

AN EFFECTIVE CBVR SYSTEM BASED ON MOTION, QUANTIZED COLOR AND EDGE DENSITY FEATURES

Kalpana Thakre¹, Archana Rajurkar² and Ramchandra Manthalkar³

¹Department of I.T., Sinhgad College of Engineering, Pune

Kalpana_sunil@yahoo.com

²Professor and Head, Department of I.T. & CSE, MGM College of Engineering, Nanded

Archana_rajurkar@yahoo.com

³Professor, Department of Electronics, SGGSI&T, Nanded

R_manthalkar@gmail.com

ABSTRACT

Rapid development of the multimedia and the associated technologies urge the processing of a huge database of video clips. The processing efficiency lies on the search methodologies utilized in the video processing system. Usage of inappropriate search methodologies may make the processing system ineffective. Hence, an effective video retrieval system is an essential pre-requisite for searching a relevant video from a huge collection of videos. In this paper, an effective content based video retrieval system based on some dominant features such as motion, color and edge is proposed. The system is comprised of two stages, namely, feature extraction and retrieval of similar video clips for the given query clip. Prior to perform the feature extraction, the database video clips are segmented into different shots. In the feature extraction, firstly, the motion feature is extracted using Squared Euclidean distance. Secondly, color feature is extracted based on color quantization. Thirdly, edge density feature is extracted for the objects present in the database video clips. When a video clip is queried in the system, the second stage of the system retrieves a given number of video clips from the database that are similar to the query clip. The retrieval is performed based on the Latent Semantic Indexing, which measures the similarity between the database video clips and the query clip. The system is evaluated using the video clips of format MPEG-2 and then precision-recall is determined for the test clip.

.Keywords

Content based video retrieval (CBVR) system, shot segmentation, motion feature, quantized color feature, edge density, Latent Semantic Indexing (LSI).

1. INTRODUCTION

Recent developments in computer technology, especially in the area of storage technology, have led to a considerable growth in the quantity and quality of multimedia databases. This has provided a good platform for the demand for information retrieval from these large databases. In general, multimedia contents comprise audio and video signals and have no indexes. So, conventional text-based information retrieval methods cannot be directly employed to multimedia contents. But, due to the high cost, manual indexing to these databases is generally impractical [2].

The traditional text-based search experiences from the subsequent drawbacks: (1) Manual annotations are time consuming and costly to implement. With the increase in the number of media in a database, the complexities in determining the required information also increases. To manually annotate all attributes of the media content is a difficult task [8]. (2) Manual

annotations fall short in handling the difference of subjective perception. The phrase “a picture is worth a thousand words” describes that the textual description is not necessary for representing subjective perception. Acquiring the entire attributes of the content of any media is unachievable [2]. For this reason, we require a good search technique – Content-Based Video Retrieval System (CBVR). In other words, content-based is defined as the search which will examine the original the image contents. Here, content relates to color, shapes, textures, or any other information that can be obtained from the image directly [12].

Recently, CBVR system has been widely studied. In CBVR, vital information is automatically taken out by employing signal processing and pattern recognition techniques for audio and video signals [1]. Content-based video retrieval [6] offers a valuable set of techniques and is mainly associated with video annotation by the use of key frames. Generally, image annotation has been associated with modelling latent visual concepts in images [3, 5] or joint occurrences of local descriptors and tags [4]. Multiple instance learning methods have also been employed to discover local features related with a concept [7].

Video retrieval can be categorized into two divisions, that is, segmentation and the retrieval system design. Segmentation in video retrieval includes splitting larger video units (scenes, clips) into smaller units (shots, key-frames). The variations between adjacent frames have to be discovered to segment the video shot [17]. The video retrieval system can be broadened into two major parts: a module for the extraction of representative features from video segments and defining an appropriate similarity model to arrange similar video clips from video database [10]. A content-based visual query system needs a number of basic components with visual feature extraction, feature indexing data structure, distance (or similarity) measurement, fast search methods, integration of different visual features, and integration of visual features as well as text-based indices [11]. Consequently, the design of retrieval system comprises of three distinct sub-sections that consist of visual feature extraction (color, texture and shape), multidimensional indexing and the techniques of querying the system [13]. Feature extraction is the elementary foundation for content-based video retrieval.

Further, feature extraction techniques can be categorized as domain-specific features or general features. General features contain shape and color whereas domain specific relates to application-dependent features; for example, face recognition in an identification system [13]. The three major facets of feature extraction are presently being researched. Color extraction has been the most widely utilized feature, because it provides a consistent depiction even in the presence of variations of light, scope and angle [16]. More specifically, the Color Histogram Technique is employed to extract this characteristic [14].

