Content Based Video Retrieval using trajectory and Velocity features

Dr. S. D. Sawarkar [1], V. S. Kubde [2]
[1]Principal, DMCE,Airoli,Navi-Mumbai,India
[2]Department Of Information Technology, DMCE,Airoli,Navi-Mumbai,India
[1]Sudhir_Sawarkar@yahoo.com, [2]Vidya802000@yahoo.com

Abstract: The Internet forms today's largest source of Information containing a high density of multimedia objects and its content is often semantically related. The identification of relevant media objects in such a vast collection poses a major problem that is studied in the area of multimedia information retrieval. Before the emergence of content-based retrieval, media was annotated with text, allowing the media to be accessed by text-based searching based on the classification of subject or semantics.In typical content-based retrieval systems, the contents of the media in the database are extracted and described by multi-dimensional feature vectors, also called descriptors. In our paper to retrieve desired data, users submit query examples to the retrieval system. The system then represents these examples with feature vectors. The distances (i.e., similarities) between the feature vectors of the query example and those of the media in the feature dataset are then computed and ranked. Retrieval is conducted by applying an indexing scheme to provide an efficient way to search the video database. Finally, the system ranks the search results and then returns the top search results that are most similar to the query examples.Therefore, a content-based retrieval system has four aspects: feature extraction and representation, dimension reduction of feature, indexing, and query specifications. With the search engine being developed, the user should have the ability to initiate a retrieval procedure by using video retrieval in a way that there is a better chance for a user to find the desired content.

Index Terms: Compression, Content Based video retrieval, motion vector, trajectory.

I. INTRODUCTION

Recent advances in multimedia technologies allow the capture and storage of video data with relatively inexpensive computers. Furthermore, the new possibilities offered by the information highways have made a large amount of video data publicly available .However, without appropriate search techniques all these data are hardly usable. Users are not satisfied with the video retrieval systems that provide analogue VCR functionality. They want to query the content instead of the raw video data. For example, a user analysing a soccer video will ask for specific events such as goals. Content-based search and retrieval of video data becomes a challenging and important problem. Therefore, the ed for tools that can manipulate the video content in the same way as traditional databases manage numeric and textual data is significant. This paper presents our approach for content-based video retrieval. It is organised as follows. In the next section, we give an overview of related work. The third section describes our approach with emphasis on the video modelling as one of the most critical processes in video retrieval. The architecture of a content-based video retrieval system is presented in the fourth section. The fifth section draws conclusion.

II.THE STATE OF THE ART OVERVIEW

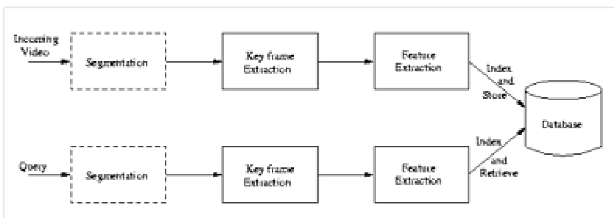
There are plenty of image video retrieval systems [1] based on the spatial features such as color, texture and shape, where as systems using motion related information which distinguishes the image from a video are rare in literature **.** There are few research work done in compressed domain video indexing and retrieval [3,4,5]. They use features such as motion vectors and DCT coefficients. Since the features are directly extracted from the compressed MPEG video, the costly overhead of decompressing and operating at pixel level is avoided. So far all the compressed domain video indexing and retrieval techniques available in the literature [3,4, 51 are operating at the frame level without considering the contents of the video. The underlying semantic content of the video is given by the characteristics of the objects present in the video. So it is essential to describe the video objects for efficient indexing and retrieval of the video.Though there are few papers [6] incorporate the above task in pixel domain, the computational cost involved in the pixel domain process is very high. The video indexing and retrieval problem has been addressed by researchers in a number of ways. A survey of digital video parsing and indexing technologies, mainly in the pixel domain, is presented in the paper by Ahanger and Little [7], including a discussion of research trends in video indexing and requirements of future data delivery systems. Topics such as video data indexing, video data modeling, information extraction, and video scene segmentation are also presented..In other work, Zhang et al. [8]describe techniques for use in the pixel domain for dealing with the representation of shot content, as well as content-based retrieval techniques using key frames and temporal properties of shots. They also present techniques for video parsing in the pixel domain followed by key frame extraction. The representation of shot content is based on several types of features, including color histogram and moment features, texture features, shape features, and edge features. Similar work can be found in [9] by Nagasaka and Tanaka who present pixel-domain techniques for performing full-video search for specified objects using features derived locally**.** Two recent papers by Flickner et al. [10] describe the QBIC system, which performs content-based retrieval based on color, shape, texture, and sketches in large image and video databases. Ardizzone et al. [11, 12] also deal with content-based video indexing based on motion, color, and texture and other global features. All the aforementioned papers present techniques in the pixel domain, but much less has been done in the compressed domain. .Idris and Panchanathan [13] propose an algorithm based on vector quantization (VQ) for indexing of video sequences in compressed form. During compression, the image is decomposed into vectors and mapped to a finite set of code words and encoded using adaptive VQ. Each frame is represented by a set of labels and a codebook which are used to generate indices..A generic paper on compressed-domain video indexing techniques by Chang [14] describes some of the issues involved in addressing such a problem in the DCT, wavelet, and sub band transform compressed domains.

