ISSN:0975-9646

Mohd. Aasif Ansari et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015, 1740-1745

Enhanced Video Retrieval and Classification of Video Database Using Multiple Frames Based on Texture Information

Mohd. Aasif Ansari[1], Dr. Hemlata Vasishtha[2]

[1]Ph.D. Scholar, [2]Professor, Faculty of Engineering & Technology,
Shri Venkateshwara University, Gajraula, India

Abstract— **Content Based Video Retrieval (CBVR) has been increasingly used to describe the process of retrieving desired videos from a large collection on the basis of features that are extracted from the videos. The extracted features are used to index, classify and retrieve desired and relevant videos while filtering out undesired ones. Videos can be represented by their audio, texts, faces and objects in their frames. An individual video possesses unique motion features, color histograms, motion histograms, text features, audio features, features extracted from faces and objects existing in its frames. Videos containing useful information and occupying significant space in the databases are under-utilized unless there exist CBVR systems capable of retrieving desired videos by sharply selecting relevant while filtering out undesired videos. Results have shown performance improvement when features suitable to particular types of videos are utilized wisely. Various combinations of these features can also be used to achieve desired performance. Many researchers have an opinion that result is poor when images are used as a query for video retrieval. Here, instead of using a single image or key frames, multiple frames of the video clip being searched are used. Technique used for CBVR system shown in this paper yields significantly acceptable and higher retrieval results. The system is implemented using MATLAB. Performance of the system is assessed using a database containing 1000 video clips of 20 different categories with each category having 50 clips. The performance is shown using two different types of texture features- features extracted using Gabor filter and Kekre's Fast Codebook Generation Algorithm (KFCG).**

Keywords- **CBVR, Multiple Frames, Gabor, KFCG, MATLAB**

I. INTRODUCTION

With lack of satisfaction from textual based video retrieval, the idea of content based video retrieval has been the attention for researchers since long time. In the beginning of content based video retrieval, they tried to retrieve videos using an image. However, video retrieval using query by image is not successful as it cannot represent a video. A video is a sequence of images and audio. A query video provides rich content information than that provided by a query image. Finding the relevant video by sequentially comparing the low level visual features of key frames of the query video with those of key frames of videos in database provide long pending solution to yield better result [6] of video retrieval. Finding similarity measure requires key frames matching and hence computing key frame features including color histogram, texture and edge features, etc., to calculate distance parameter. These huge computations cause long response time to the users and thus, the problem of high computation cost in computing visual features of videos is persistent. Apart from this, considerations for motion features, temporal, sequence and duration of shots in a video pose a challenge for the research area [5]. The structural and content attributes obtained through content analysis, segmentation, video parsing, abstraction processes and the attributes entered manually are referred to as metadata. Video is indexed on a table using the metadata using clustering process which categorizes video clips or shots. Clustering process categorizes video clips or shots using metadata to form an index table of videos into different visual categories. Researchers have developed various tools and schemes to index, enquire, browse, search and retrieve videos from large databases but effective and robust tools are still lacking to test with large databases [6]. Due to these limitations [5], [6] a majority of video searches and retrievals still relies on keyword or text attributions. Face detection is assessed for image and video analysis. It was experimented in a commercial system [15]. It was found that accuracy of face recognition in video collection of the type mentioned in the system [8] was too poor to prove to be useful. Overall a large number of queries do not yield satisfactory results as mentioned [8] about one third of the queries were unanswerable by any of the automatic systems participating in the video retrieval track [16]. No system or method was able to provide relevant results. An integrated video retrieval system is proposed [2] where a video shot is represented not by key frame only but by all frames to extract more visual features of a shot. Color and motion features are integrated to fully exploit the spatio-temporal information contained in a video [29]. To overcome these drawbacks, i.e. considering lower efficiency of CBVR systems using a single image and very high computational cost of CBVR systems using key frames and the problem of availability of effective tools for CBVR systems using clustering process and to strike a balance between the efficiency and computational cost, visual features from multiple frames of a video clip are used in the system proposed here instead of a single frame or key frames or all frames of a clip. Also, it is learnt from the evaluation of video information retrieval that good image retrieval leads to good performance of video retrieval system when query is an image or an image from the query video [8]. Computational cost point of view, the system proposed in this paper is cost effective along with acceptable as well as significantly higher results.

