An Evolving Approach on Video Frame Retrieval Based on Color, Shape and Region

D.Shanmuga Priyaa
Research Scholar, Computer Science Department
Karpagam University
Coimbatore, India
mona2324@gmail.com

T.Nachimuthu
Research Scholar, Computer Science Department
Karpagam University
Coimbatore, India
mano_haran@rediffmail.com

Dr.S.Karthikeyan
Asst.Prof, Department of Information Technology
College of Applied Sciences, Sohar
Sultanate of Oman
skaarthi@gmail.com

Abstract— **This paper proposes a new methodology for matching of objects in video based on the color, shape and region. The objects are segmented and indexed based on the similarity between the frames. The similarity feature such as color, shape and region are measured for the objects between two videos are matched and they are displayed. The similarities between two frames are resulted from three major features such as color, shape and region in order to solve the problem of objects retrieval in video. Applications of our method are the retrieval of objects in video-shot collections or grouping of shots containing the same protagonist into video scenes. MPEG-7 data set are presented and the results are discussed. The experimental results proved the feasibility and effectiveness of the proposed algorithm.**

Keywords- Video objects matching, Object-based segmentation, fuzzy, feature extraction.

I. INTRODUCTION

Multimedia mining systems automatically extract semantic information (knowledge) from multimedia files. Multimedia database systems store and manage a large collection of multimedia objects such as text, audio, image and video [1]. With the increased availability of multimedia data over the Internet, the retrieval of the audiovisual documents has become a challenging task. Recently, computer vision has transition from understanding single images to analyzing image sequences, or video understanding. Video understanding deals with understanding of video sequences such as recognition of gestures, activities, facial expressions, etc. The main shift in the classic paradigm has been from the recognition of static objects in the scene to motion-based recognition of actions and events. Video understanding has overlapping research problems with other fields, therefore blurring the fixed boundaries. The aim of data mining is to identify interesting patterns in data. Video mining can be defined as the unsupervised discovery of patterns in audio-visual content. This task is especially challenging when the data consist of video sequences (which may also have audio content), because of the need to analyze enormous volumes of multidimensional data. To this effect the MPEG-7 standard [2] aims to provide standardized core technologies allowing the description of the audiovisual data. The basic visual features that MPEG-7 standardizes are color, texture, shape and motion [2]. In video, the shape, the size and the structure of objects change mainly due to camera motion, object motion and occlusion phenomena. Thus, the structure of the same object at different times in a video may present significant differences. In video processing one of the issues is Object-based segmentation. The step further towards the retrieval of objects from video is based on similarity. The Figure 1 shows the model of applying multimedia mining in different multimedia types.

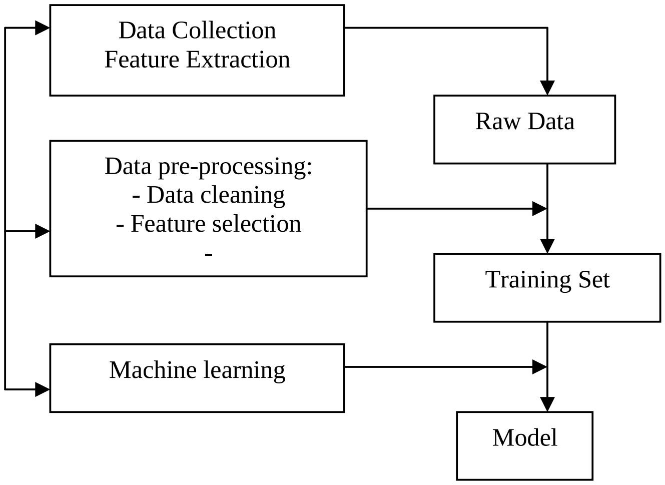

Figure 1. Multimedia mining process

The following are the requirements of the video mining system.

1. It should be unsupervised.

2. It should not have any assumptions about the data

3. It should uncover interesting events.

The existing works in video mining are categorized as mining similar motion patterns and mining similar objects. Video object retrieval is the detection of object from large videos. The second category is used to mine frequently appearing objects in videos. The following steps are most significant in case of content based video retrieval system. They are segmentation, feature extraction and feature grouping. A video segmentation algorithm attempts to divide a video sequence into l subgroups termed as “shots”. In recent times a number of techniques, varying from color histogram to block based approaches with motion compensation have been proposed for video segmentation [3]. This paper deals with the problem of object retrieval from video based on color similarity, shape similarity, region similarity using fuzzy.

