Sports Video Characterization Using Scene Dynamics

R.S. Jadon
MCA Deptt.
MITS Gwalior
rsjadon@hotmail.com

Santanu Chaudhury
EE Deptt.
IIT Delhi
santanuc@ee.iitd.ac.in

K.K. Biswas
C.S.E. Deptt
IIT Delhi
kkb@cse.iitd.ac.in

Abstract

Dynamic changes of object positions provide an important clue for video characterization. In the present work, we exploit the dynamic information present over different frames of a sports video to characterize the change in the configuration of players across different frames. For scene dynamic characterization firstly location of players are detected by using motion based segmentation. We then construct a polygon with the players placed at the vertices of the polygon. Next the change in the shape of the polygon across different frames is computed in terms of the difference in the moments of the polygon shape. The difference is computed for seven moment invariants. The mean of these difference values, called mean-difference, is used as the key feature for characterizing the scene dynamics. An evolutionary learning based fuzzy rule based system is developed for characterizing sports sequences using duifference values.

1. Introduction

Human beings are being able to categorize sequences based on number of cues such as background music, voices and many kinds of visual cues. This is particularly true of sports video, where a view of goal post, cricket pitch, badminton court or green field is enough for categorization. These are spatial features which can be extracted from key frames. However, domain knowledge based temporal features are also equally important for characterizing sports sequences. Different approaches have been suggested in the past for semantic categorization of video sequences using domain knowledge. The features extracted for this purpose are task specific. Zhang et al. [11] have used models of specific types of programs such as TV news. Model based or knowledge based approaches have been also proposed for analysis of sports video like soccer [5] and tennis [10]. Use of a rule based system has been proposed in [12] for categorization of basketball video into different classes using visual and motion based characteristic features. They have used an inductive decision-tree learning method to arrive at a set of if-then rules. An event based indexing scheme for broadcast sports video by combining visual, auditory and textual (closed caption) cues has been proposed in [3]. Assfalg et al. [2] have proposed a method for semantic annotation of sports video by combining several low level visual primitives. Recently Chen Jianyun et al. [9] have suggested a unified framework for semantic content analysis of sports video using basic semantic units (BSU). Most of the approaches suggested in literature are based on low level visual or audio features.

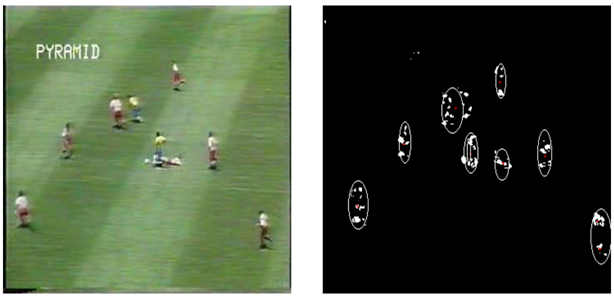
In the present work we have proposed a novel approach for characterizing sports video using semantic interpretation of scene dynamics. For sports video sequences like football and cricket, the relative motion of the players can be estimated. In the fotball game for example the relative motion will be much more than in the game of cricket. Knowing the object geometry, it is possible to precompute spatial and visibility relationtionships that are present in the subsequent frames. 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 the variation of moments of polygon shape change is computed. Thus the 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. Based on this knowledge, a fuzzy framework is designed in which the spatial configuration change is fuzzified so that it can be categorized as no-change, a large slow change, a small slow change and large fast change or small fast change. A set of fuzzy rules is than evolved using genetic algorithm based learning [7] for characterizing cricket and football sequences. Rest of the paper is organized as follows: We briefly describe the motion based segmentation technique in section 2. In section 3 the scene dynamics characterization scheme is discussed. Experimental results are given in section 4. Finally we conclude in section 5.

2. Motion Based Segmentation for Player Detection

Motion based segmentation is an integral part of many image sequence analysis problems like 3-D motion and structure estimation [1] ans scene segmentation [4]. In the present work we have used optical flow based motion segmentation. In addition to the color and texture information available in still images, a video sequence provides temporal information. While it is hard to extract semantically meaningful objects based only on color and texture cues, motion cues facilitate segregation of objects from the background. For detecting some of the dynamic objects like players in the field, motion-based segmentation turns out to be more effective as compared to color based segmentation. We have used simple optic flow based segmentation. We first compute the motion vectors u and v between a pair of frames using Horn and Shunk's optical flow computation [6]. We then compute the magnitude of motion as follows:

$$\mu_{i,j} = \sqrt{u_{i,j}^2 + v_{i,j}^2}$$

where $\mu_{i,j}$ is the magnitude at $(i,j)^{th}$ location. A threshold is then applied for detecting the regions with high motion content in a frame. The motion induced by camera movement is small as compared to the movement of players, thus an appropriate threshold can give motion based segmentation regions correctly. Motion based segmentation for a football video sequence is shown in Fig.1. The detected regions are explicitly marked with oval shape figures in Fig.1(b), each one corresponding to a player in Fig.1(a).



