A Fuzzy Theoretic Approach for Semantic Characterization of Video Sequences

R.S. Jadon, CSE Dept., IIT Delhi, New Delhi-16, rsjadon@cse.iitd.ernet.in

Santanu Chaudhury, EE Dept., IIT Delhi, New Delhi-16, santanuc@ee.iitd.ernet.in

K.K. Biswas, CSE Dept., IIT Delhi, New Delhi-16, kkb@cse.iitd.ernet.in

Asma Shakil, EE Dept., IIT Delhi, New Delhi-16, asmashakil@hotmail.com

Abstract

Most of the existing video processing systems consider the problem of classifying the video sequences on the basis of syntactic features alone. These similarity measures may be adequate only if the goal is to find frames with similar distribution of color, segmented regions or other low-level characteristics. We have worked on classification as a mechanism to support for indexing video sequences based upon intrinsic features of video data and/or their semantic content. First we segment video sequences into shots using a fuzzy framework. An opportunistic scheme is then used for processing the video sequences. Accordingly either mosaic-based or frame-based representation is used. We also compute various syntactic features like nature of cuts, camera motion, object motion and dynamics of scene. These features are evaluated for well defined domains using fuzzy classifier. Domain-based semantic modeling and heuristic based object recognition are used to semantically define and recognize video sequences. Experimental analysis illustrates the effectiveness of our system in offering a novel approach for video classification.

1 INTRODUCTION

Collection of digital images and video have become sources of useful information. Content-based technology for indexing and retrieval is then required to effectively use these digital repositories. The central problem in creating a robust and extensible system for processing video information lies in representing and visualizing video content. Due the to recent growth of information and entertainment delivery services based on direct broadcasting satellites and the future prospects for video delivery by data networks such as Internet, applications like interactive television, personalized-news and video-on-demand will become reality very soon. In such an environment, the task of finding the programs that best suit one's own interests will become critical, requiring assistance from smart automated information and entertainment appliances.

Such applications can however, become a reality only if robust methods are developed for content characterization, browsing, filtering and sorting. The first step for any type of video processing is shot segmentation. The segmentation boundaries can be either abrupt transitions or gradual transitions. Zhang et al.[3] proposed a threshold method working on difference of color histograms. Hamparur et. al.[6] present a model-based method, most suited to well established video processes. In our earlier work[5], a fuzzy theoretic approach has been proposed for video segmentation. This scheme provides a mechanism for characterizing scene transitions through soft-decisions. Subjectivity involved in the video partitioning process can be adequately captured by this framework. This algorithm can clearly distinguish between abrupt and gradual transitions. In the present work, we have proposed a scheme for fuzzy logic based semantic characterisation of video sequences which makes use of the result of fuzzy segmentation.

We are interested in the design of feature-spaces which can capture semantics of video sequences in terms of dynamics of scenes and presence or absence of an object. A fuzzy theoretic specification of image based features and spatial relations between image components has the ability to accomodate variations inherent in the different categories of the video sequences. Here, we have explored the use of fuzzy features and fuzzy rules for semantic categorisation of video sequences. We have also proposed the use of an evolutionary learning algorithm to learn rules from a given set of examples. Use of a learning algorithm has made the system adaptive.

For content characterization and categorization of video an appropriate scheme for representing video sequences is required. We have adopted an opportunistic approach for this purpose. For general camera shots, a frame-based-representation is used while for panning shots a mosaic-based representation is used. To characterize non-panning sequences we have developed a scheme for the fuzzification of spatial relationships that are present between the objects across different frames. To achieve this we have computed the spatial-configuration change, and fuzzified the resulting values. The overall scheme is represented in figure[1].

Rest of the paper is organized as follows: In section 2 we discuss camera motion detection scheme. In section 3 mosaic based video characterization is discussed. Section 4 contains the video classification based on scene dynamics. The learning paradigm is discussed in section 5. Experimental results are given in section 5 and finally we conclude in section 6.

2 CAMERA MOTION DETECTION

Camera motion is detected using optical flow computation for sequences in which gradual transitions have been reported. Optical flow computation involves representing the difference between two consecutive frames as a set of motion vectors. During a camera pan these vectors will predominantly have the same direction. In the subsequent subsections we will discuss the optical flow computation and fuzzification for camera motion detection.

2.1 Optical Flow Computation

Horn and Shunk's[4] algorithm has been used for optical flow computation. It gives motion vectors u and v in X-direction and Y-direction respectively. The direction of motion at each pixel in the image is obtained by computing the phase value as follows:

$$\theta = atan(u/v)$$

We then compute the histogram of these phase values, over the range $-\pi/2$ to $+\pi/2$. The bins with higher values provide an estimation of the direction of dominant motion. The quantification for histogram values is done using fuzzy-logic.