Considerable work has been done to augment the proficiency of this technique, combining research into perceptually weighted histograms to remove the problem of large histogram distances between perceptually similar images [15]. By employing semantic features, it is feasible to identify high-level semantic actions and to encode more semantic details, which will allow users to determine solutions in a better way and in lesser time [9]. Yet, the effectiveness remains an open problem. In order to retrieve the similar videos for a given query video, numerous alternatives are available from random browsing to explore by the query clip. This preference of alternatives increases when the content-based retrieval of video is analyzed. A properly designed system would enable the user to query by a single image, and also by a shot, scene or entire clip of video. The aforementioned details illustrate that the content-based search and retrieval of video data becomes a demanding and vital problem [2].

Here, we propose an effective CBVR system based on some dominant features such as motion, color and edge. Prior to perform the feature extraction, the database video clips are segmented

into different shots. The system is comprised of two stages, namely, feature extraction and retrieval of similar video clips for the given query clip. In the feature extraction, motion features is extracted using Squared Euclidean distance whereas color feature is extracted based on color quantization. Then, edge density feature is extracted for the objects present in the database video clips. When a video clip is queried, the second stage of the system retrieves a given number of video clips from the database that are similar to the query clip. The retrieval is performed based on the LSI, which measures the similarity between the database video clips and the query clip. The rest of the paper is organized as follows: Section 2 makes brief review about the recent research works available in the literature. Section 3 describes the proposed CBVR system with required mathematical formulations. Section 4 discusses about the implementation results and Section 5 concludes the paper.

2. RELATED WORKS

Che-Yen Wen et al. [18] have applied the technology of moving-object tracking to content base video retrieval. They have utilized the background subtraction to detect moving pixels, overcome the shadow problem, then they have used connected components labelling and morphological operations to eliminate noise and mend the moving pixels. Then, with color histograms, color similarity and motion vector, they have extracted the target's image and information for content base video retrieval in the database. Their method has been applied to image frame retrieval in single-CCD (Charge Coupled Device) or multi-CCD surveillance systems. For multi-surveillance retrieval, abrupt environment changes (such as light), CCD shift, viewing angle and position will cause detection and retrieval error.

Heng Tao Shen et al. [19] have presented a graph transformation and matching approach to the video subsequence identification problem, with extension to identify the occurrence of potentially different ordering or length due to content editing. With a batch query algorithm to retrieve similar frames, the mapping relationship between the query and database video was first represented by a bipartite graph. The densely matched parts along the long sequence were then extracted, followed by a filter-and-refine search strategy to prune some irrelevant sub sequences. During the filtering stage, maximum size matching was deployed for each sub-graph constructed by the query and candidate subsequence to obtain a smaller set of candidates. During the refinement stage, sub-maximum similarity matching was devised to identify the subsequence with the highest aggregate score from all candidates, according to a robust video similarity model that incorporated the visual content, temporal order, and frame alignment information. The performance studies were conducted on a long video recording of 50 hours have validated that their approach was promising in terms of both search accuracy and speed.

Kai Sun et al. [20] have proposed a system to understand video affective content automatically. The primary task was to transform the abstract concept of emotion into the form which could be handled by the computer easily. An improved V-A emotion space was proposed to address the aforesaid problem. It has unified the discrete and dimensional emotion model by introducing the typical fuzzy emotion subspace. Fuzzy C-mean clustering (FCM) algorithm was adopted to divide the V-A emotion space into the subspaces and Gaussian mixture model (GMM) was used to determine their membership functions. Based on their proposed emotion space, the maximum membership principle and the threshold principle were introduced to represent and recognize video affective content. A video affective content database has created to validate their proposed model. Their experimental results have shown that the improved emotion space could be used as a solution to represent and recognize video affective content.

Liu David and Chen Tsuhan [21] have presented a framework for content-based video retrieval. They have utilized an unsupervised learning method to automatically discover and locate the object of interest in a video clip. Their unsupervised learning algorithm has alleviated the need

for training a large number of object recognizers. Regional image characteristics were extracted from the object of interest to form a set of descriptors for each video. An ensemble-based matching algorithm compared the similarity between two videos based on the set of descriptors each video contains. Videos containing large pose, size, and lighting variations were used to validate their approach.