(a) (b)

(a)Original-Frame (b)Segmented Frame

Figure 1: Motion Based Segmentation: Players Detected are marked with elliptical regions

3. Scene Dynamics Characterization

Dynamic changes of object positions is an important feature for video characterization. In the present work, we exploit the dynamic information present over different frames of a sports video to characterize the change in the configuration of players across different frames. For scene dynamic characterization firstly location of players are detected as explained earlier. We then construct a polygon with the players placed at the vertices of the polygon. As the players move in the field, the topology of the polygon changes. The change in the shape of the polygon across different frames is computed in terms of the difference in the moments of the polygon shape. The difference is computed for seven moment invariants. The mean of these difference values, called mean-difference, is used as the key feature for characterizing the scene dynamics. This phenomenon of polygonal shape change is illustrated for some of the consecutive frames for a football sequence in Fig. 2. For sequences with fast relative motion between the players (e.g. football sequences) the shape of the polygon changes rapidly while the change is slow for rather sedate games (like cricket).

Motivated by this heuristic we fuzzify the **spatial configuration change** so that it can be categorized as a **'no-change' or 'a large slow change' or 'a small slow change' or 'a large fast change' or 'a small fast change'**. The following features are chosen for this purpose:

- Normalized range of mean-difference values so that it can be labeled as very small, small, medium, large and very large.

- Normalized coefficient of variation of the mean-difference values which is labeled as being either small, medium, or large.

Normalized range is calculated as the difference between the largest and smallest normalized mean-difference values. A normalized-mean-difference value is computed by dividing the mean-difference value with the maximum of mean-difference values for a particular sequence. Similarly, Normalized coefficient of variation is calculated as the coefficient of variation of normalized-mean-difference values. Some typical rules for characterizing the video sequences based on spatial configuration change are given below:

- *Rule1: If normalized-range is SMALL and normalized-coefficient-of-variation is SMALL then the sequence is UNCHANGED*
- *Rule2: If normalized-range is MEDIUM and normalized-coefficient-of-variation is SMALL then the sequence is UNCHANGED*
- *Rule3: If normalized-range is MEDIUM and normalized-coefficient-of-variation is MEDIUM then sequence is SLOW CHANGE*