2.2 Fuzzification of camera motion

The histogram values are fuzzified for the purpose of camera motion estimation. We first identify three fuzzy labels for histogram values; small, large and very large. The fuzzy boundaries for these classes are determined statistically as a percentage of the total number of pixels. The basic heuristic used is that in a panning sequence large number of pixels move in one direction.

The histogram of the phase value for motion vectors suggests the dominant direction. In exact-reasoning, the problem lies in estimating the number of pixels moving in one direction which can significantly quantify the camera motion. The difficulty exists due to two reasons, first the motion vector computation is highly sensitive to noise, and second none of the existing algorithms for optical flow computation give exact estimation of motion vectors. Thus there is always an inherent fuzziness involved in determining the camera motion. In the present work this uncertainty is modeled using fuzzy inferencing. Once the fuzzy labels are computed for histogram values of the phase of motion, the set of fuzzy-rules can be designed to detect panning sequences. Some of the typical rules for this purpose are as follows:

- *Rule 2.1: If LARGE-NUMBER of histogram values are SMALL and SMALL NUMBER of histogram values are LARGE then it is NO-PANNING sequence.*

- *Rule 2.2: If LARGE-NUMBER number of histogram values are SMALL and SMALL-NUMBER of histogram values are VERY-LARGE then it is PANNING SEQUENCE.*

Thus panning sequences can be identified with certain membership value. A high membership value is a positive indication of panning.

3 MOSAIC BASED VIDEO CHARACTERIZATION

A mosaic is a collection of overlapping images put together with coordinate transforms relating the different image coordinate systems. The video information distributed over many frames at the cost of very high temporal redundancy is converted into compact representation in terms of mosaics. A number of techniques are available in the literature for creation of mosaics such as [9]. Mosaics [1] can represent video information given that the constraints of camera model, motion and lightness changes are accounted for.

3.1 Content Extraction from Mosaics

For the purpose of identification of *Mosaics*, we statistically characterize different regions with different key domain-independent parameters covering shape, color and location. This task is accomplished by calculating these features so that a given image or a mosaic can be effectively classified or categorized. We have developed a segmentation algorithm based on the hue, saturation and intensity (HSI) color space for segmenting the mosaics. This is because of the advantage of

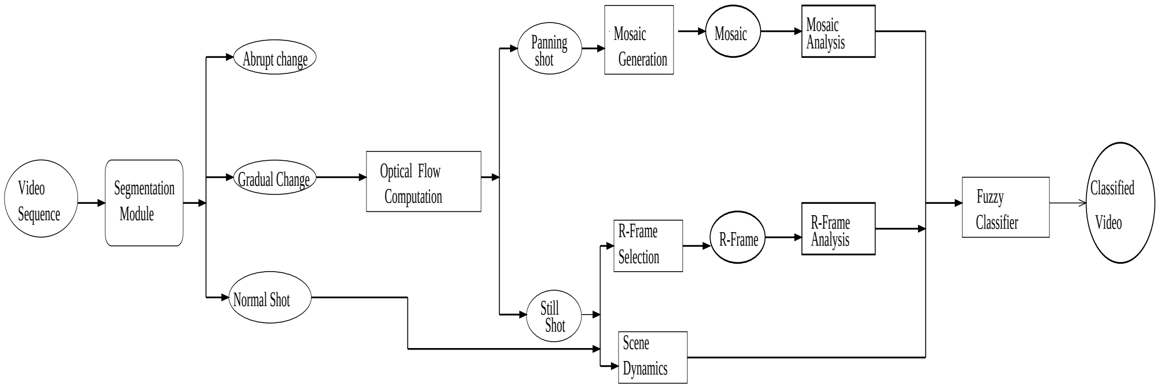
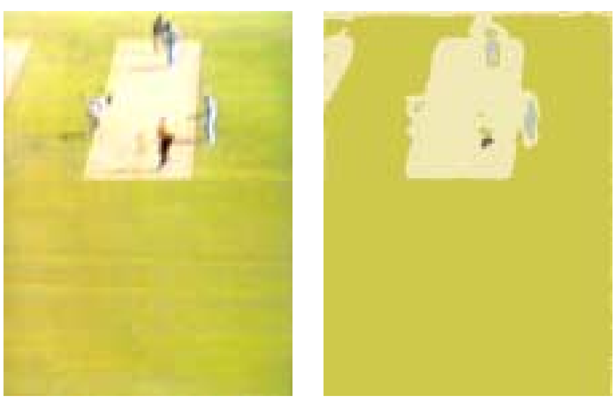


