSeek [10], PhotoBook [13] etc.

A. Colored Pattern Appearance Model (CPAM)

The visual object patterns of the images can be represented by different kinds of models like RGB, HSV, YCbCr etc. According to the opponent color theory [18], the human color vision system has three visual streams.

$$\mathbf{vectorC} = \sum_{row}\sum_{col} \mathbf{cb(row, col)} \Big/ \mathbf{4 \times localmean} + \sum_{row}\sum_{col} \mathbf{cr(row, col)} \Big/ \mathbf{4 \times localmean} \quad (1)$$

$$\mathbf{minDistP} = \mathbf{min}\left(\sum_{j=1}^{vectorSizeP} \sum_{i=1}^{vecorP} \left(\mathbf{vectorP(j) - codeBookP(i,j)}\right)^2 \right) \quad (2)$$

$$\mathbf{minDistC} = \mathbf{min}\left(\sum_{j=1}^{vectorSizeC} \sum_{i=1}^{vecorC} \left(\mathbf{vectorC(j) - codeBookC(i,j)}\right)^2 \right) \quad (3)$$

The pattern color separable model of human color vision [19] [20] [22] [23] model suggests that the value of one neural image is the product of three terms like the stimulus color

direction, the spatial patterns of the stimulus and to the stimulus strength.

In CPAM [17] [21] [24], based on the spatial and spectral characteristics of a group of neighboring pixels in a color image a colored image pattern is defined. The visual appearance of a small image block is modelled by three components: the stimulus strength (S), the spatial pattern (P) and the color pattern (C). For a small image area, the stimulus strength S is approximated by the local mean of the Y component as in eqn 1.The pixels in Y is normalized by S form the spatial pattern by the eqn 2 and eqn 3. Because Cb and Cr have lower bandwidth, they are sub-sampled by a factor of 2 in both dimensions. The sub-sampled pixels of Cb and Cr are normalized by S, to form the color pattern (C) component of the appearance model. Normalizing the pattern (eqn. 4) and color channels (eqn. 5) by the strength has the effect of removing (to a certain extent) the effects of lighting conditions, making the visual appearance model somewhat color constant.

$$\mathbf{ASPH} = \frac{\mathbf{histP}}{\sum \mathbf{histP}} \tag{4}$$

$$\mathbf{CSPH} = \frac{\mathbf{histC}}{\sum \mathbf{histC}} \tag{5}$$

By separating achromatic and chromatic signals, the work can be done on two low-dimensional vectors rather than one very high dimensional vector.

B. Histogram of Oriented Gradients (HOG)

The local object appearance and shape of an image can be represented by the proper distribution of local intensity gradients. This can be implemented by dividing the image into small cells, for each cell a local 1-D histogram of gradient directions over the pixels of the cell is calculated [28] and it is normalized. This normalized descriptor blocks can be termed as Histogram of Oriented Gradient (HOG) descriptors [27] - [29]. The variation of the spatial position to scaling and rotation can be achieved by extracting descriptors from only salient points in the scale space of the image following rotation normalization. The steps involved are: (1) Scale-space extrema detection, (2) Orientation assignment and (3) Descriptor extraction.

Fig. 1 shows an example patch with their corresponding gradient directions. In [27] [29], the authors were successful in applying the HOG for face recognition.

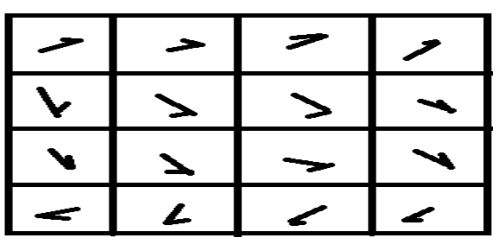

Fig. 1 An example patch with theircorresponding gradient directions

The proposed work tries to use the HOG to the images of varied image databases, since this descriptor characterizes local object appearance and shape of the images well.

III. PROPOSED WORK

Image retrieval using local object appearance and spatial color layout is done by the combined approach of Colored Pattern Appearance Model (CPAM) and Histogram of Oriented Gradients (HOG) in the proposed work. Fig 2 shows the algorithm for Colored Pattern Appearance Model and fig. 3 shows the algorithm for Histogram of Oriented Gradients. Fig. 4 gives the schematic diagram of CPAM. Although many variations of color space exist, the YCbCr space has been chosen as the color space [24].

