Using Fuzzy Logic for Event Detection in Soccer Video

Nguyen Ngoc Thanh, Le Ngoc Giao
Research Institute of Posts and Telecoms, Hanoi, Vietnam
e-mail: nndiep@hn.vnn.vn

Abstract

Video event detection has become an essential application in multimedia computing. For sports video, salient events are usually detected by analyzing video sequence by specific decision rules. However in many kinds of sports video (e.g. soccer), the game contains continuous actions, in which the boundaries of shots, scenes are uncertain. So the conventional analyzing methods using crisp decisions are not efficient. Fuzzy logic is a natural approach that can tackle this problem. In this paper, we present a new approach using fuzzy technique for event detection in soccer video. The experiment shows encouraging results for this method

I. INTRODUCTION

Recently, the great development of telecommunication media including TV broadcasting and Internet technology has made the explosion of broadcast information. The large amount of multimedia content causes the needs of abstraction generation that helps one can understand and retrieve interested contents quickly. Manual generation of video abstract is a tedious, time-consuming, and costly process. Therefore automatic tools for video abstract generation are very necessary.

In general, a video abstraction is constituted from salient events detected in a video. A general framework for event detection contains two main steps: extracting audiovisual features from raw video data and analyzing these features to find semantic video segments. Expert knowledge is usually used to bridge the gap between the low-level features and the semantic events of video.

Following this framework, many approaches have been suggested in literatures [1,2,3,4,5,6]. Most of these schemes are based on the crisp estimation of syntactic video features. However, due to variations in video conditions, crisp estimates do not always provide a robust mechanism. Moreover expert knowledge contains uncertain, incomplete and vague information that is very difficult to apply in crisp estimations. To overcome these problems we propose a soft-computing approach using fuzzy logic for video event detection. Goal-detection for soccer game video is performed to illustrate this approach.

The fuzzy set theory introduced by Zadeh [7] has been utilized in many areas of research, from control to database and expert system [8]. Recently fuzzy techniques are efficiently applied in the field of image processing. Fuzzy logic is a very efficient method to overcome the uncertain and insufficient conditions of data. Fuzzy logic offer powerful tools to represent and process human knowledge. Concept of fuzzy model can be used at two levels in the field of video processing. In the feature level it is used in representing an input pattern as an array of membership values denoting the degree of possession of certain properties and. In semantic level it is used in representing linguistic phrases (or knowledge) by fuzzy rules.

Our event-detection system for soccer video contains three main steps. In the first step, audiovisual features are extracted from video. Three features are selected from MPEG-7 standards for soccer video. They are Dominant Color, Camera Motion, and Audio Power descriptors. The Dominant Color descriptor is useful to detect objects in frames: field region, overlay block, and objects (player, referee). In the second step, these audiovisual features are represented in fuzzy model (feature fuzzification). In the last step, fuzzy rules based on knowledge are formed to combines fuzzied features and give decisions to detect specified events.

The rest of the paper is organized as follows. We briefly discuss audiovisual features used for soccer-event detection in section 2. Section 3 presents the goal-event detection system using fuzzy technique. It includes fuzzy conceptualization for numerical features and inference system. The initially experimental results are reported in section 4. Finally, section 5 presents the conclusions.

II. FEATURE EXTRACTION

Selection of appropriate features is the first step for processing video data. It is the critical issue, which decides the efficiency of the video processing system.

There are two kinds of feature: low-level features and high-level feature. The low level features are domain independent and extracted directly from video sequence. Low level features can be used in many domains and does not contain any semantic meaning about video. The high-level features depend on specific domain of video and are specified with the events that need to be detected.

2.1 Low-level features

For soccer game video, we choose three low-level features: color histogram, motion activities, and audio power. The definition and extraction method of these features follows MPEG-7 standards (Multimedia Content Description Interface) [9]. The following subsections will describe briefly these feature and their applications in soccer video domain.