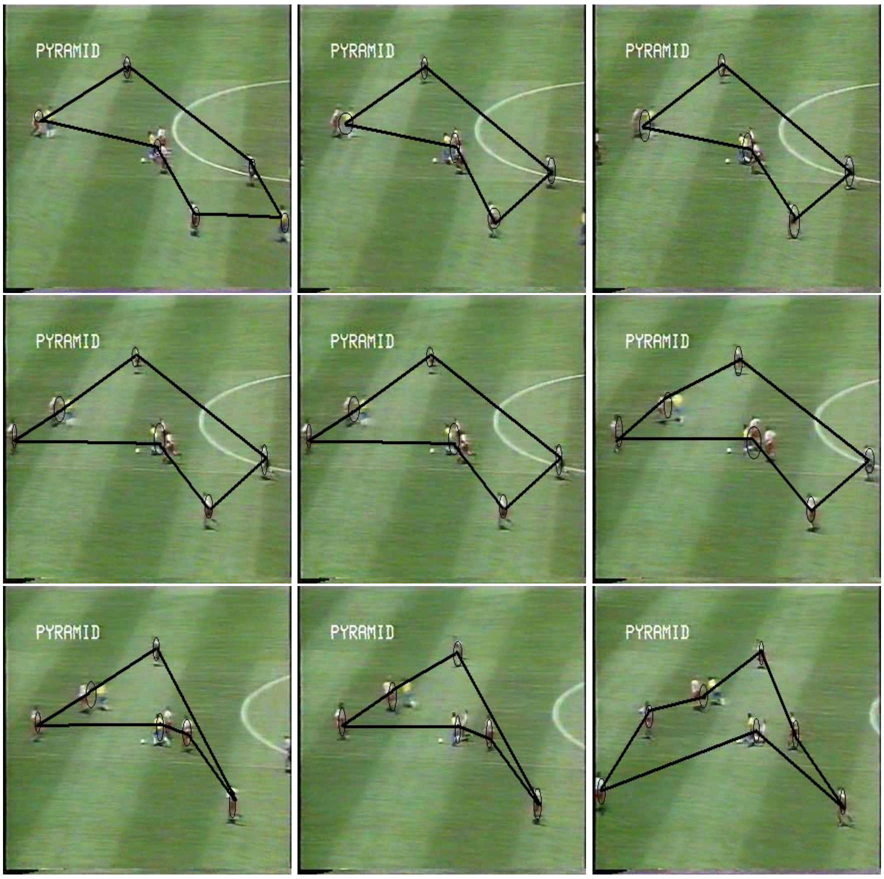


Figure 2: Polygonal Shape Change for a Football Sequence

- *Rule4: If normalized-range is LARGE and normalized-coefficient-of-variation is MEDIUM then the sequence is SLOW CHANGE*
- *Rule5: If normalized-range is large and normalized-coefficient-of-variation is LARGE then the sequence is FAST CHANGE*
- *Rule6: If normalized-range is VERY LARGE and normalized-coefficient-of-variation is LARGE then change is FAST CHANGE*

We further compute the mean-difference value between the first and last frame of the sequence, we call this cumulative difference. This is also fuzzified. Finally rules are evolved for characterizing the scene dynamics using the cumulative difference and spatial configuration change. Some of the typical rules evolved for final characterization are given below

- *If SLOW-CHANGE is Unchanged and CUMULATIVE-DIFFERENCE is Small then NO-CHANGE*
- *If SLOW-CHANGE is Large and CUMULATIVE-DIFFERENCE is Small then SMALL-SLOW-CHANGE*
- *If FAST-CHANGE is Large and CUMULATIVE-DIFFERENCE is Small then SMALL-FAST-CHANGE*

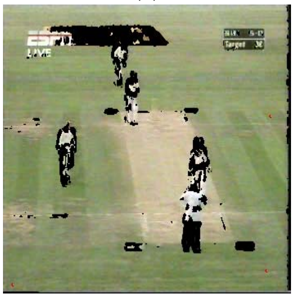
Scene dynamic analysis involves the study of spatial change of player positions in the field. Thus this analysis is highly domain dependent and sensitive to player identification. In the current-state-of art for object recognition like player detection with high accuracy is not possible. However, for our purpose just player location in the field is sufficient.

We first segment the video into shots using a fuzzy theoretic approach [8]. For each shot we compute the scene dynamic labels as: **no-Change, small-slow-change, small-fast-change, large-slow-change and large-fast change** as explained above. We then use a fuzzy rule based system for characterizing the video sequence using scene dynamics, some of the rules found are:

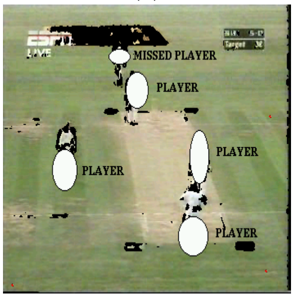
- If spatial-configuration-change is unchanged then it is cricket sequence.

(a)

(b)



(c)



(d)

(a)Frame1 (b)Frame2
(c)Motion Based Segmented Frame (d)Players Detected

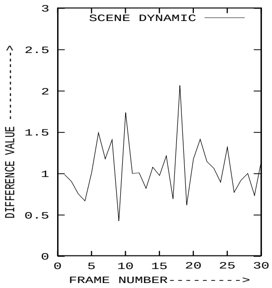
Figure 3: Player Detection Results

- If spatial-configuration-change is small-fast-change then it is football sequence.
- If spatial-configuration-change is small-slow-change then it is cricket sequence.
- If spatial-configuration-change is large-fast-change then it is football sequence.
- If spatial-configuration-change is large-slow-change then it is football sequence.