Algorithm: Colored Pattern Appearance Model

Inputs: Image, width of the image,
height of image and blocksize.

1. Initialize vectorSize as blocksize X blocksize.
2. Initialize a three dimension matrix named temp to store the data to be used
3. Store the R,G and B component details in the third dimension of the temp
4. Set up transformation marix T
5. Transform to YCbCr color model.
6. Find the local mean of all YCbCr components
7. Calculate vectorC by eqn (1)
8. Find vectorP by normalising YCbCr with localmean
9. minDistP and minDistC are calculated by eqn (2) and eqn (3)
10. Histogram values for the corresponding pattern and color code, histP and histC can be calcuated by eqn (4) and eqn (5)

Fig. 2: Algorithm for Colored Pattern Appearance Model

A. Normalization of Color

The evaluations on several input pixel representations including grayscale, RGB and LAB color spaces optionally with power law (gamma) equalization. These normalizations have only a modest effect on performance, perhaps because the subsequent descriptor normalization achieves similar results. We do use color information when available.

B. Gradient Computation

The testing on the gradients are computed using Gaussian smoothing with several discrete derivative masks. Several smoothing scales were tested including none. Simple 1-D [-1, 0, 1] masks work best. While the larger masks are used it decreases the performance and if the smoothing is done in images the performance decreases significantly.

C. Descriptor Block Formation

The strengths of gradient vectors can be varied for various local variations and contrast change. So the local contrast normalization is very essential to get a proper descriptor block.

Algorithm : Hisogram of Oriented Gradients

Inputs: Image, HOG window sizes $nwin_x$ and $nwin_y$ per bound box and histogram bins, B.
L: num of lines ; C: num of columns

1. Find Stepx and Stepy as

$$\mathbf{Stepx} = \left\lfloor \frac{\mathbf{C}}{\mathbf{nwinx} + 1} \right\rfloor$$

$$\mathbf{Stepy} = \left\lfloor \frac{\mathbf{C}}{\mathbf{nwiny} + 1} \right\rfloor$$

2. For each block
   a. Find the change in pixel intensities in X and Y directions, Gx and Gy
   b. Find the magnitude and angles of the same by

$\theta = \tan^{-1} \frac{Gy}{}$

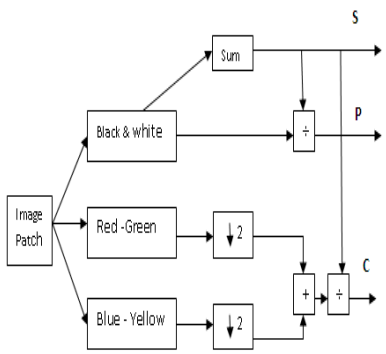

Fig. 4 Colored Pattern Appearance Model

IV. Experimental Results

We conducted an extensive empirical study to 1) evaluate the existing approaches to answer queries and compare it to the multiple expansion approach, and 2) evaluate the proposed techniques, with various variables and conditions. Precision and recall are the two commonly used performance analysis variables. The Fig 5 gives the Average Precision and Average Recall values of various images in the dataset which varies for different window sizes. For almost all type of image groups the algorithm works fairly well. The dataset groups Horses, Flowers, Dinosaurs, Buses and Elephants in the dataset gets 100% precision and recall. But the precision and recall of Food is very poor. The minute variations in the intensity variations make poor result for Food images.

The results of proposed method can be compared with that of HOG and CPAM. Fig 6 shows the comparison of Average Precision and Average Recall for various number of retrieval results. The Average Precision and Average Recall gives better result for top 3 retrievals which gets lowered as the count of retrieved results increases. While making the comparison of Average Precision and Average Recall for various window sizes, the retrieval result of proposed method gives better performance than the other two (i.e. HOG and CPAM) at least by 8%.

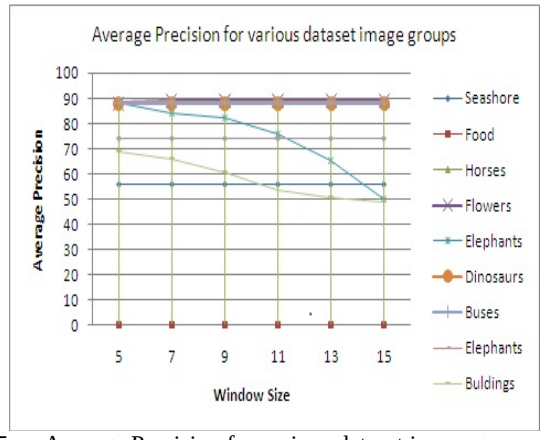

Fig 5 a: Average Precision for various dataset image groups with various window sizes.