In section 2 features and features extraction algorithms are discussed; section 3 discuses about similarity measure; section 4 shows the methodology to calculate result parameters in the proposed CBVR system. Proposed CBVR system is elaborated in section 5 and the result charts are shown in section 6; problems and challenges posed to this CBVR system are discussed in section 7 and the conclusion is presented in section 8.

II. FEATURES AND FEATURES EXTRACTION

A. *Extraction of Gabor Features*

For effective video indexing, classification and retrieval visual features embedded in video data is exploited. Three primary features to be extracted are color, texture and motion for effective video indexing. These features are represented by color histogram, Gabor texture features and motion histogram respectively [4]. Edge histogram and texture features are one of the most reliable data for effective video retrieval application. Gabor filters can also be used to obtain textural properties of texts which are distinct and distinguish them from its background in the image [7]. Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and a specific direction. Expanding a signal using this basis provides a localized frequency description, therefore capturing local features/energy of the signal. Texture features can then be extracted from this group of energy distributions. The scale (frequency) and orientation tuneable property of Gabor filter makes it especially useful for texture analysis. The filters of a Gabor filter bank are designed to detect different frequencies and orientations. They can be used to extract features on key points detected by interest operators [17]. From each filtered image, Gabor features can be calculated and used to retrieve images. The algorithm for extracting the Gabor feature vector is shown in fig. 1 and the related equations (1 - 4) are also shown below [18], [20].

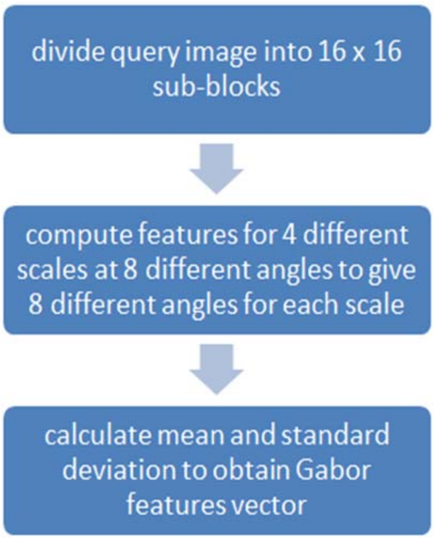

Fig. 1 Gabor Filter Algorithm

For a given image I(x,y), the discrete Gabor wavelet transform is given by a convolution:

$$W_{mn} = \sum_{x_1}\sum_{y_1} I(x_1, y_1)\, g_{mn} * (x - x_1, y - y_1) \quad -(1)$$

where * indicates complex conjugate and m, n specify the scale and orientations of wavelet respectively. After applying Gabor filters on the image with different orientation at different scale, an array of magnitudes is obtained:

$$E(m,n) = \sum_{x}\sum_{y} |W_{mn}(x,y)| \quad - \;(2)$$

These magnitudes represent the energy content at different scale and orientation of the image. The main purpose of texture-based retrieval is to find images or regions with similar texture.

The standard deviation σ of the magnitude of the transformed coefficients is:

$$\sigma_{mn} = \sqrt{\frac{\sum_x \sum_y (|W_{mn}(x,y)| - \mu_{mn})^2}{P\ X\ Q}} \quad - \;(3)$$

Where μ is the mean of magnitude and given as

$$\mu_{mn} = \frac{E(m,n)}{P\ X\ Q}$$

A feature vector f (texture representation) is created using mn as the feature components [19], [14]. M scales and N orientations are used and the feature vector is given in equation (4)

$$f = [\sigma_{00}, \sigma_{01}, \sigma_{02} \dots \sigma_{(M-1)(N-1)}] \;-(4)$$

$f_{Gabor} = \frac{f-\mu}{\sigma}$ where μ is the mean and σ is the standard deviation of *f*.

B. *Extraction of KFCG Features*

In vector quantization, the image is divided into smaller non-overlapping blocks of n x n (= k) pixels. k pixels of each such block are arranged to form a vector. These vectors are represented by the closest codewords and the set of all these codewords is known as codebook. Number of bits required to represent each codeword is much smaller than to represent image vectors [28]. Hence, compression is achieved. There are many methods to generate codebook where the Linde-Buzo-Gray (LBG) algorithm, is most commonly used [27]. In this method, block vectors are known as training vectors. The training vectors are divided into different clusters formed by process of iteration. Centroids of these clusters are the codebook vectors [23]. The corresponding codebook vectors represent all the training vectors in that cluster. The codebook vector closest to a training vector [21] is selected to represent it as a codebook vector. A set of codewords represents codebook vectors through a codebook which is used to encode as well as decode the images [22].