Figure 1: An Overview of Video Characterization System

the HSI color space in characterizing colors; for example the distinction in shades of green like light green, dark green, pale green etc. can be done by considering only hue values. Segmentation is done by transformation from RGB to HSI space through a quantization and thereby grouping the regions falling in the similar hue, saturation and intensity values. Our segmentation approach is somewhat similar to the technique described in [8]. The color match similarity metric is also taken as described in [8]. Some segmentation results are as shown in Figure 2 for cricket mosaic.



(a) (b)

Figure 2: Segmentation results: (a)Original cricket mosaic, (b) Segmentation using using HSI model

Then several region based features like centroid, area, perimeter, moment invariants, etc. are extracted for further analysis.

3.2 Fuzzy Attributes

For categorizing a sports video sequence, different fuzzy variables like green score, pitch score, yellow color match, rectangular shape match, man score, elliptical score, surroundedness in grass etc. are considered. We have discussed the extraction and representation of such attributes in detail in [2].

3.3 Development of Fuzzy Rules for Characterization

The motivation for fuzzy rules can be in the terms of human perception of knowledge. There is a need to represent this knowledge based on our past experiences and skill. For example a cricket scene may consist of a field with a certain degree of green color, a pitch somewhat rectangular in shape (probably slight yellowish-brown in color and surrounded by grass), and some players in the ground. Such knowledge can be used for further rule evaluation and classification of mosaics or video. For the current approach we have employed several rules as mentioned in [2].

4 VIDEO CHARACTERIZATION BASED ON SCENE DYNAMICS

Dynamic changes of object positions also provide an important clue for video characterization. For example in a football game, the changes will be much more than in a game of cricket. Since panaromic mosaics cannot capture the dynamic information content of the video we propose to combine frame based attributes with mosaic based information for fuzzy categorization. For example for sports video sequences like football and cricket, the relative motion of the players can be estimated.

Knowing the object geometry, it is possible to precompute spatial and visibility relationships that are present in the subsequent frames. For computing object shape feature the set of invariant moments are computed. In the current approach we are basically constructing a polygon to represent the configuration of players in a sequence of video frames to get an overall shape representation. Each vertex of the polygon represents a player. Next, the error in variation of moments of the polygon shape change is computed. Thus for sequences with fast relative motion between the objects, the shape of the polygon changes and the amount of change is significant, while for rather static objects the amount of change is less.

The spatial configuration change has been fuzzified, so that it can be categorized as **no-change, a large slow change, a small slow change, a large fast change or a small fast change**. Following features are chosen for this purpose:

- Normalized range of difference in moment values so that it can be labeled as small, medium, large and very large.
- Normalized coefficient of variation of the difference in moment values which is labeled as being either small, medium, or large.

Some typical rules for video classification based on spatial configuration change are given below:

- *Rule4.1: If range is SMALL and coefficient of variation is SMALL then the sequence is UNCHANGED*
- *Rule4.2: If range is MEDIUM and coefficient of variation is SMALL then the sequence is UNCHANGED*
- *Rule4.3: If range is MEDIUM and coefficient of variance is MEDIUM then sequence is SLOW CHANGE*
- *Rule4.4: If range is LARGE and coefficient of variation is MEDIUM then the sequence is SLOW CHANGE*
- *Rule4.5: If range is large and coefficient of variation is LARGE then the sequence is "FAST CHANGE"*
- *Rule5.5: If range is VERY LARGE and coefficient of variation is LARGE then change is FAST CHANGE*

We further compute the difference between the first and last frame of the sequence, we call this the second moment difference. This difference is also fuzzified. Finally the following rules are used to determine the overall dynamics of the scene:

- *Rule5.6: If sequence is SLOW CHANGE and second difference is SMALL then the change is SMALL SLOW CHANGE*
- *Rule5.7: If sequence is FAST CHANGE and second difference is SMALL then the change is SMALL FAST CHANGE*