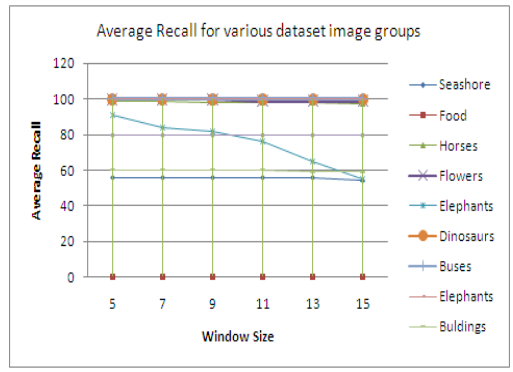

Fig 5 b: Average Recall for various dataset image groups with various window sizes

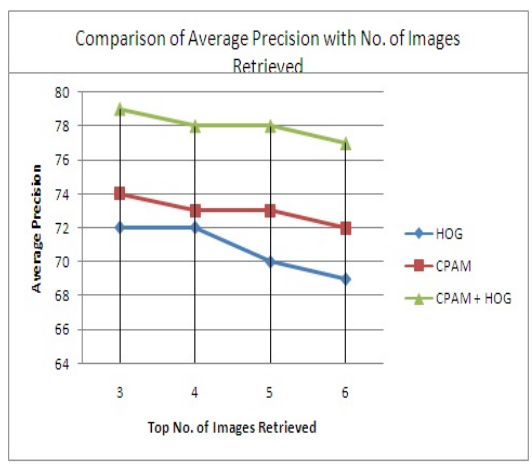

Fig 6 a: Average Precision for images with the varying window sizes.

The evaluation of the system's performance can also be done with a single combined metric derived from the Precision and Recall. Here we used the F-score(Fβ), defined as follows:

$$\mathbf{F\beta} = (1 + \beta^2)\frac{\mathbf{precision.recall}}{\beta^2\mathbf{precision} + \mathbf{recall}} \quad (6)$$

β defines weight should be given to the variables recall and precision. F-score should be a number between 0 and 1, with 1 representing a perfect retrieval system that is completely robust and completely discriminant (100% precision and 100% recall).

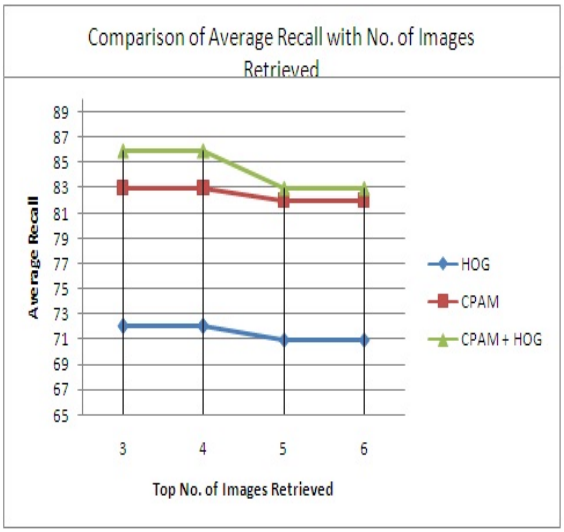

Fig 6 b: Average Recall for images with the varying window sizes

We use β=0.5, which gives twice as much importance to precision as to recall. Fig 7a shows the comparison of F-score for images with the number of images retrieved and Fig 7b shows the comparison of F-score for images with the varying window sizes. The comparison of F-score for various methods given in Fig 7 gives better result for proposed work at least 0.78%.

