

Available online at <http://www.mecs-press.net/ijwmt>

A Comprehensive CBVR System Based on Spatiotemporal Features Such as Motion, Quantized Color and Edge Density Features

Kalpana S. Thakre^a, Dr. Archana M. Rajurkar^b

^aAP, Information Technology Dept., Sinhgad College of Engineering, Wadgaon, Pune, India

^bHead, Dept. of CSE and IT, M.G.M. College of Engineering, Nanded, India

Abstract

Rapid development of the multimedia and the associated technologies urge the processing of a huge database of video clips. The processing efficiency depends on the search methodologies utilized in the video processing system. Use of inappropriate search methodologies may make the processing system ineffective. Hence, an effective video retrieval system is an essential pre-requisite for searching relevant videos 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 evaluated using the video clips of format MPEG-2 and then precision-recall is determined for the test clip.

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

© 2011 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science

1. Introduction

Multimedia information systems are increasingly important with the advent of broadband networks, high-powered workstations, and compression standards. Since visual media requires large amounts of storage and processing, there is a need to efficiently index, store, and retrieve the visual information from large volume of data and this data needs to be transmitted for further processing or retrieval. Compressed data will efficiently reduce the storage and the transmission of the compressed data will also make the video accurate and perfect.

Video has both spatial and temporal dimensions and video index should capture the spatial-temporal contents of the scene. In order to achieve this, a video is first segmented into shots, and then key frames are identified and used for indexing and retrieval [2].

Most of the research in the area of video retrieval is being carried out. Video Retrieval in compressed domain is still a young and rapidly evolving field in the area of multimedia. Various factors have contributed towards the triggering of interest in this field. They are 1) much of the multimedia content available today is in

* Corresponding author.

E-mail address: ^akalpana_sunil@yahoo.com; ^bArchana_rajurkar@yahoo.com

compressed format already and most of the new video and audio data produced and distributed will be in standardized, compressed format. Using compressed-domain features directly makes it possible to build efficient and real-time video indexing and analysis systems. 2) Some features, such as motion information, are easier to extract from compressed data without the need of extra, expensive computation. Of course, most features can be obtained from uncompressed data as well, usually with a higher precision but at a much higher computational cost. 3) In practical systems, trade-off between efficiency and accuracy can be explored.

Video retrieval can be categorized into two parts, 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)[1]. The variations between adjacent frames have to be discovered to segment the video shot [2]. 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 [3]. A content-based visual query system needs a number of basic components with visual feature extraction [4], 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 [5]. 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 [6]. 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 [6]. 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 [7]. More specifically, the Color Histogram Technique is employed to extract this characteristic [8] and [9].

Here, we propose an effective CBVR system based on some dominant features such as motion, color. 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. 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 describes the proposed CBVR system. Section 3 discusses about the implementation results and Section 4 concludes the paper.

2. 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. This process is detailed as follows.

2.1. Shot Segmentation

In the process of shot segmentation, the entire video clips are separated into ‘chunks’ or video shots[10]. 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.

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 frames that belong 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.

2.2. Feature Extraction

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

2.3. 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. 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.

2.4. 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 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.

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

The obtained blocks 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.

2.5. Retrieval of video clips based on query clip

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.

2.6. 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.

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

2.7. Similarity measure by LSI

In this section we perform the LSI based similarity measure.

For every video clip, the feature vector which we have computed earlier, the transpose of this each feature vector extracted is computed. This will convert every feature vector in the column vector form. The 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 vector for color feature of the 0th video clip is concatenated below the element of the 0th column of the feature matrix. Same process is performed for all the feature vectors for all the video clips. 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 matrix A is subjected to SVD decomposition. Hence, the database video clips that are similar to the query clip are retrieved based on the LSI.

3. Results And Discussions

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 in the intermediate process and the retrieval process of the proposed CBVR system is captured.

Moreover, the proposed system is evaluated by determining the precision and recall using “(1)” and “(2)” respectively. The determined precision and recall for a given query image is plotted and the graph that depicts the precision vs recall is computed.

$$precision = \frac{\text{No. of retrieved videos that are relevant to the query clip}}{\text{Total no. of retrieved videos}} \quad (1)$$

$$recall = \frac{\text{No. of retrieved videos that are relevant to the query clip}}{\text{Total no. of available videos that are relevant to query clip}} \quad (2)$$

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

4. Conclusion

In this paper, we have proposed an effective CBVR system based on the dominant features such as motion feature, quantized color 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 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] 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.
- [2] Richard Hallows, "Techniques used in the content-based retrieval of digital video", 2nd Annual CM316 Conference on Multimedia Systems, based at Southampton University, UK.
- [3] 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).
- [4] Chia-Hung Wei, Chang-Tsun Li, "Content-based multimedia retrieval - introduction, applications, design of content-based retrieval systems, feature extraction and representation", 2004.
- [5] 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.
- [6] Yong Rui, Thomas S. Huang, and Shih-Fu Chang, "Image Retrieval: Current Techniques, Promising Directions and Open Issues. Jan 7 1999.
- [7] Aigrain, P., Zhang, H.J., Petkovic, D., "Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review", *MultiToolApp*(3), No. 3, November 1996, pp. 179-202. 9611.
- [8] M. J. Swain and D. H. Ballard, "Color Indexing", *International Journal of Computer Vision*, Vol.7, pp.11 - 32, 1991.
- [9] 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.
- [10] 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.