Ghoshal et al. [22] have empirically demonstrated that even within the relatively narrow domain of news videos collected from a variety of news programs and broadcasters, the assumption of distributional similarity of visual features did not hold across programs from different broadcasters. That was manifested in considerable degradation of ranked retrieval performance on novel sources. They have observed that the concepts whose spatial locations remain relatively fixed between various sources were also more robust to source mismatches, and vice versa. They have also showed that a simple averaging of multiple visual detectors is more robust than any of the individual detectors. In addition, they have showed that for certain sources using only 20% of the available annotated data can bridge roughly 80% of the performance drop, while others can require larger amounts of annotated data.

Sebastine et al. [23] have the aim to provide a semantic Web based video search engine. Currently, they did not have scalable integration platforms to represent extracted features from videos, so that they could be indexed and searched. The task of indexing extracted features from videos is a difficult challenge, due to the diverse nature of the features and the temporal dimensions of videos. They have presented a semantic Web based framework for automatic feature extraction, storage, indexing and retrieval of videos. Videos were represented as interconnected set of semantic resources. Also, they have suggested a ranking algorithm for finding related resources which could be used in a semantic Web based search engine.

3. THE PROPOSED CONTENT BASED VIDEO RETRIEVAL SYSTEM

The proposed CBVR system is comprised of the three processes, namely, shot segmentation, feature extraction and retrieval of video clips based on query clip. The processes are detailed as follows.

3.1. Shot segmentation

In the process of shot segmentation, the entire video clips are separated into ‘chunks’ or video shots. Consider the database video clips as v_i ; $0 \leq i \leq N_v - 1$, in which each clip is constituted of f_{ij} ; $0 \leq j \leq N_{f_i} - 1$ frames of size $M \times N$. In other words, the shot segmentation can also be defined as the grouping of consecutive frames based on the captured shots. In the proposed retrieval system, the shot segmentation is performed by applying biorthogonal wavelet transformation to every frame of a video clip and then by calculating the L2-norm distance between every frame. Firstly, all the frames of every i^{th} video clip is transformed to biorthogonal wavelet transformation domain as

$$F_{ij}(x, y) = X(x, y)f_{ij}(x, y)X^{-1}(x, y); 0 \leq x \leq M - 1, 0 \leq y \leq N - 1 \quad (1)$$

where, F_{ij} is the frame f_{ij} in the wavelet domain and X is the transformation matrix of the biorthogonal wavelets. Then, the frames f_{ij} and f_{ij-1} are chosen as same shot when it satisfies the condition $L2_{ij} \leq S_T$. If $L2_{ij} > S_T$, then the frames f_{ij} and f_{ij-1} belongs to different

International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 2, April 2011
shots, where, S_T is the threshold to separate shots and $L2_{ij}$ is the L2 norm distance between f_{ij} and f_{ij-1} that can be determined as

$$L2_{ij} = \sqrt{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (|F_{ij}(x, y) - F_{ij-1}(x, y)|)^2} \quad (2)$$

By checking all the consecutive frames, they are separated based on its belonging shots. Hence, N_s numbers of shots are obtained for every i^{th} video clip and $f_{ikl}^{(j)}$ be the frame that belongs to the k^{th} shot of i^{th} video clip. Once the frames of all the database images are segmented based on shots, they are subjected to the process of feature extraction.

3.2. Feature Extraction

In the process of feature extraction, here, some dominant features such as Motion feature, Quantized color feature and Edge density are determined. The feature extraction process is described as follows.

3.2.1. Motion feature extraction

Motion features are any components in the video clip that exhibits motion i.e. the components that shows movement in the consecutive frames. They are extracted by dividing the frames of a particular shot into several blocks and then by identifying the blocks that exhibits motion in the consecutive frames. Hence, each frame of the k^{th} shot is sub-divided into n_b blocks (each of size $m \times n$) and then the SED is determined for the blocks of two consecutive frames as follows

$$SED_{ikl}^{(j)}(c) = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (b_c^{(j)}(x, y)_{ikl} - b_c^{(j)}(x, y)_{ik(l+1)})^2 \quad (3)$$

Based on the SED, the blocks that has moving object are identified and their index are appended as follows

$$[b_M]_{ikl}^{(j)} \ll b_{c_{ik(l+1)}}^{(j)} \quad ; \text{ if } SED_{ikl}^{(j)}(c) < M_T \quad (4)$$

Thus obtained blocks $[b_M]_{ikl}^{(j)}$ for every l^{th} frame belongs to k^{th} shot of a video clip is constituted by a moving object. Hence, the blocks are stored as the motion feature of the i^{th} video clip and they are stored in the feature database as a vector.