Kekre's Fast Codebook Generation Algorithm is used for image compression [25] [26]. This algorithm takes lesser time to make codebook for vector quantization process [24]. The image is divided into small blocks of 2 x 2 pixels each. Each of these pixels contains red, green and blue components using one byte for each color component so that the size of each block is twelve bytes. Content of these blocks is arranged to form a vector, known as a training vector. The image is now represented by a set of such training vectors. A complete set containing all training vectors forms one cluster C1 with its centroid as a code vector V1 as shown in fig. 2. During first iteration, the first cluster C1 is divided into two smaller clusters, C11 and C12. If the first element of a training vector of cluster C1 is less than the first element of code vector V1, it will belong to cluster C11. If not, it will belong to C12 as shown in fig. 3.

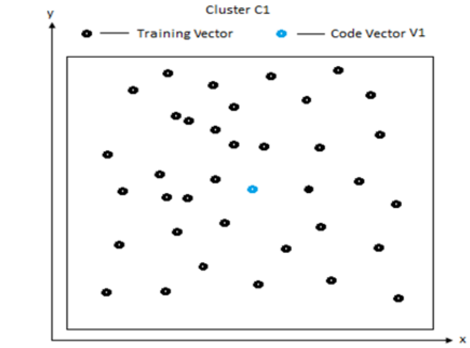

Fig. 2 Cluster C1

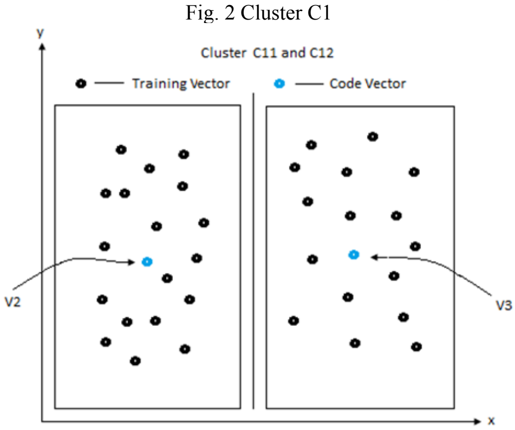

Fig. 3 Cluster C11 and C12

Then, centroid of cluster C11 is obtained, called as Code Vector V2. Similarly, centroid of cluster C12 is obtained, called as Code Vector V3. Using V2 and V3, the clusters C11 and C12 are again divided into two clusters each to form four clusters totally, as shown in fig. 4.

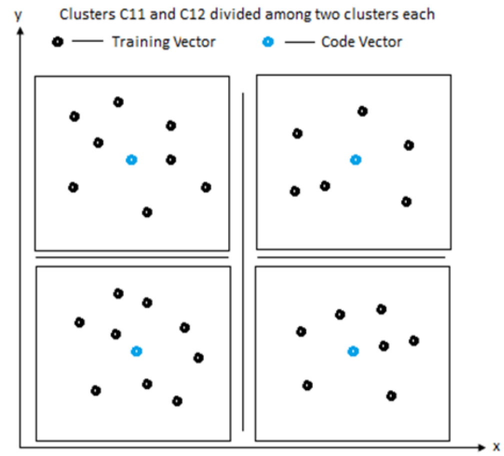

Fig. 4 Clusters divided to form four clusters

Through multiple iterations, these clusters are again divided to form smaller clusters with their own centroids as their corresponding code vectors by comparing second and third elements of training vectors belonging to these clusters with the corresponding elements of code vectors of corresponding clusters. Iteration process is continued depending upon size of the codebook to be created. In this paper, we generate codebook for each frame containing 16 code vectors having 12 elements each.