5 EVOLUTIONARY LEARNING OF RULES

Further, we have implemented an evolutionary learning algorithm which uses genetic algorithms as the means to evolve an automatic fuzzy system with the membership function shapes and types (i.e. the database), the fuzzy rule-set and the number of rules inside it[7]. We have used this evolutionary learning mechanism to evolve a Fuzzy Rule Based System(FRBS) to be used in the process of shot detection in the video sequences and a FRBS for the classification of mosaics and frames based on their content (as discussed in the previous sections). This approach of automating the fuzzy system design is particularly attractive in our context because it eliminates the need of incorporating human reasoning into designing the most optimum set of fuzzy rules, choosing appropriate membership function shapes and class boundaries which is the most tedious task in the present situation because of the large ambiguities in the data to be classified. The main steps in genetic algorithm based learning are:

- **Population of Potential Solutions**

  In our case the individuals in the genetic population are basically FRBSs. Thus each chromosome encodes the complete information of the number of rules in a fuzzy system, the membership function types and the class boundaries for each fuzzy label of each fuzzy variable and also the complete set of rules in that particular fuzzy system.

- **Set of Evolution Operators** The genetic operators of selection, crossover and mutation are applied appropriately to the encoded chromosome to evolve new generations.

- **Performance Index** The performance index we have chosen for the evaluation of the individuals in the current population consists of assigning to each individual (i.e. each FRBS) a reward in proportion to the number of patterns it can correctly classify out of a total of a number of training patterns fed to it.

Sport	Test Patterns	Hand Crafted Rules (No. correct)	Evolved FRBS (No. correct)
Cricket	17	15	16
Golf	7	7	6
Football	12	2	7

Table 1: Classification of Mosaics/Frames

Shots (Total 36)	Hand-Crafted Rules	The Evolved FRBS
Correct	11	31
Missed	4	0
Incorrect	21	5

Table 2: Comparison for Shot Detection

Finally the effectiveness of the evolved Fuzzy System is tested by using it for classifying a number of test patterns consisting of mosaics of different sports and also for shot detection, the results of which are tabulated in Tables[1],[2]. As can be seen from these tables, classification using the evolved Fuzzy System yields better results in most cases than that using hand-crafted fuzzy rules and membership functions.

6 EXPERIMENTAL RESULTS

The overall system is implemented on SGI(IRIX 6.3) Workstation employing Image Vision(il) Digital Media(dmedia) and X11 toolkit. We have experimented with nearly 100 sports clips.

6.1 Camera Motion Detection Results

As mentioned earlier the inputs for this module are obtained from segmentation module. We have computed the optical flow only for those subsequences, where gradual transitions are seen. Some of the results for camera motion detection are shown in Table[3]. As can be seen in the table for some of the sports video sequences the fuzzy membership value for panning is very high. Such sequences are the best candidates for mosaic-based characterization.

6.2 Scene Dynamics Results

Graphs characterising scene dynamics for football and cricket sequences are shown in Figure 3.

From the graphs it is clear that there is a large variation in the moment difference for the football sequence(frames) as compared to that of cricket. In other words we can say that,in general, there is a large variation associated with the football. These attributes have been modeled with the help of fuzzy rules as mentioned above.

File Name	PAN Membership	NON-PAN Membership
(cam1.Jpeg)	0.750811	0.110117
(cam2.Jpeg)	0.511601	0.210315
(foot1.Jpeg)	0.663001	0.233447
(hockey.Jpeg)	0.856714	0.011445
(educ.Jpeg)	0.113671	0.655542
(cric1.Jpeg)	0.655912	0.110231
(zee1.Jpeg)	0.011142	0.781113
(golf1.Jpeg)	0.766651	0.110211

Table 3: Results showing Fuzzy scores for Camera Motion Detection

S.No	File Name	Cricket	Golf	Football
1	(c1.ppm)	0.608050	0.466407	0.280638
2	(c2.ppm)	0.949325	0.230734	0.438150
3	(c3.ppm)	0.590072	0.283898	0.272341
4	(c4.ppm)	0.678893	0.329069	0.313321
5	(c5.ppm)	0.907895	0.397286	0.483226
6	(c6.ppm)	0.604081	0.452852	0.278806
7	(c7.ppm)	0.419162	0.452209	0.193460
8	(g1.ppm)	0.354282	0.783548	0.163515
9	(g2.ppm)	0.510706	0.783637	0.441208
10	(fb3.ppm)	0.399765	0.556414	0.184507

Table 4: Results showing fuzzy scores of mosaics for Cricket,Golf and Football sequences with Hand-Crafted Rules