II. RELATE D WORK

The earlier work for color similarity deals with the simplest feature such as mean color vector of a region in the RGB space as the colorimetric feature [4]. They assumed that if two regions strongly differ on one of the features, they are not similar. The property is called as absorbing property [4]. The shape similarity measures benefits from absorbing property. They used the region’s oriented bounding box (OBB) properties to characterize a region shape [4]. The approach developed is based on matching of region adjacency graphs (RAG) of pre-segmented objects [4]. The region features (texture, color, shape) are not strongly relevant due to the resolution. The problem of object matching can be expressed in terms of directed acyclic graph (DAG) matching [5]. The mainstream of the presented feature extraction algorithms decide on one or more key frames as being representative of each shot. Feature extraction techniques such as Wavelets or Gabor filters are extensively used to extract the features from these frames. In the final stage, the shot features are grouped into clusters to correspond to relevant objects which appear in those shots. In a query problem, features selected from the query region will be compared with existing cluster features for possible matches [6].

Fourier descriptors, compactness, and eccentricity features are commonly used for shape retrieval in the content based image retrieval systems [7] [8]. A new method for object based image retrieval, efficient subimage retrieval (ESR) is introduced by [9].

An effective video mining technique was described by Missaoui et al. in [10]. Their paper is dedicated to revisiting image and video mining techniques from the perspective of image modeling approaches, which amount to the theoretical basis for these techniques. The most important areas belonging to image or video mining are: image knowledge extraction, content-based image retrieval, video retrieval, video sequence analysis, change detection, model learning, as well as object recognition. Conventionally, these areas have been developed independently, and hence have not benefited from some common sense approaches which provide potentially optimal and time-efficient solutions. Two different types of input data for knowledge extraction from an image collection or video sequences are considered: original image or symbolic (model) description of the image. Several basic models are described briefly and compared with each other in order to find effective solutions for the image and video mining problems. They include feature-based models and object-related structural models for the representation of spatial and temporal entities (objects, scenes or events).

Fang et al. in [11] projected a fuzzy logic approach for detection of video shot boundaries. In their paper, they proposed a fuzzy logic approach to put together hybrid features for detecting shot boundaries inside general videos. These features include color histogram intersection, motion compensation, texture change, and edge variances. The fuzzy logic approach contains two processing modes, where one is dedicated to detection of abrupt shot cuts including those short dissolved shots, and the other for detection of gradual shot cuts. These two modes are unified by a mode-selector to decide which mode the scheme should work on in order to achieve the best possible detection performances. The advantages of their contribution can be highlighted as: (i) a range of features can be integrated by fuzzy logic operation to exploit their individual strength collectively; and (ii) while directly thresholding features remains sensitive to noises, selecting threshold in fuzzy domain provides a buffered operation and thus makes the detection more reliable. Experimental results support that the proposed algorithm is effective in video segmentation benchmarked by three existing algorithms and measured by precision and recall rates.

III. SEGMENTATION

In computer vision, segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels, also known as superpixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. An effective video segmentation requires proper feature selection and an appropriate distance measure. Different features and homogeneity criteria generally lead to different segmentations of the same video, for example, color, texture, or motion segmentation. Depending upon the application specific video segmentation methods should be considered. Factors that affect the segmentation method include the following:

Real-time performance: If segmentation must be performed in real time, for example, for rate control in videotelephony, then simple algorithms that are fully automatic must be used. Semiautomatic, interactive algorithms can be used for off-line applications such as video indexing or off-line video coding to obtain semantically meaningful segmentations.

Precision of segmentation: If segmentation is needed for object-based video authoring/editing or shape similarity matching, then it is important that the estimated boundaries align with actual object boundaries perfectly.

Scene complexity: Complexity of video content can be modeled in terms of amount of camera motion, color and texture uniformity within objects, contrast between objects, smoothness of motion of objects, objects entering and leaving the scene, regularity of object shape along the temporal dimension, frequency of cuts and special effects, etc.

In this paper the regions are segmented by using Seeded Region Growing algorithm. The seed image is a partly segmented image which contains uniquely labeled regions (the seeds) and unlabeled pixels (the candidates, label 0). Seed regions can be as large as and as small as one pixel. If there are no candidates, the algorithm will simply copy the seed image into the output image. Otherwise it will aggregate the candidates into the existing regions so that a cost function is minimized. Candidates are taken from the neighborhood of the already assigned pixels, where the type of neighborhood is determined by parameter neighborhood which can take the values FourNeighborCode() (the default) or EightNeighborCode(). The algorithm basically works as follows:

1. Find all candidate pixels that are 4-adjacent to a seed region. Calculate the cost for aggregating each candidate into its adjacent region and put the candidates into a priority queue.