III. Similarity Measure

Queries are classified by categories sorted out according to type of features used or type of example data. The query is found out by calculating similarity between feature vector [9], [10] stored in the database and the query features. The similarity is obtained using the frames separated out from the enquired clip with those of the videos stored in the database. Video retrieval result depends greatly on video similarity measures. The videos are retrieved by measuring similarity between the features extracted from multiple frames associated with query video and videos from the database. The similarity is obtained by matching these features. Measuring similarity by matching features is most convenient and direct method [1]. It is measured by the average distance between features of corresponding frames [12]. In query by example frames like the one used in the system shown in this paper similarity measure to find relevant videos usually low level feature matching is used. Video similarity can be measured at different levels of resolution or granularity [13]. A video clip is retrieved by finding similarity between the query video and videos of the database by using the Euclidean distance between them. Euclidean distance is calculated to rank the retrieved videos. The video from the database corresponding to the frame similar to the query frame is higher in rank if the Euclidean distance is smaller [3], [7]. The equation for Euclidean distance between a query frame q and a database frame d is shown in equation (5)

$$\text{Euclidean Distance} = \sqrt{\sum_{n=1}^{N}(V_{dn} - V_{qn}).(V_{dn} - V_{qn})} \quad - \ (5)$$

Where V_{dn} are the feature vectors of database frame d and V_{qn} are the feature vectors of query frame q each having size N.

IV. Result Evaluation Method

The performance of video retrieval is evaluated with the same parameters as it is evaluated in image retrieval [11]. Recall and precision are the two parameters [2] as given in equations (6) and (7).

$$Recall = \frac{DC}{DB} \quad - (6)$$

$$Precision = \frac{DC}{DT} \quad - (7)$$

$DC = number\ of\ similar\ clips\ detected\ correctly$
$DB = \ number\ of\ similar\ clips\ in\ the\ database$
$DT = total\ number\ of\ detected\ clips$

Crossover points are calculated using the above mentioned two parameters to find the performance of the proposed system.

V. Proposed CBVR System

A CBVR system is proposed in this paper in which multiple frames are obtained for the query clip and the videos' database instead of using single frame or key frames or all frames [2]. Features are extracted from these frames. The similar and most relevant videos are obtained based on similarity measure using minimum Euclidean Distance. Significantly acceptable results have been obtained using this system. A typical methodology is used in this system where a video is retrieved based on a query clip. Here, database is processed offline. The video clip is represented by features extracted from multiple frames from the videos. These features from the query clip are compared with features in the database. Video sequences are ranked according to the distance measures and similar videos are retrieved as shown in fig. 5. As mentioned above, multiple frames based retrieval yields acceptable results without the complexity of finding key frames to represent a shot.

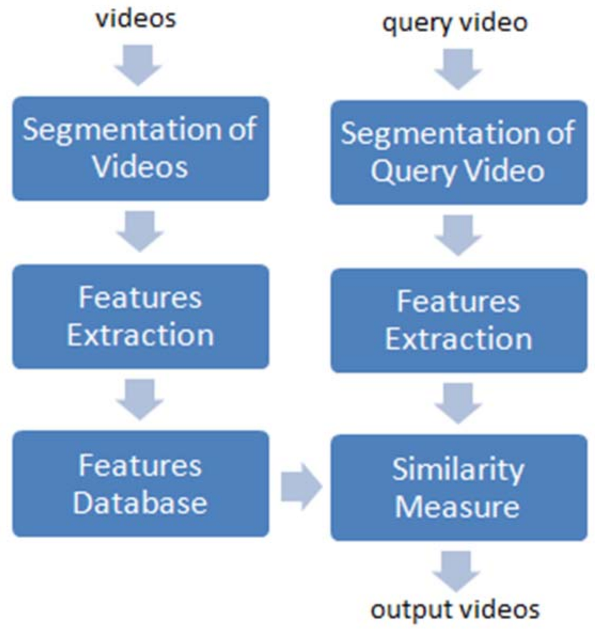

Fig. 5 Proposed CBVR system

A process flow of the proposed CBVR system is shown in fig. 5. Multiple frames are obtained during segmentation. Features are then extracted for each frame and stored in features database. Features of the corresponding frames from the query video are also extracted and then compared with features stored in the database. The output video is obtained by finding the similarity measure between features of query video and the features stored in the database.

VI. Results

A. Database

The technique using multiple frames with Gabor Magnitude and KFCG Algorithm are applied to a video database having 1000 videos with 20 categories of 50 videos each as shown in fig. 6. The similarity measure between the query clip and videos in database is found out using Euclidean Distance. The precision and recall values are computed by grouping the number of retrieved videos sorted according to the minimum Euclidean Distance.