The current techniques, proposed in literature, mostly deal with uncompressed multimedia objects (images and videos). There are several techniques proposed for shot detection and segmentation of compressed video These techniques use block comparison metrics, which measure the differences between DCT coefficients of blocks in two frames.. We assume that images and videos in multimedia databases are stored in compressed form (JPEG for images or MPEG and M-JPEG for videos). We propose a fast retrieval and indexing algorithm that can be very efficiently used for content-based search on the Internet.

Our approach to compressed domain indexing and retrieval can be described in three parts --- segmentation, indexing, and query processing. First, video segmentation divides the incoming video into shots or scenes, and selects one or more key or representative frames for each shot. A shot in a video clip is defined as a maximal sequence of frames resulting from a continuous uninterrupted recording of video data.

These shots may be further subdivided into scenes, for instance, if significant camera motion 2 is present and the view of the camera switches to a different subject.

Second, features are extracted from the key frames supplied by the segmentation process and used to create a database index. The features used are derived from the DWT coefficients and the motion vector information available in the MPEG compressed video.Finally, the database is accessed using the features derived from a query clip as an index



In this paper we try to extract various features of the video objects from the compressed video shot. Though defining a semantic video object is a difficult task, in most of the cases a video object can be defined **as** a coherently moving image region. The readily available motion vector information from the compressed MPEG video is used for segmenting the video objects from background.

III. SYSTEM OVERVIEW

The proposed system initially takes the compressed MPEG video shots and extract motion vectors by partial decoding of the MPEG stream. The proposed system consists of the following three stages (i) Motion vector processing (ii) Object segmentation and tracking (iii) Feature extraction. Fig.1 shows the various parts ***of*** the proposed system. Since the motion vectors obtained from the MPEG video are noisy, they can not be directly used for object segmentation. To increase the reliability of motion information, the motion vectors of few neighbouring frames on either side of the current frame are also used, the details of this process is given in the next section. The segmentation stage takes the motion information obtained from the previous stage for segmenting the coherently moving video objects . The segmented objects are tracked temporally to get the trajectory information of the object. Section **4** describes the aforementioned segmentation and tracking phase. Finally from the segmented object and the consecutive tracking, the features of the corresponding object such **as** velocity, location of object center and motion information are computed for indexing the scene. The global features such as motion activity and camera motion are directly obtained from the motion vector processing block.

IV.COMPRESSION

Introduction Compression Algorithm (WWT)
This program is depends on Wavelet and Walsh transform for transformation, and then using
Arithmetic coding for compress an image.
My Compression algorithm consists of the following steps:
A) Two Levels Discrete Wavelet Transform
B) Apply 2D Walsh-Hadamard Transform on each 8x8 block of the low-frequency sub-band
C) Split all DC values form each transformed block 8x8
D) Compress each sub-band by using Arithmetic coding
For example
If suppose original image is of size 1.97 Mb
Than after walsh and wavelette transform the size is reduced to 75 Kb and the image gets converted from bmp to wwt format.