3.2.2. Extraction of Quantized Color feature

Color quantization or color image quantization is a process that lessens the number of individual colors employed in an image or frame of a video clip, generally with the intent that the new image should be as visually identical to the original image. Here, with the aid of the color quantization process, the color features are extracted from the

International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 2, April 2011
 shot segmented video clips. To accomplish this, firstly, all the frames of every shot of the video clip is converted from RGB color space to L^*a^*b color space as follows

$$L^* = \begin{cases} 116 \sqrt[3]{Y/Y_0} - 16 & ; \text{if } Y/Y_0 > 0.008856 \\ 903.3 \frac{Y}{Y_0} & ; \text{otherwise} \end{cases} \quad (5)$$

$$a^* = 500(\psi(X/X_0) - \psi(Y - Y_0)) \quad (6)$$

$$b^* = 200(\psi(Y/Y_0) - \psi(Z - Z_0)) \quad (7)$$

In Eq. (5), Eq. (6) and Eq. (7), X_0, Y_0 and Z_0 are the tristimulus values of the reference white. The $\psi(t)$ and the XYZ model of the RGB color space are obtained as

$$\psi(t) = \begin{cases} \sqrt[3]{t} & ; \text{if } t > 0.008856 \\ \frac{7.787 t + 16}{116} & ; \text{otherwise} \end{cases} \quad (8)$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = T \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

where, T is the Transformation Matrix and generally, it can be given as

$$T = \begin{bmatrix} 0.412435 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \quad (9)$$

The color space converted frames are then divided into blocks as done earlier. Then, each block of every frame is subjected to DCT transformation as

$$B_c^{(j)}(u, v)_{ikl} = \alpha_u \alpha_v \sqrt{\frac{2}{m}} \sqrt{\frac{2}{n}} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} b_c^{(j)}(x, y)_{ikl} \cos\left(\frac{\pi u}{2m}(2x+1)\right) \cos\left(\frac{\pi v}{2n}(2y+1)\right) \quad (10)$$

where,

$$\alpha_u = \begin{cases} \sqrt{1/2} & ; \text{if } u = 0 \\ 1 & ; \text{if } 1 \leq u \leq m-1 \end{cases} \quad (11)$$

$$\alpha_v = \begin{cases} \sqrt{1/2} & ; \text{if } v = 0 \\ 1 & ; \text{if } 1 \leq v \leq n-1 \end{cases} \quad (12)$$

The obtained blocks $B_c^{(j)}(u,v)_{ikl}$ which are in DCT domain are scanned in zigzag fashion. While, zigzag scanning of a block, the first N_c elements of the block are extracted as $\left[b_c^{zigzag} \right]_{ikl}^{(j)}$ and it is stored as the quantized color feature of the block.

3.2.3. Edge density extraction

The edge density feature is an attribute of a video clip that can indicate the clip frames by means of magnitude of the edge of any object present in the clip. To extract the feature, firstly, the shot segmented video clip is resampled so that the frames of the shot segmented video clip accomplish the size of $M_r \times N_r$. The resampled frames of the shot segmented video clip are subjected to gray scaling operation and so that every frame of the shot segmented video clips that are in RGB color space is converted to gray scale. Then, two pixel distances are determined in Eq. (13) and Eq. (17) as follows

$$d_{1ikl}^{(j)} = \left| f_{gray}^{(j)}(x, y) - f_{gray}^{(j)}(x-1, y-1) \right|; 1 \leq x \leq M_r - 1, 1 \leq y \leq N_r - 1 \quad (13)$$

Once the distance is calculated, an edge preserving operation is performed based on the obtained distance and so three classes of edges are obtained as follows

$$E_{1ikl}^{(j)}(x, y) = \begin{cases} G_{\max} & ; \text{if } d_{1ikl}^{(j)}(x, y) > G_{LT} \\ G_{\min} & ; \text{otherwise} \end{cases} \quad (14)$$

$$E_{2ikl}^{(j)}(x, y) = \begin{cases} G_{\min} & ; \text{if } d_{1ikl}^{(j)}(x, y) > G_{LT} \\ G_{\max} & ; \text{otherwise} \end{cases} \quad (15)$$

$$E_{3ikl}^{(j)}(x, y) = \begin{cases} G_{\max} & ; \text{if } d_{1ikl}^{(j)}(x, y) > G_{HT} \\ G_{\min} & ; \text{otherwise} \end{cases} \quad (16)$$