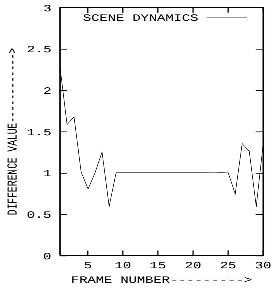
4. Experimental Results

We now present the results for scene dynamic computation. As mentioned earlier scene dynamics are computed from image sequences. A shot (as detected in the segmentation module) is an ideal candidate for this purpose. We therefore compute the scene dynamics for shots. Player detection results for a pair of consecutive frames for a cricket shot are illustrated in Fig. 3. Note, in Fig. 3(d) one player region not correctly characterized, is marked as **missed player** while correctly detected players are marked as **player**.

The graphs characterizing the scene dynamics for typical football and cricket shots are shown in Fig.4. From the graphs it is clear that there is a large variation in the moment difference values for the football sequence as compared to that of cricket. We observe that, in general, in a football sequence the spatial configuration change between players is more than in a cricket sequence.



(a)



(b)

Figure 4: Variation of difference values (a)Football (b)cricket

Table 1 shows the results of fuzzy analysis for scene dynamics for some of the typical video sequences. Note, that the spatial configuration change for football and sprinting sequence are always categorized as Large-Slow-Change or Large-Fast-Change as compared to cricket where it is either No-Change or Small-Slow-Change.

5. Conclusions

In this paper we have proposed a fuzzy theoretic approach for characterizing sports sequences using scene dynamic based features. This characterization is based on the domain knowledge. We have developed a fuzzy rule based system system for motion based segmentation and scene dynamics interpretation. This work can be extended for sematic interpretation of many other similar problem domains.

References

[1] G. Adiv. Determining three-dimentional motion and structure from optical flow generated by several moving objects. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 7:384–401, 1985.

[2] J. Assfalg, M. Bertini, C. Colombo, and A. D. Bimbo. Semantic annotation of sports video. *IEEE Multimedia*, pages 52–60, April-June 2002.

[3] N. Babaguchi, Yoshihiko, and T. Kitahashi. Event based indexing of broadcasted sports video by intermodal collaboration. *IEEE Tran. on Multimedia*, 4(1):68–75, March 2002.

[4] M. Chang, M. Sizan, and A. Tekalp. An algorithm for simultaneous motion estimation and scene segmentation. In *Proc. Int. Conf. ASSP, Adelaide, Australia*, April 1994.

[5] Y. Gong, C. Chuan, H.Zhang, and M. Sakauchi. Automatic Parsing of Tv Soccer Programs. *IEEE Transations on Multimedia*, pages 167–172, 1995.

[6] B. Horn and B. Shunk. *Robot Vision*. MIT PRESS, 1986.

Sequence Type	No-Change	Small-Slow Change	Small-Fast Change	Large-Slow Change	Large-Fast Change
Cricket	0.66	0.33	0.00	0.00	0.00
Cricket	0.60	0.37	0.07	0.00	0.00
Football	0.00	0.00	0.00	0.00	0.13
Football	0.00	0.00	0.00	0.000	0.68
Football	0.00	0.00	0.00	0.12	0.00
Sprinting	0.00	0.00	0.00	0.60	0.00

Table 1: Fuzzy Characterization Results for Scene Dynamics: Membership values for Spatial Configuration Change

[7] R. Jadon. *An Evolutionary Learning-Based Fuzzy Theoretic Approach for Characterizing Video Sequences*. PhD thesis, Deptt. Of Computer Science and Engg, IIT Delhi, October 2002.

[8] R. Jadon, S. Chaudhury, and K. Biswas. A Fuzzy Theoretic Approach for Video Segmentation Using Syntactic Features. *Pattern Recognition Letters*, 22:1359–1369, November 2001.

[9] C. Jianyan, L. Yunhao, L. Songyang, and W. Lingda. A unified framework for semantic content analysis in sports video. In *Proc. 2nd International Conf. on Information Technology for Application(ICITA 2004)*.

[10] G. Sudhir, J. Lee, and A. Jain. Automatic Classifiaction of Tennis Video for High-level Content-based Retrieval. In *Proc. IEEE International Workshop on CAIVD, Bombay, India*, 3 Jan 1998.

[11] H. Zhang, Y. Gong, S. Smoliar, and S. Y. Tan. Automatic parsing of news video. In *Proc. Of Int. Conf. On Multimedia Computing and Systems, Boston, Massachusetts, USA*, pages 45–54, May 1994.

[12] W. Zhou, A. Vellaikal, and C.-C. J. Kuo. Rule-based Video Classification System for Basketball Video Indexing. In *Proc. ACM Multimedia Workshop Marina Del Rey CA USA*, pages 213–216, 2000.