V.MOTION VECTOR PROCESSING

MPEG compressed video provides one motion vector (usually noisy) for each macro block of size 16 **x** 16 pixels. To remove noise, the motion vectors are subjected to the following steps
(i) Motion Accumulation
(ii) Determination of representative motion vectors.
Motion Accumulation Initially, the motion vectors are scaled appropriately to make them independent of frame type **[3].** This is accomplished by dividing the motion vectors by the difference between the corresponding frame number and the reference frame number (in the display order). Then, the motion vectors are rounded to nearest integers. In the case of bidirectionally predicted macroblocks, reliable motion information is obtained either from the forward or backward motion vector depending upon which of the reference frames **(LT)** is closer. If backward prediction is chosen, the sign of the motion vector is reversed after normalization.
The motion vector obtained from current macroblock of the current frame n is assigned to the center pixel ***(k, I)*** of that macroblock. Let *mi'(n* - c) and mf(n - ***c)*** represent the motion vectors along the horizontal and vertical directions for the macroblock centered at ***(k,*** I) in frame (n - ***c).*** Then, the new position for this macroblock in the current frame can be estimated as:

$$(\hat{k},\hat{l}) = (k,l) + \sum_{f=n-1}^{n-c} \left(m_x^{kl}(f), m_y^{kl}(;)\right)$$

The motion vector (mk'(n - ***c),*** mr(n - ***c))*** in (n - c)th frame is assigned to the new position **(i*i*),** w ith respect to the current frame. Fig. 2 explains the above process for the case of c = 1.The motion accumulation is also done by tracking the frames in forward direction from the current frame. This is achieved by keeping few future frames in the buffer. In forward tracking the motion vectors are accumulated according to the following equation:

IJECSE
International Journal of Electronics and Computer Science Engineering
WWW.IJECSE.ORG

$$(\hat{k},\hat{l}) = (k,l) - \sum_{f=n+1}^{n+c} \left(m_x^{kl}(f), m_y^{kl}(f)\right) \qquad (2)$$

Here, the motion vector (m!'(n + ***c),*** m:'(n + ***c))*** in (n + c)th frame is assigned to the new position **(k,** *i)* with respect to the current frame. Each frame approximately provides one additional motion vector per macroblock.

VI.FEATURE EXTRACTION

The features required for indexing the video shots are extracted from each frame after segmentation. The global shot attributes and the object based features used for indexing the video shots are explained below.

1. Motion Activity: This descriptor gives an idea about the 'pace of action' in the video segment. The value of the descriptor is high for more dynamical shots such as sport sequence and low for video-telephony sequences. The motion activity is measured by the standard deviation of the
magnitudes of motion vectors **[7].**

2. Object Area: This gives the rough estimate of area the object measured from the object mask obtained from the segmentation stage. The object area is represented as the fraction of object macroblocks available in the frame.

3. Velocity of the Object: The object velocity **a** describes the speed of the object along horizontal and vertical direction.

4. Object Trajectory: The trajectory of the object is represented by two second order polynomials, one each for the horizontal and the vertical directions. Both trajectories arecomputed from the motion trail of the object center, represented by a sequence ***{~[i*y*]*[i,]}** **, i E 1,.** . . ,N , where ***I*** indicates the temporal location of the object. Each shot is divided into blocks containing N number of frames (with an overlap of one frame) and the trajectories of the object for both horizontal and vertical directions are obtained by fining, in least square sense, a second order polynomial to the above trail.

VII. QUERY RESULTS AND DISCUSSIONS

The user is asked to enter a video clip as query'Our program will do the processing on it Like first the compression will be done by DWT
That will be displayed ,later the motion vector processing will be done on it, which will continue with feature extraction of the query video like trajectory, velocity and area. These features will be compared with the features of the videos in the database , and matching video from the database will be displayed.

VIII. CONCLUSIONS

The use of motion information in the compressed domain allows for rapid analysis of the content of the video. A video indexing and retrieval system based on the motion information obtained from
the compressed MPEG video has been presented. The retrieval can be performed by considering the spatial features such as DWT coefficients that can be easily extracted from the MPEG video.