Fig. 6 Video database of 1000 videos with 20 categories

B. Results

The graphs shown below in fig. 7 and fig. 8 represent the retrieval results obtained for retrieving video clips of "Mother Teresa" while the graphs of fig. 9 and fig. 10 correspond to video clips of "classical dance". These two categories are among the 20 categories of video clips from the video database of 1000 videos. The results obtained are much appreciable for all the categories but two of them are demonstrated here. The results are obtained using two types of features, one extracted using Gabor wavelet transform and the other using KFCG algorithm.

1) Results for the video clip of "Mother Teresa"

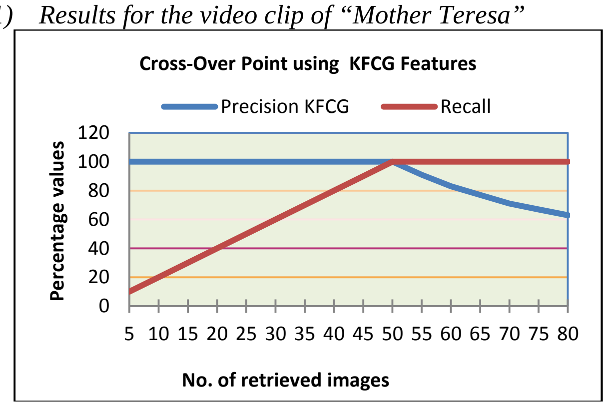

Fig. 7 Precision & Recall plotted for KFCG algorithm based CBVR technique using multiple frames

Fig. 7 shows results obtained by CBVR system based on KFCG algorithm using multiple frames. There is significant improvement in results using this technique, the best results are obtained. The above precision-recall curve shows crossover point of 100%.

Fig. 8 below shows results obtained by CBVR system based on Gabor features using multiple frames. There is significant improvement in results using this technique too. The results obtained are quite satisfactory. The above precision-recall curve shows crossover point of 98%.

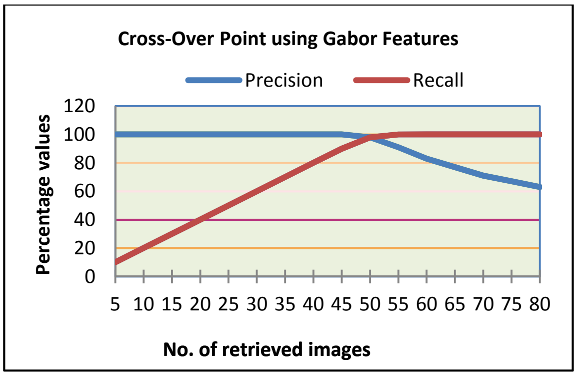

Fig. 8 Precision& Recall plotted for Gabor featuresbased CBVR technique using multiple frames

2) Results for the video clip of "Classic Dance"

Fig. 9 shows results obtained by CBVR system based on KFCG algorithm using multiple frames. The precision-recall curve shown below has crossover point of 82%. Also, the nature of the graph shows that the precision and recall values are high and indicate that this technique is an efficient technique showing acceptable results.

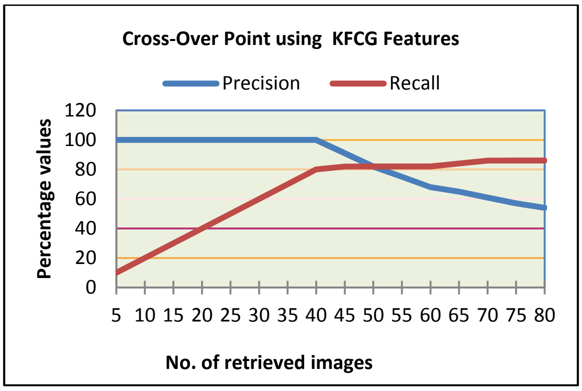

Fig. 9 Precision & Recall plotted for KFCG algorithm based CBVR technique using multiple frames

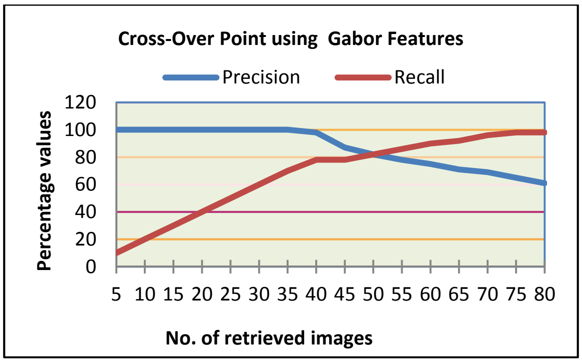

Fig. 10 Precision & Recall plotted for Gabor features based CBVR technique using multiple frames

Fig. 10 shows results obtained by CBVR system based on Gabor features using multiple frames. Here also, the results are appreciable. The above precision-recall curve too shows crossover point of 82%.