$$d_{2ikl}^{(j)}(x, y) = \left| f_{gray}^{(j)}(x, y+1) - f_{gray}^{(j)}(x, y) \right|; 0 \leq x \leq M_r - 2 \text{ and } 0 \leq y \leq N_r - 2 \quad (17)$$

$$E_{1ikl}^{(j)}(x, y) = \begin{cases} G_{\max} & ; \text{if } E_{1ikl}^{(j)}(x, y) = 0 \text{ and } d_{2ikl}^{(j)}(x, y) > G_{LT} \\ G_{\min} & ; \text{otherwise} \end{cases} \quad (18)$$

$$E_{2_{ikl}}^{(j)}(x, y) = \begin{cases} G_{\min} & ; \text{if } E_{1_{ikl}}^{(j)}(x, y) = 0 \text{ and } d_{2_{ikl}}^{(j)}(x, y) > G_{LT} \\ G_{\max} & ; \text{otherwise} \end{cases} \quad (19)$$

$$E_{3_{ikl}}^{(j)}(x, y) = \begin{cases} G_{\max} & ; \text{if } E_{3_{ikl}}^{(j)}(x, y) = 0 \text{ and } d_{2_{ikl}}^{(j)}(x, y) > G_{HT} \\ G_{\min} & ; \text{otherwise} \end{cases} \quad (20)$$

The edges $E_{1_{ikl}}^{(j)}$, $E_{2_{ikl}}^{(j)}$ and $E_{3_{ikl}}^{(j)}$ obtained in Eq. (14), Eq. (15) and Eq. (16) are based on the Eq. (13) and based on Eq. (14), Eq. (16) and Eq. (17), the edges $E_{1_{ikl}}^{(j)}$, $E_{2_{ikl}}^{(j)}$ and $E_{3_{ikl}}^{(j)}$ are obtained. The final edge obtained in the Eq. (20,) $E_{3_{ikl}}^{(j)}$ is subjected to determine the edge density. This can be accomplished by determining the edge density matrix as follows

$$\rho_{E_{ikl}}^{(j)}(p, q) = C_{ikl}^{\max} \left(\frac{n_b p}{2} + q \right) + 1; 0 \leq p \leq n_b / 2 - 1, 0 \leq q \leq n_b / 2 - 1 \quad (21)$$

where, $C_{ikl}^{\max} \left(\frac{n_b p}{2} + q \right)$ is the number of G_{\max} values present in the c^{th} block ($c = \frac{n_b p}{2} + q$) of the edge $E_{3_{ikl}}^{(j)}$. It is to be noted that the $E_{3_{ikl}}^{(j)}$ is the edge obtained for the l^{th} frame of the k^{th} shot that belongs to the i^{th} video clip. Thus obtained $\rho_{E_{ikl}}^{(j)}$ is stored as the edge density feature vector of the corresponding video clip.

3.3. Retrieval of video clips based on query clip

In the retrieval, the database video clips that are similar to the query clip are retrieved by means of measuring the similarity in between the query clip and the database video clips. When a query clip is given to the proposed retrieval system, all the aforesaid features are extracted as performed for the database video clips. Then, with the aid of LSI, similarity is measured between every database video clip and the query clip.

3.3.1. Similarity measure by LSI

Prior to perform the LSI based similarity measure, the transpose of each feature vector extracted for every video clip is determined so as to obtain the feature vector as column vector. The obtained column vectors for the motion feature of all the database video clips are concatenated and then by appending zeros in the necessary locations, a feature matrix is generated. Then, the column vectors of the next feature, color, are appended just below the particular location of the feature matrix. In other words, the column

International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 2, April 2011

vector for color feature of the 0th video clip is concatenated below the element of the 0th column of the feature matrix. Similarly, feature vectors for all the video clips are performed. The same process is repeated for the final feature vector, edge density. Hence, a feature matrix A of size $\hat{N} \times N_v$ is obtained. When a query clip is given, all the aforesaid features are extracted. Then, the feature vector is converted to column vector and then all the feature vectors are concatenated below (as stated above). Hence, a column feature vector $N_q \times 1$ is obtained for the query clip.