2.1.1 Color Histogram

The Color Histogram is represented by Scalable Color descriptor in MPEG-7. It is described in MPEG-7 standard that the scalable color descriptor can be interpreted as a Haar transform-based encoding scheme applied across values of a color histogram in the HSV color space [9]. Through processing Color Histogram, we can get the following information:

- Color area represents grass color in a soccer video and the grass region in each frame of video. Algorithm to get this information is described in [3].
- Histogram difference (HD): A color-histogram difference value HD_{ij} represents the difference value of color components between frame i and j. HD_{ij} is computed using normalized histogram intersection as follows:

$$HD_{ij} = 1 - (1/3)\left[\begin{array}{l} \sum_{k=1}^{nH} \min\left(F_{Hk}^{i}, F_{Hk}^{j}\right) + \\ \sum_{k=1}^{nS} \min\left(F_{Sk}^{i}, F_{Sk}^{j}\right) + \\ \sum_{k=1}^{nV} \min\left(F_{Vk}^{i}, F_{Vk}^{j}\right) \end{array}\right]$$

where nH, nS, nV is the number bins in histograms of channels H, S, V respectively. F_{Hk}^{i} is the histogram value at bin k of channel-H of frame i. Similar terms are defined for V channel and S channel.

This measure ensures that for frames which are nearly similar, HD turns out to be close to zero, while for dissimilar frames HD is closer to one. HD is a feature representing for the changing speed of frames in a shot of video.

2.1.2 Motion activities

The Motion Activities descriptor captures intuitive notion of human about actions in video sequence like "intensity of action" or "pace of action". Examples of high activity include scenes such as "goal scoring in soccer mach", "a high-speed car chase", while examples of low action shots are "news reader shot", "an interview scene", and so on. This descriptor enables us to accurately express the activity of a given video sequence. Motion Activity descriptor includes several attributes that enable to describe motion of video sequence more efficiently. These attributes are Intensity of activity, Direction of activity, Spatial distribution of activity, and Temporal distribution of activity. In our application, Intensity of activity is used because it represents the high motion of objects in goal scene.

2.1.3 Audio Power

The excitement of the commentator is a valuable feature for detection of goal events. This feature can be capture by Audio Power descriptor of MPEG-7. Audio Power descriptor is a basic low-level audio descriptor of MPEG-7. It describes the power of the audio signal in the form of the instantaneous power of the samples in the frame.

2.2 High-level features

As discuss in section 2.1.1, color histogram feature is used specify the grass color in the field of soccer game, and then to find field region of each frame. The field regions can contain some objects such as players, referees, the line, goal, and overlay block. We can segment two kinds of objects in this region based on their attributes. The first kind is players and referees and the second is overlay blocks that show information about the match.

The players and referees are objects that have different colors to the background color (grass color); moreover their positions are not fixed through a video sequence; and there exist some motion vectors at their position. From these objects, we get their medium size. The object whose size is far from medium size is eliminated. Finally we have the medium size of the real objects. Normalize to the size of video frame, the values of this feature is in range from 0 to 1.

The overlay blocks that show some information about the game also have different color to the background, but their position is fixed through continuous frames, and there is no motion vector at their positions. Fuzzy logic is used to find the existence potential of overlay blocks in each frame, and it is a high-level feature that is used for goal-event detection.

III. FUZZY RULE-BASED SYSTEM FOR GOAL EVENT DETECTION

This section presents two steps for event detection based on fuzzy theory. In the first step, concept of fuzzy set is used to process the input features (fuzzy conceptualization). The second step combines these fuzzy features by fuzzy rules as a way for knowledge representation.

3.1 Fuzzy conceptualization

To use fuzzy logic and fuzzy systems for problem solving, the problem must be represented in fuzzy terms. This process is called conceptualizing in fuzzy terms [10]. The objective here is to represent input and output values as linguistic variables. A linguistic variable is a variable which takes fuzzy values and has a linguistic meaning. Features extracted in the previous section are fuzzied like that:

1. Color-Histogram difference (HD)

Color-histogram difference for a video sequence is average changing of all consecutive frames.

$$HD = \frac{1}{N}\sum_{i=1}^{N} HD_{i,i+1}$$

where $HD_{i,i+1}$ is the color-histogram difference between frame i and $i+1$.

HD is fuzzied by labeling as "small", "moderate", "large".