Fig a

Fig b

Fig c

Fig d

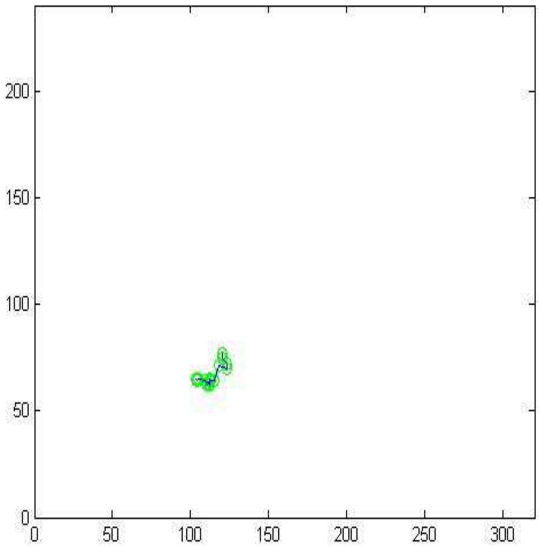
Fig e

ISSN 2277-1956/V1N2-2460-2466

Fig f

Fig g

Fig a Shows a video of moving bus,
Fig b,c,d,e shows frames formed after segmentation
Fig below shows trajectory of the whole video .



IX. REFERENCES

[1] M. Flickner, H. Sawhney, W. Niblack, J. Aashley, Q. Huang,B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic,D. Steele, and P. Yanker, "Query by image and video content:The QBlC system:' IEEE Computer Magazine, vol. 28, pp. 23-32, Sept. 1995.
[2] Sylvie Jeannin, Radu Jasinschi, Alfred She, T. Naveen, Benoit Mory, and Ali Tabatabai, "Motion descriptors for contentbased video representation:' Signal Processing: Iimage Communication, vol. 16, pp. 59-85, 2000.[3] V.

ISSN 2277-1956/V1N4-2460-2466

Kobla, D. S. Doermann, and K-I. Lin, "Archiving, indexing and retrieval of video in compressed domain:' in SPIE Conference on Multimediu Storage and Archiving Systems, 1996, vol. 2916, pp. 78-89.
[4] V. Kobla, D. S. Doermann, K-I. Lin, and C. Faloutsos, "Compressed domain video indexing techniques using dct and motion vector information in MPEG video:' in SPIE Conference on Multimedia Storage anddrchiving Systems, 1997, vol. 3022, pp. 20&2211.
[5] K. Yoon, D. F. DeMenthon, and D. Doermann, "Event detection from MPEG video in the compressed domain," in Int.Con/: on Pattern Recognition, Barcelona, Spain, 2000."An integrated approach for content-based video object segmentation and retrieval:' IEEE
[6] D. Zhong and S. F. Cbang, , G. Ahanger and T.D.C. Little. A survey of technologies for parsing and indexing digital video. Journal of Visual Communication and Image Representation, 7:28{ 43, 1996).
[7] H.J. Zhang, C.Y. Low, S.W. Smoliar, and J.H. Wu. Video parsing, retrieval and browsing: An integrated and content-based solution. In Proc. of the ACM Multimedia Conference, pages 15{24, 1995.}
[8] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: the QBIC system. IEEE Computer, 28(9):23{32, September 1995.}
[9] E. Ardizzone, M. La Casia, V. Di Gesu, and C. Valenti. Content-based indexing of image and video databases by global and shape features. In Proc. of the International Conference on Pattern Recognition, pages 140{144, 1996.
[10] E. Ardizzone, M. La Casia, and Davide Molinelli. Motion and color-based video indexing and retrieval. In Proc. of the International Conference on Pattern
[11] F. Idris and S. Panchanathan. Indexing of compressed video sequences. In Proc. of the SPIE Conference on Storage and Retrieval for Still Image and Video
[12] S.F. Chang. Compressed-domain techniques for image/video indexing and manipulation . In Proc. of the IEEE International Conference on Image Processing.