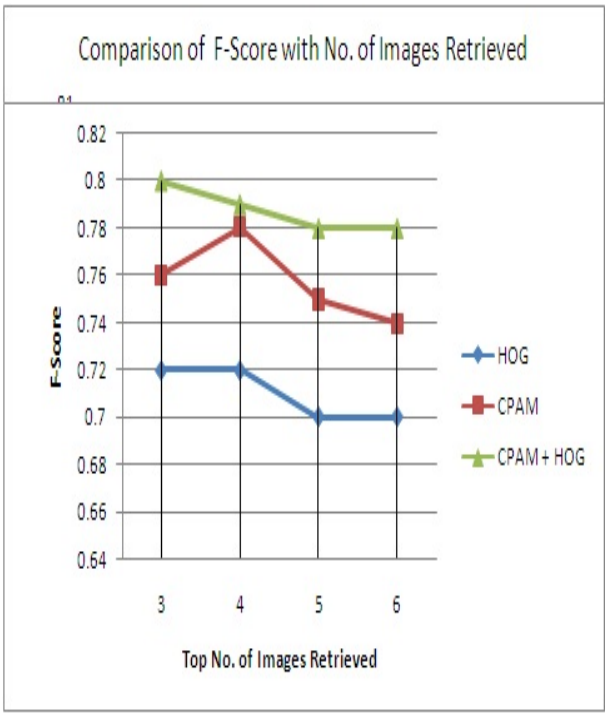

Fig 7 a: F-score for images with the no. of images retrieved.

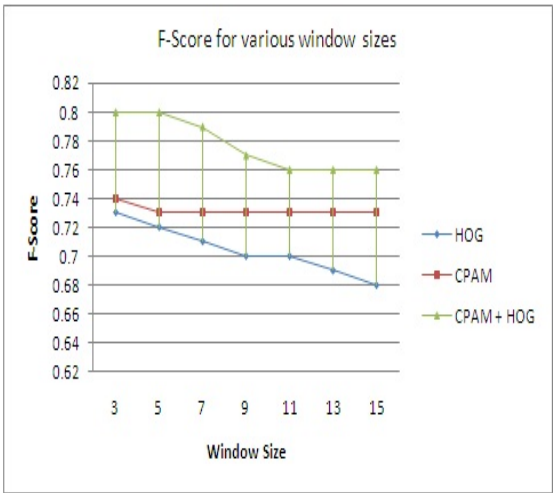

Fig 7 b: F-score for images with the varying window sizes.

V. Conclusion

The modeling of image content based on CPAM and the local spatial and spectral descriptor HOG fits the need for efficient retrieval based on the joint similarity between individual entities and their mutual relationships. We use databases of more than one thousand images which vary in its characteristics. The color model used helps to fully exploit the human vision system. Analytical comparison and experimental results with the view of gamma/color normalization, gradient computation, spatial / orientation binning, normalization and descriptor blocks show that the proposed work provides improvement and also using the algorithms in a regular grid helped us to get the better results which compensate errors in images due to occlusions, pose and illumination changes.

References

[1] James Z.Wang, Jia Li, Gio Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture LIbraries," IEEE Trans. PAMI, Vol.23, No.9, Sep 2001.

[2] A.Smeulders, M. Worring, S.Santini, A.Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.22, No.12, pp. 1349-1380, Dec. 2000.

[3] S.Belongie, J.Malik, and J.Puzicha, "Shape Matching and Object Recognition using Shape Contexts," IEEE Trans. Pattern Anlysis Machine Intelligence, Vol.24,No.4, pp.509-522, Apr.2002.

[4] B.S.Manjunath and W.Y. Ma,"Texture Features for browsing and Retrieval of Image Data," IEEE Trans. On Pattern analysis and Machine Intelligence, Vol.18, pp.837-842, Auust 1996.

[5] B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, and A. Yamada," Color and Texture descriptors," IEEE Systems for Video Technology, Vol.11, pp.703-715, June 2001.

[6] Mani Malek Esmaeili, Mehrdad Fatourechi and Rabab Kreidieh Ward, "A Robust and Fast Video Copy Detection System Using Content-Based Fingerprinting," IEEE Trans. On Information Forensics and Security, Vol.6, No.1, pp. 213-225, March 2011. (f score)

[7] Noureddine Abbadeni, "Computational Perceptual Features for Texture Representation and Retrieval," IEEE Trans. On Image Processing, Vol.20, No.1, pp.236-246, January 2011.

[8] Grant J.Scott, Mathew M.Klaric,Curt H.Davis and Chi-Ren Shyu, "Entropy-Balanced Bitmap Tree for Shape-Based Object Retrieval From Large-Scale Satellite Imagery Databases," IEEE Trans. On

Geoscience and Remote Sensing, Vol.49, No.5,pp.1603-1616, May 2011.

[9] Yi-Feng Pan, Xinwen Hou and Cheng-Lin Liu,"A Hybrid Approach to Detect and Localize Texts in Natural Scene Images," IEEE Trans. On Image Processing, Vol.20, No.3, pp.800-813, March 2011.