VII. Problems and Challenges

The video content is represented by spatial and temporal characteristics of videos. In spatial domain, features are obtained from frames to form feature vectors from different parts of the frames. In temporal domain, video is segmented into its elements like frames, shots, scenes and video clips and features like histograms, moments, textures and motion vectors represent the information content of these video segments [7]. Drawback of techniques employing key frames matching is that temporal information and the related information between the key frames in a shot is lost. Content based video retrieval systems using query by image or query by clips using images or frames is implemented with low level features present in these images. Because of this, different objects present against similar backgrounds in frames belonging to different videos can yield confusing or false retrievals. Also, the low level features of the frames belonging to different videos can also yield false retrievals due to their corresponding low level features matching.

VIII. Conclusion

It can be concluded from discussion in the previous sections that encouraging results are obtained and comparatively higher efficiency is achieved by using multiple frames instead of single frame or key frames representing a shot. Also, computational cost is lower for the system proposed here than that when using key frames to represent shots of a video. Query by example image is popular for content based image retrieval. Low level features are used for retrieval. The retrieval performance and the usefulness of these systems is restricted to the queries having distinct low level visual features but they do not address to the problems of video retrievals using semantic information for the query. Also, an efficient solution is needed to address the problems for the queries having similar backgrounds and showing confusing results. Automatic retrieval systems should be the focus and it requires more attention from researchers for improved retrieval results. A trend to reduce computational cost is needed to project commercialized systems for video indexing, classification and retrieval to facilitate the availability of low cost, fast and efficient CBVR systems. Capability of these systems can be magnified by reaching huge video databases that exist and are accessible on the web. The accessible databases should empower the users with options to accurately select the desired videos only while filtering out the relevant but undesired as well as irrelevant videos so that valuable, moral, ethical and informative data becomes accessible efficiently, quickly and at low cost.

References

[1] Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng, Maybank S., "A Survey on Visual Content-Based Video Indexing and Retrieval", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41-6,797-819, 11/2011

[2] Liang-Hua Chen, Kuo-Hao Chin, Hong-Yuan Liao, "An integrated approach to video retrieval", Proceedings of the nineteenth conference on Australasian database-Volume 75, 49–55, 2008

[3] B V Patel, B B Meshram, "Content Based Video Retrieval Systems", International Journal of UbiComp, vol 3, No. 2, pg 13-30, 2012

[4] Yining Deng, B.S. Manjunath, "Content-based Search of Video Using Color, Texture, and Motion", IEEE, pg 534-537, 1997

[5] Ja-Hwung Su, Yu-Ting Huang, Hsin-Ho Yeh, Vincent S. Tseng , "Expert Systems with Applications", 37, pg 5068-5085, 2010

[6] Nicu Sebe, Michael S. Lew, Arnold W.M. Smeulders, "Video retrieval and summarization", Computer Vision and Image Understanding, vol. 92, no. 2-3, pg 141-146, 2003

[7] C. V. Jawahar, Balakrishna Chennupati, Balamanohar Paluri, Nataraj Jammalamadaka, "Video Retrieval Based on Textual Queries", Proceedings of the Thirteenth InternationalConference on Advanced Computing and Communications, Coimbatore, Citeseer, 2005

[8] Alexander G. Hauptmann, Rong Jin, and Tobun D. Ng, "Video Retrieval using Speech and Image Information", Electronic Imaging Conference (EI'03), Storage Retrieval forMultimedia Databases, Santa Clara, CA, January 20-24, 2003.

[9] Swain M.J. and Ballard, B.H. “Color Indexing,” Int’l J. Computer Vision, vol. 7, no. 1, pp. 11-32, 1991.

[10] Hafner, J. Sawhney, H.S. Equitz, W. Flickner, M. and Niblack, W. “Efficient Color Histogram Indexing for Quadratic Form Distance,” IEEE Trans. Pattern Analysis and Machine Intelligence, 17(7), pp. 729-736, July, 1995.