2. While (priority queue is not empty and termination criterion is not fulfilled)

1. Take the candidate with least cost from the queue. If it has not already been merged, merge it with it's adjacent region.

2. Put all candidates that are 4-adjacent to the pixel just processed into the priority queue.

The color segmentation by using the Enhanced Hue Saturation Value (EHSV) color space. This transformation is used to find a pixel classification allowing to separate the pixels with a significant color from the pixels for which their color is not obvious and rather similar to a grey level. Once such a classification is achieved, proceeds to an enhancement of the "real" colors and of the grey levels. The chosen color space is the HSV space and this space is splitted into two distinct regions.

Figure 2. Multimedia mining process

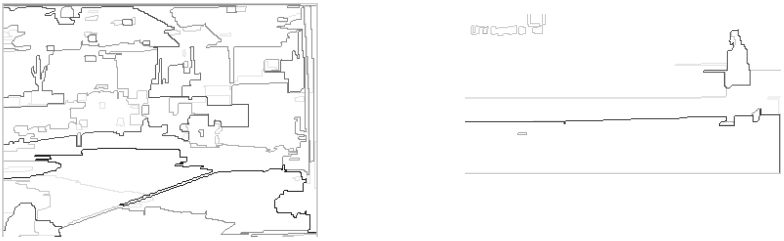

Figure 3. Extracted Objects

IV. FEATURE EXTRACTION

In order to define similarity between graphs, set of features are used to describe an object. Most of the approaches consider color as the major querying feature. Shape and Texture are the other usual features in the Content Based Image Retrieval (CBIR) framework. In this paper the object is retrieved based on the similarity of color, shape and region.

Color Region Similarity

The most frequently used feature for querying and retrieving multimedia data is color. Color histograms are common tools used in video retrieval. The similarity distance between two color image regions A and B is calculated as a weighted sum of the distances between the luminance L and the chrominance descriptors Ci = 1, 2, ….K. While the distance DL measures the similarity of shape, the distance DC evaluates the matching of the spatial distribution of a specific dominant

$$D_{CART}\left(A,B\right) = \alpha D\left(L\left(A\right), L\left(B\right)\right) + \sum_{i=1}^{K} \beta_i D\left(Ci\left(A\right), Ci\left(B\right)\right)$$

where $Ci(A)$ is the binary image associated to the ith dominant color of A and $\alpha + \sum_{i=1}^{K} \beta_i$ =1 .

Shape Similarity

The shape deformation feature vector of a video object is retrieved using Angular Circular Local Motion (ACLM) descriptor. The ACLM descriptor is computed by dividing a video object into M angular and N circular segments and computing the variances, σ2n,m , for each angular circular segment (n,m), in the temporal direction using the Video Object Planes (VOPs) of the video objects as follows:

$$\sigma^2_{n,m} = \frac{1}{S\left(n,m\right)^2 K} \sum_{k=0}^{K-1} \left(P_{n,m} - \mu_{n,m}\right)^2$$

$$\mu_{n,m} = \frac{1}{K} \sum_{k=0}^{k-1} P_{n,m}$$

$$P_{n,m} = \sum_{\theta=\theta m}^{\theta_{m+1}} \sum_{\rho=\rho_n}^{\rho_{n+1}} VOP_k(\rho, \theta)$$

where K is the number of the VOPs of the video object, VOPk is the binary shape map of the video object at kth instant, VOPk (ρ ,θ) is the value of the binary shape mask in VOPk at the position (θ ,ρ) in the polar coordinate system centered at the mass center of VOPk. S (n, m) is the area, θm is the start angle, and ρn is the start radius of the angular circular segment (n, m). They are defined by

$$S(n,m) = \frac{\pi\left(\rho_{n+1}^2 - \rho_n^2\right)}{M}$$

$$\theta_m = m \times \frac{2\pi}{M} \quad , \qquad \rho_n = n \times \frac{\rho_{\max}}{N} \quad , \qquad \rho_{\max} = \max_{VOP_k \in VO} \left\{\rho_{VOP_k}\right\}$$

where M and N are the number of angular and circular sections, respectively, VOPk is the kth video object plane of the video object, and θVOPk is the radius of the tightest circle around VOPk that is centered at the mass center of VOPk. The local motion feature matrix R is formed by σ2n,m, where σ2n,m is the normalized variance of the pixels that fall into the segment (n,m). The feature matrix R is re-ordered so that the angular segment with the largest variance is in the first column of R. This is achieved by first summing the columns of R to obtain the 1×M projection vector →A and then finding the maximum element of A, which corresponds to the angular segment m that has the largest variance. Then, the left the columns of R are circularly shifted by m to obtain a rotation invariant feature matrix.