[10] T. Gevers and A. Smeulders, "PicToSeek: A Content Based Image Search System for the World Wide Web,"Proc. Visual 97, Knowledge Systems Institute, Chicago, 1997, pp. 93-100.

[11] Bjorn Johansson, "QBIC (Query By Image Content)", November 2002, [Online Document].

[12] B. S. Manjunath, "NeTra: A toolbox for navigating large image databases,' Multimedia Systems 7, pp.184–198 ,1999.

[13] Pentland A, Picard R, Sclaroff S: Photobook – tools for content-based manipulation of image data-bases. Proceedings SPIE 2185:34–47, 1994.

[14]. J. Han and K. Ma, "Fuzzy Color Histogram and Its Use in Color Image Retrieval", IEEE Trans. On Image Processing, vol. 11, pp. 944 – 952, Aug. 2002.

[15]. Hirata K. and Kato T. "Query by visual example – content-based image retrieval", In Proc. Of Third International Conference on Extending Database Technology, EDBT"92, 1992, pp 56-71.

[16] Lining Zhang, Lipo Wang and Weisi Lin,"Semisupervised Biased Maximum Margin Analysis for Interactive Image Retrieval," IEEE Trans. on Image Processing, Vol.21, No.4, pp.2294-2306, April 2012.

[17] Ning Zhou, William K.Cheung, guoping Qiu and Xiangyang Xue, "A Hybrid Probabilistic Model for Unified Collaborative and Content-Based Image Tagging," IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol.33, No.7, pp.1281-1294,July 2011.

[18] Poirson and B. Wandell, "Appearance of colored patterns: pattern-color separability", J. Opt. Soc. Am. A, vol. 10, pp. 2458 – 2470, 1993.

[19] P. K. Kaiser and R. M. Boynton, Human Color Vision, Optical Society of America, W ashington DC, 1996.

[20] D. Marr, Vision: A Computational Investigation into Human Representation and Processing of Visual Information. Freeman and Co., New York, San Francisco, 1982

[21] L. Harvey, Jr. and M. Gervais, "Internal representation of visual texture as the basis for judgment of similarity", Journal of experimental Psychology: Human Perception and Performance, vol. 7, pp. 741 – 753, 1981

[22] Blakemore and F. W. Campbell, "On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images", Journal of Psychology, vol. 204, pp. 237 – 260, 1969

[23] R. Rao and D. Ballard, "An active vision architecture based on iconic representations", Artificial Intelligence, vol. 78, pp. 461-505, 1995

[24] G. Qiu, "Image coding using a Colored Pattern Appearance Model", Visual Communication and Image Processing 2001, January 2001, San Jose, CA, USA

[25] M. Lewicki and B. Olshausen, "Probabilistic framework for the adaptation and comparison of image codes", J.Opt. Soc. Am. A, vol. 16, pp. 1587 – 1601, 1999

[26] Schiele and J. L. Crowley, "Recognition without correspondence using multiresolution receptive field histogram", International Journal of Computer Vision, 36 (1), 31 -50, 2000

[27] Navneet Dalal and Bill Triggs, "Histograms of Oriented Gradients for Human Detection,"

[28] Jonathan Brookshire," Person Following using Histograms of Oriented Gradients",

[29] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using Histograms of Oriented Gradients," Pattern Recognition Letters 32, pp. 1598–1603,2011.

[30] Jae Young Choi, Yong Man Ro, IEEE, and Konstantinos N. Plataniotis," Color Local Texture Features for Color Face Recognition," IEEE Trans. On Image Processing, Vol.21, No.3, pp. 1366-1380,March 2012.

[31] J. R. Smith and S.-F. Chang, " Automated image retrieval using color and texture", Columbia University, Tech. Rep CU/CTR 408-95-14, July 1995.

[32]. J. Huang, S. R. Kumar, M. Mitra and W. J. Zhu, R. Zabih, "Image Indexing Using Color Correlograms", Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 762 – 768, June 1997.

[33] N. R. Howe and D. P. Huttenlocher, "Integrating Color, Texture and Geometry for Image Retrieval", Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. II, pp. 239-246, June 2000.

[34] Dr.H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, Vaishali Suryavanshi,"Improved Texture Feature Based Image Retrieval using Kekre"s Fast Codebook Generation Algorithm", Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010.