[11] Aljahdali, S., Ansari, A., Hundewale, N., "Classification of Image Database Using SVM With Gabor Magnitude", International Conference on Multimedia Computing and Systems (ICMCS), 2012 , vol., no., pp.126,132, 10-12 May 2012

[12] P. Browne and A. F. Smeaton, “Video Retrieval Using Dialogue, Keyframe Similarity and Video Objects”, IEEE Int. Conf. Image Process., vol. 3, pp. 1208–1211,Sep. 2005.

[13] R. Lienhart, “A System For Effortless Content Annotation To Unfold The Semantics In Videos,” in Proc. IEEE Workshop Content-Based Access Image Video Libraries, pp. 45–49, Jun. 2000.

[14] C. Faloutsos, R. Barber, M. Flicker, J. Hafner, W. Niblack, D. Petkovic and W. Equitz, "Efficient and effective querying by image content", J. lntelL Inf. Systems 3, 231-262, 1994.

[15] Visionics Corporate Web Site, FaceIt Developer Kit Software, http://www.visionics.com, 2002.

[16] The TREC Video Retrieval Track Home Page, http://www-nlpir.nist.gov/projects/trecvid/

[17] Arti Khaparde, B. L. Deekshatulu, M. Madhavilath, Zakira Farheen, Sandhya Kumari V, "Content Based Image Retrieval Using Independent Component Analysis”, IJCSNS International Journal ofComputer Science and Network Security, VOL.8 No.4, April 2000.

[18] H.B. Kekre, V.A. Bharadi, S.D. Thepade, B.K. Mishra, S.E. Ghosalkar, S.M. Sawant, "Content Based Image Retreival Using Fusion of Gabor Magnitude and Modified Block Truncation Coding," icetet, pp.140-145, 2010 3rd International Conference on Emerging Trends in Engineering and Technology, 2010.

[19] Flickner M. et al, “Query by image and video content: the QBIC system”, IEEE Computer 1995, Volume 28, Number 9, pp 23-32, 1995.

[20] Sinora Banker Ghosalkar, Vinayak A.Bharadi, Sanjay Sharma, Asif Ansari, "Feature Extraction using Overlap Blocks for Content based Image Retreival" International Journal of Computer Applications (0975-8887), Volume 28-No.7, August 2011.

[21] Yu-chen, Hu,Bing-Hwang, Su Chih-Chiang Tsou.”Fast VQ codebooksearch algorithm for grayscale imagecoding”,Science Direct Image andvision computing 26(2008) 657-666.

[22] Asmita A. Bardekar, Mr. P.A.Tijare, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (6) , 2011, 2584-2589. ISSN:0975-9646

[23] JariKangas, ”Increasing the Error Tolerance in Transmission of Vector Quantized Images by Self-Organizing Map”,Helsinki University of Technology, Finland.

[24] Dr. H B Kekre, T K Sarode, V A Bharadi, A AAgrawal, R J Arora , M C Nair, "Iris Recognition Using Discrete Cosine Transform and Kekre’s Fast Codebook Generation Algorithm", Pages 36-42, ACM New York, NY, USA ©2010

[25] H. B. Kekre, Tanuja K. Sarode, “New Fast Improved Codebook Generation Algorithm for Color Images using Vector Quantization,” International Journal of Engineering and Technology, vol.1, No.1, pp.: 67-77, September 2008.

[26] H. B. Kekre, Tanuja K. Sarode, “Fast Codebook Generation Algorithm for Color Images using Vector Quantization,” International Journal of Computer Science and Information Technology, Vol. 1, No. 1, pp.: 7-12, Jan 2009.

[27] Manoj Kumar, Poonam Saini, "Image Compression With Efficient Codebook Initialization Using LBG algorithm", International Journal of Advanced Computer Technology (IJACT), ISSN:2319-7900, vol. 2, no. 2

[28] Mukesh Mittal, Ruchika Lamba, "Image Compression Using Vector Quantization Algorithms: A Review", International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277 128X), Volume 3, Issue 6,June 2013.

[29] Aasif Ansari, Muzammil H. Mohammed,“Content Based Video Retrieval Systems - Methods, Techniques, Trends and Challenges”, International Journal of Computer Applications (ISBN : 973-93-80885-36-9),Volume 112 – No. 7, February 2015