In the first process of LSI based similarity measure, the A is subjected to SVD decomposition. Using the SVD theorem, the matrix A is decomposed as

$$A_{\hat{N} \times N_v} = U_{\hat{N} \times \hat{N}} S_{\hat{N} \times N_v} V_{N_v \times N_v}^T \quad (22)$$

where, $S \in R_{m \times n}$ is a diagonal matrix with nonnegative diagonal elements known as the singular values, $U \in R_{m \times n}$ and $V \in R_{m \times n}$ are orthogonal matrices. The columns of matrices U and V are labeled as the left singular vectors and the right singular vectors respectively and they can be given as

$$U^T U = I_{\hat{N} \times \hat{N}} \quad (23)$$

$$V^T V = I_{N_v \times N_v} \quad (24)$$

A query vector co-ordinate is determined for the query vector as follows

$$q_{co} = q^T \times U \times S^{-1} \quad (25)$$

where, q^T is the transpose of the query vector. Based on the query vector, similarity is determined as

$$d_{3_x} = \frac{1}{N_v} \sum_{y=0}^{N_v-1} \frac{q_{co}^T(x) \cdot (V^T(x, y))}{(q_{co}^T(x)) \cdot (V^T(x, y))} \quad ; 0 \leq x \leq N_v - 1 \quad (26)$$

where, d_{3_x} is the similarity coefficient determined for the x^{th} database video clip. Then, N_r number of similar database video clips are retrieved based on d_{3_x} . This can be accomplished by retrieving the video clips that has maximum d_{3_x} . Hence, the database video clips that are similar to the query clip are retrieved based on the LSI.

4. RESULTS AND DISCUSSION

The proposed retrieval system is implemented in the MATLAB platform (version 7.8) and tested using the database video clips of MPEG-2 format. The frame results obtained

International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 2, April 2011
in the intermediate process of the proposed CBVR system is depicted in the following figures.



(a)



(b)

Figure 1: Results obtained from shot segmentation (a) frames belongs to one shot (b) frames belong to another shot.



(a)

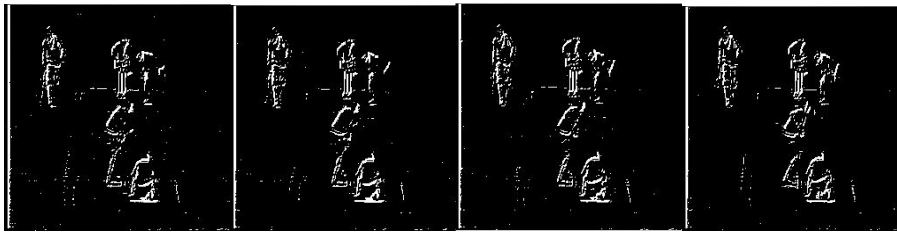


(b)

Figure 2: Sample output from color space conversion: (a) frames in RGB color space (b) frames in L*a*b* color space.



(a)



(b)

Figure 3: Sample output obtained from Edge detection: (a) actual frames (b) edge detection frames



(a)



(b)



(c)



(d)



(e)



(f)

Figure 4: Sample results obtained from the proposed CBVR system (a) some frames that belongs to query clip (b) (c) (d) (e) and (f) some frames belongs to the clips that are retrieved as relevant to the query clip.

In Fig. 1, the frames of a video clip that are shot segmented are depicted. It consists of the frames of two different shots. Fig.2 comprises of some video frames in RGB color space and in LAB color space, that are converted for the purpose of extracting the quantized color feature. The results given in the Fig. 3 depicts the edges of object, which occupies in few frames of the video clip. Fig. 4 illustrates that some frames belong to the query clip as well as the frames of the database video clips that are retrieved as similar to the query clip. Moreover, the proposed system is evaluated by determining the precision and recall using the Eq. (27) and Eq. (28) respectively. The determined precision and recall for a given query image is plotted and the graph that depicts the precision vs recall is given in the Fig. 5

$$precision = \frac{\text{No. of retrieved videos that are relevant to the query clip}}{\text{Total no. of retrieved videos}} \quad (27)$$

$$recall = \frac{\text{No. of retrieved videos that are relevant to the query clip}}{\text{Total no. of available videos that are relevant to the query clip}} \quad (28)$$