V. Experimental Results

The database consists of two videos, first the videos are segmented into frames. The objects are retrieved based on similar frames and also the object is retrieved based on the color, region, and shape similarity for moving object. The performance is evaluated in the context of query by example. A retrieved object is considered a correct match if it represents the same object as the query.

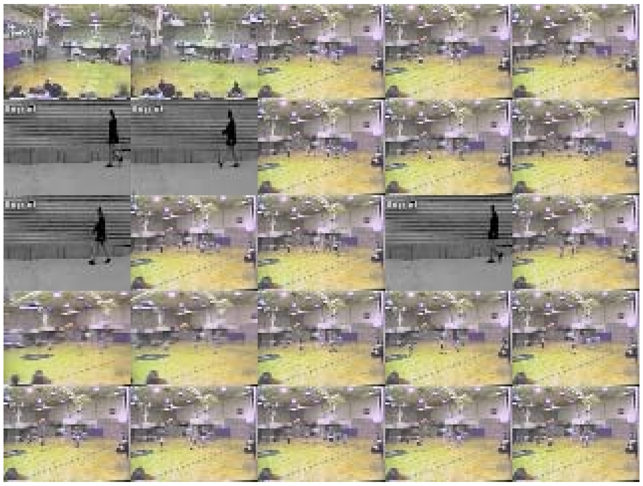

Figure 4. Retrieval of Objects Based on Color, Shape, and Region Similarity

VI. Conclusions

Thus in this paper we have presented a new approach for the retrieval of object from videos based on similarities of color, region, and shape. In the case of a large database, a filtering step based on global features of objects such as global histogram can be combined with our matching method in order to reduce the number of objects to test. In future this may be expanded to retrieve the video objects based on color, shape, region using fuzzy.

ACKNOWLEDGMENT

The authors thank the management of Karpagam University for their kind support to do this article successfully.

REFERENCES

[1] Yoshitaka A. and Ichikawa T., A survey on content-based retrieval for multimedia databases. IEEE Trans.on Knowledge and Data Engineering, Vol 11, 1999, pp. 81-93.

[2] S.Jeannin, “MPEG-7 Visual part of experimentation Model Version 9.0”,ISO/IEC JTC1/SC29/WG11/N3914, Mpeg Meeting, Pisa, Italy, Jan. 2001.

[3] S.V.Porter, “Video segmentation and indexing using motion estiomation,” Ph.D. Thesis, University of Bristol, Bristol, 2003.

[4] F. Chevalier, J.P. Domenger, J. Benois-Pineau, and M. Delest, “Retrieval of objects in video by similarity based on graph matching”, Pattern Recognition Letters, 2007.

[5] F. Chevalier, J.P. Domenger, and M. Delest, “A Heuristic for the Retrieval of Objects in Low resolution Video”, An international Workshop on Content Based Multimedia Indexing, 2007, conducted by IEEE.

[6] A.Anjulan, and N.Canagarajah, “Object based video retrieval with local region tracking”, Elsevier, Signal Processing Image Communication “, Vol 22, pp.607-621, 2007.

[7] Eakins, J.P., Boardman, J.M., and Shields, K., “Retrieval of trade mark images by shape feature – the ARTISAN project”, Proceedings of International Conference on Electronic Library and Visual Information Research, pp. 101-109, 1994.

[8] Wang, J., Yang, W., and Acharya, R., “Efficient access to and retrieval from a shape image database”, IEEE Workshop on Content-Based Access to Image & Video Libraries, pp. 63-67, 1998.

[9] Christoph H. Lampert, “Detecting Objects in Large Image Collections and Videos by Efficient Subimage Retrieval”, International Conference on Computer Vision (ICCV), Kyoto, Japan, 2009.

[10] Rokia Missaoui and Roman M. Palenichka, “Effective image and video mining: an overview of model-based approaches,” Proceedings of the 6th international workshop on Multimedia data mining: mining integrated media and complex data, pp. 43-52, 2005.

[11] Hui Fang, Jianmin Jiang and Yue Feng, “A fuzzy logic approach for detection of video shot boundaries,” Journal of Pattern Recognition Society, Elsevier, vol. 39, pp. 2092-2100, 2006.