S.No	File Name	Cricket	Golf	Football
1	(c1.ppm)	1.000000	0.890966	0.679438
2	(c2.ppm)	1.000000	0.327932	0.583780
3	(c3.ppm)	1.000000	0.438391	0.168582
4	(c4.ppm)	1.000000	0.532265	0.270754
5	(c5.ppm)	1.000000	0.788140	0.577267
6	(c6.ppm)	1.000000	0.789475	0.577267
7	(c7.ppm)	1.000000	0.877481	0.911877
8	(g1.ppm)	0.729749	1.000000	0.270754
9	(g2.ppm)	0.729579	1.000000	0.372925
10	(fb3.ppm)	1.000000	1.000000	0.883780

Table 5: Results showing fuzzy scores of mosaics for Cricket,Golf and Football sequences with Evolved FRBS

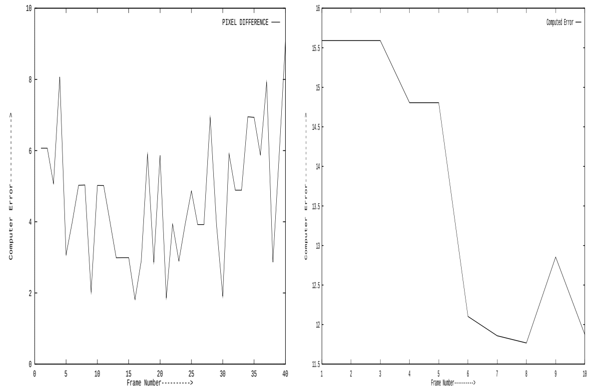


Figure 3: Graph showing error for (a)Football sequence (b)cricket sequence

6.3 Overall Classification Results

In this section we shall analyse the effectiveness of our fuzzy classification scheme and validity of our learning algorithm. The classification and shot detection results of the evolved fuzzy systems for video of different sports are presented in Tables[1],[2]. As can be seen from these tables, classification using the evolved Fuzzy System yields better results in most cases than that using hand-crafted fuzzy rules and membership functions.

The fuzzy membership scores of mosaics of some sequences obtained by our classification scheme are as shown in tables 4 and 5. Entries 1-7 are cricket mosaics, 8,9 are golf mosaics and 10 is a football sequence. These membership values show that with evolved fuzzy rule based system we have obtained better disambiguation even though misclassification errors have not been eliminated. In terms of overall statistics, using a reasonable set of examples for learning, we have achieved an overall accuracy of around 85% using evolutionary FRBS, while its only 75% for hand-crafted rules.

7 CONCLUSION

In this paper we have proposed a generic fuzzy scheme for characterizing video sequences. Starting with the problem of segmenting the video data we have suggested schemes upto content based classification of video data. We have primarily focussed on domain-based video categorization problems and chosen the sports video. Large video database is populated for experimentation purpose. From the results of the fuzzy classification it is clear that the categorization and classification of the sequences is done effectively. The approach for the spatial configuration change can be further employed as a frame based approach along with mosaics as an aid to represent games like football, volleyball etc.

References

[1] P. Anandan and M. Irani. Video Indexing Based on Mosaic Representations. In *Proceedings of IEEE*, volume 86, pages 905–921, May 1998.

[2] Pranav Bhushan, R.S. Jadon, and Santanu Chaudhury. Mosaic Based Characterization Of Video Sequences Using Fuzzy Inferencing. In *Proceedings of International Conference on Multimedia and Systems*, pages 65–69, Auguest 13-15 2000.

[3] A. Kankanhalli H.J. Zhang and S.W. Smoliar. Automatic Partitioning of Full-motion Video. *Multimedia Systems*, 1(1):10–28, 1993.

[4] B.K.P. Horn and B.G. Schunk. *Robot Vision*. MIT PRESS, 1986.

[5] R.S. Jadon, Santanu Chaudhury, and K.K. Biswas. Video Segmentation Using Fuzzy Logic. In *Proc. of the Indian Conference on Computer Vision, Graphics and Image Processing(ICVGIP'98), IIT Delhi, Dec'98*, pages 364–370, 21-23 Dec 1998.

[6] A. Hampapur R. Jain and T. Weymouth. Production Model Based Digital Video Segmentation. *Journal of Multimedia Tools and Applications*, pages 9–46, March 1995.

[7] Yuhui Shi, Russell Eberhart, and Yaobin Chen. Implementation of Evolutionary Fuzzy Systems. *IEEE Transactions on Fuzzy Systems*, 7(2), April 1999.

[8] J.R. Smith and Shih-Fu Chang. VisualSEEK: a fully automated content-based image query system. *ACM Multimedia*, November 1996.

[9] Richard Szeliski. Video Mosaics for Virtual Environments. *IEEE Computer Graphics and Applications*, 16(2):22–30, March 1996.