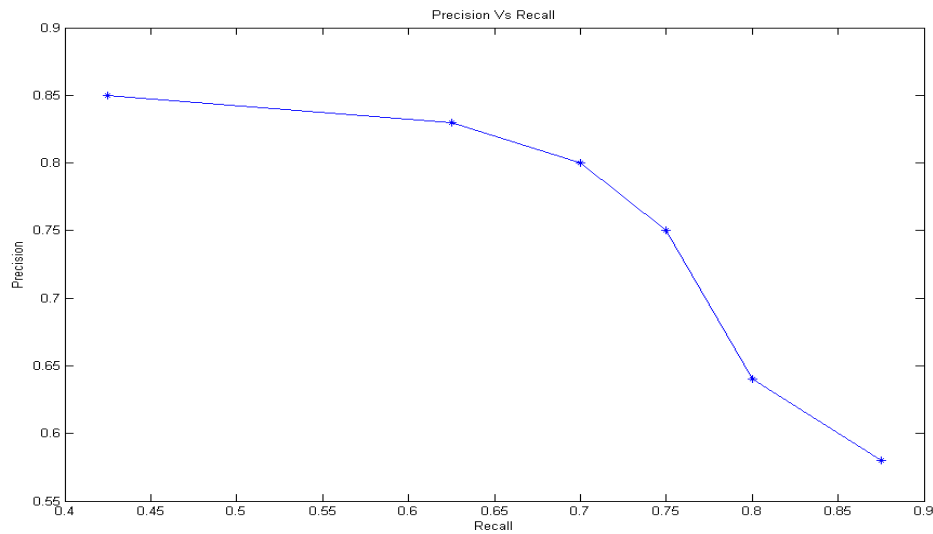


Figure 5: Precision-recall plot for the proposed CBVR system

The depicted results as well as the precision-recall calculation show the effectiveness of the proposed CBVR system.

5. CONCLUSION

In this paper, we have proposed an effective CBVR system based on the dominant features such as motion feature, quantized color feature and edge density feature. The results have shown that the proposed system retrieves the database video clips that are relevant to the query clip in an effective manner. The efficacy of the proposed system has also been shown by the precision-recall values determined for a given query clip. The proposed system is effective mainly because of the features that are proposed to extract from any video clip. The classic color feature, color histogram, is replaced by quantized color feature. The extracted features, motion feature, quantized color feature as well as the edge density feature have the capability of differentiating the video clips. Moreover, the LSI which has been utilized in measuring the similarity has performed well in retrieving the video clips that are relevant to the given query clips.

References

- [1] Koichi Shinoda, Nguyen Huu Bach, Sadaoki Furui, Naoki Kawai, "Scene recognition using hidden markov models for video database", http://www.ks.cs.titech.ac.jp/publication/2005/LKR2005_2-019.pdf
- [2] Petkovic, Milan, Jonker, Willem, "Content-based video retrieval", Kluwer Academic Publishers, Boston, Monograph, 2003, 168 p., Hardcover ISBN: 978-1-4020-7617-6
- [3] Barnard K. and Duygulu P. and Forsyth D. and de Freitas N. and Bleib D. and Jordan M., 'Matching Words and Pictures', J. Mach. Learn. Res., Vol. 3, pp.1107-1135, 2003.
- [4] S. L. Feng, R. Manmatha, and V. Lavrenko, "Multiple Bernoulli relevance models for image and video annotation", In proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol.2, pp II--1002--II--1009, 2004.

- [5] Fei-Fei Li and Pietro Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories", In proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol.2, pp.524-531, 2005.
- [6] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain, "Content-Based Image Retrieval at the End of the Early Years", In proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, pp.1349 - 1380, 2000.
- [7] Yang C. and Lozano-Perez T., "Image Database Retrieval with Multiple-Instance Learning Techniques", In Proceedings of the 16th International Conference on Data Engineering, pp.233-243, 2000.
- [8] Chia-Hung Wei, Chang-Tsun Li, "Content-based multimedia retrieval - introduction, applications, design of content-based retrieval systems, feature extraction and representation", 2004.
- [9] Lili, N.A., "Hidden Markov Model for Content-Based Video Retrieval", Proceedings of the 2009 Third Asia International Conference on Modelling & Simulation table of contents, Pages: 353-358, 2009, ISBN:978-0-7695-3648-4
- [10] T.N.Shanmugam and Priya Rajendran, "An Enhanced Content-Based Video Retrieval System Based On Query Clip", ISSN: 2076-734X, EISSN: 2076-7366 Volume 1, Issue 3(December 2009)
- [11] Shih-Fu Chang, "Compressed-Domain Content-Based Image and Video retrieval", Published in Symposium on Multimedia Communications and Video Coding, Polytechnic University, New York, Oct. 1995.
- [12] John Eakins, Margaret Graham, University of Northumbria at Newcastle, "Content-based Image Retrieval" (JISC Technology Applications Programme Report 39 -1999)
- [13] Yong Rui, Thomas S. Huang, and Shih-Fu Chang, "Image Retrieval: Current Techniques, Promising Directions and Open Issues. Jan 7 1999.
- [14] M. J. Swain and D. H. Ballard, "Color Indexing", International Journal of Computer Vision, Vol.7, pp.11 - 32, 1991.
- [15] Lu G. & Phillips J. (1998), "Using perceptually weighted histograms for colour-based image retrieval", Proceedings of Fourth International Conference on Signal Processing, 12-16 October 1998, Beijing, China, pp. 1150-1153.
- [16] Aigrain, P., Zhang, H.J., Petkovic, D., "Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review", MultToolApp(3), No. 3, November 1996, pp. 179-202. 9611.
- [17] Richard Hallows, "Techniques used in the content-based retrieval of digital video", 2nd Annual CM316 Conference on Multimedia Systems, based at Southampton University, UK.
- [18] Che-Yen Wen, Liang-Fan Chang and Hung-Hsin Li, "Content based video retrieval with motion vectors and the RGB color model", Forensic Science Journal, Vol.6, No.2, pp.1-36, 2007.
- [19] Heng Tao Shen, Jie Shao, Zi Huang and Xiaofang Zhou, "Effective and Efficient Query Processing for Video Subsequence Identification", IEEE Transactions on Knowledge and Data Engineering, Vol.21, No.3, pp.321-334, March 2009.
- [20] Kai Sun, Junqing Yu, Yue Huang and Xiaoqiang Hu, "An improved valence-arousal emotion space for video affective content representation and recognition", In proceedings of IEEE International Conference on Multimedia and Expo., pp.566-569, New York, June 2009.
- [21] Liu David and Chen Tsuhan, "Video retrieval based on object discovery", Computer vision and image understanding, Vol.113, No.3, pp.397-404, 2009.
- [22] Ghoshal, Khudanpur and Klakow, "Impact of novel sources on content-based image and video retrieval", in proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp.1937-1940, April 2009.