2. *Motion activities (MA)*

In MPEG-7, the intensity of activity is calculated and taken value from 1 to 5.

3. *Audio power (AP)*

$$AP = \max(AP_i)$$

AP is also fuzzied to two levels *high* and *low*.

4. *Size of objects (SO)*

$$SO = \frac{1}{N}\sum_{i=1}^{N}|SO_{i+1} - SO_i|$$

SO is fuzzied to "low" and "high"

5. *Overlay existence (OE)*

Overlay existence can be fuzzied based on the attributes of their component.

3.2 Fuzzy Rule Based System

The attributes of goal event in soccer video are that:

- It has loud speech of the commentators.
- It has high motion activities due to the running of scoring players, the moving of viewer, etc.
- The cameras change very fast from the goal-scoring action, to the main actors of these events, to the couch, audiences, and replay the goal scoring again from some different points of views. Therefore in the video segment of goal, color histogram of frames also change very fast.
- The overlay blocks occur after the goal to show information like the current score, the actor of goals, etc.

The attributes above are the common cinematic features of goal-events in soccer video. They are the knowledge to be used in detection system. This knowledge is represented in fuzzy model in the rule like that:

IF		***HD***	is *high*
	and	***MA***	is *high*
	and	***AP***	is *high*
	and	***SO***	is *high*
	and	***OE***	is *high*
THEN	Goal		

Figure 1. Rule for goal event

IV. EXPERIMENT

The first experiment has been done with a soccer 90-minute video sequence. Our method detected 4 over 5 goal events with no wrong detection. The missed goal is a penalty that has little difference situation to other goals.

V. CONCLUSION

In this paper, we presented a new approach for video event detection using fuzzy technique. Goal-event detection for soccer video is performed to illustrate this approach.

This approach has two advantages. First, the fuzzy inference is a natural way for knowledge representation that bridges the gap between the low-level audiovisual feature and the semantic events. Second, fuzzy inference is a flexible method for combining multi modalities.

In the future, more experiment will be done to get the performance of the detection system and compare to other methods. Other events will be considered such as foul, penalty, etc.

VI. REFERENCES

[1] Cees G.M. Snoek and Marcel Worring, "Multimodal Video Indexing: A Review of the State-of-the-art," ISIS technical report series, vol. 2001-20, Dec 2001.

[2] Nguyen N. Thanh, Truong C. Thang, Beet Nara Bea, and Yong Man Ro, "Video abstraction Generation Supervised by User Preference", *SPIE Electronic Imaging*, Jan. 2003.

[3] Nguyen N. Thanh, Truong Cong Thang, Tae Meon Bae, Yong Man Ro, "Soccer Video Summarization System Based on Hidden Markov Model with Multiple MPEG-7 Descriptors", *CISST '03*, vol. 2, pp. 673-678, 2003.

[4] A. Ekin and A. M. Tekalp, "Automatic Soccer Video Analysis and Summarization," in Symp. Electronic Imaging: Science and Technology: Storage and Retrieval for Image and Video Databases IV, IS&T/SPI03, Jan. 2003, CA.

[5] Xie, L., Chang, S.-F., Divakaran, and A., Sun, H., "Structure analysis of soccer video with hidden Markov models," Proc. ICASSP 2002.

[6] Peng Xu, Lexing Xie, Shih-Fu Chang, Ajay Divakazan, Anthony Vetro, Huifang Sun, "Algorithms and system for segmentation and structure analysis in soccer video," Proc. ICME 2001.

[7] L.A. Zadeh, "Fuzzy sets", Information and Control, vol. 8, pp. 338-353, 1965

[8] Kazuo Tanaka and Hua O. Wang, "Fuzzy Control Systems Design and Analysis"

[9] B.S. Manjunath, Philippe Salembier, Thomas Sikiora, "Introduction to MPEG-7 Multimedia Content Description Interface", John Wiley & Sons Ltd., 2002.

[10] R. Jadon, S. Chaudhury, and K. Biswas, "A Fuzzy Theoretic Approach for Video Segmentation Using Syntactic Features", Pattern Recognition Letters, 22:1359-1369, Nov. 2001.