- [23] Sebastine, Thuraisingham and Prabhakaran, "Semantic Web for Content Based Video Retrieval", In proceedings of IEEE International Conference on Semantic Computing, p.103-108, September 2009.

Authors

(1)



Kalpana S. Thakre received degree of B.E. (Computer Engineering) from Kavikulguru Institute of Technology and Science, Nagpur in First Division in 1993, M.E.(Computer Science and Engineering) in First Division in 2006 from G.H.Raisoni College of Engineering, Nagpur, Nagpur University and currently she is a research scholar of Shri Gurugovind Singh Institute of Information and Technology, SRTM University, Nanded. She has published 2 research papers in International journal, 12 research papers in International conferences, 3 research papers in National conferences. Currently she is Assistant Professor in the Department of Information Technology, Sinhgad College of Engineering, Pune, India. Her research interests are Content-based Image and Video Retrieval, Multimedia and Image Databases, Image Processing.

(2)

Archana M. Rajurkar received degree of B.E. (Computer Engineering) from MGM's College of Engineering, Nanded in First Division in 1991, M.E. (Instrumentation) in First Division in 1997 from SGGGS Institute of Engineering and Technology, Nanded, Marathwada University, Aurangabad and Ph.D from Indian Institute of Technology Roorkee, Roorkee, India in 2003. She has published one research paper in International journal, one paper in National Journal, 12 research papers in International conferences, 10 research papers in National conferences and one paper as a Book chapter. She joined as a faculty member in M.G.M.'s College of Engineering, Nanded in 1991. She was Head, Department of Computer Science and Engineering, M.G.M.' College of Engineering, Nanded from April 1997 to June 1999 and from November 2003 to till date. Currently she is Professor and Head in the Department of Computer Science and Engineering, M.G.M.'s College of engineering, Nanded, India. Her research interests are Content-based Image and Video Retrieval, Multimedia and Image Databases, Image Processing and Computer Vision.

She is Chairman of Board of Studies of Computer Science and Engineering at SRTM University, Nanded. She is Chairman of Research Recognition Committee (RRC) at SRTM University, Nanded. She is recognised Ph.D guide of SRTM University, Nanded. She is currently guiding 3 students for their Ph.D.

(3)

Dr. R.R.Manthalkar received B.E. (Elect. Engineering) from Shri Guru Govind Singhji College of Engineering and Technology in First Division in 1988, M.E.(Elect. Engineering)from Shri Guru Govind Singhji College of Engineering and Technology,Nanded, in First Division in 1994 and Ph.D from Indian Institute of Technology, Kharagpur, India in 2003. He has published 12 research papers in International journal and 25 research papers in International conferences. Currently he is a Professor in the Department of Electronics and Telecommunication at Shri Guru Govind Singhji College of Engineering and Technology, Nanded. His research interests are Image Processing, Texture Analysis and Computer Vision.