

Review on Event Retrieval in Soccer Video

Mr. Ganesh.I.Rathod¹, Mrs.Dipali.A.Nikam²

Assistant Professor, Department of Computer Science & Engineering,

*Dr.J.J.Magdum Engineering College, Jaysingpur
Maharashtra, India*

Assistant Professor, HOD-Department of Computer Science & Engineering,

Dr.J.J.Magdum Engineering College, Jaysingpur, Maharashtra, India

Abstract— Event retrieval is one of the important topics of research in multimedia processing. Several works related with the retrieval of sports video are presented here and finally a hybrid method is been proposed which makes indirect use of human knowledge in their retrieval process. The proposed method aims to efficiently build human knowledge directly for soccer video events retrieval by fuzzy systems..

Keywords— Video Event Retrieval, Soccer Video, Fuzzy Systems.

I. INTRODUCTION

The Multimedia collections are growing rapidly in both the professional and consumer environment, and are characterized by a steadily increasing capacity and content variety, such as movies, documentaries, sports, news, home videos, e-learning, etc. Therefore, it is more difficult to find relevant video from an increasing video database. The production and availability of large video collection necessitated video retrieval relevant to a user defined query. This has become one of the most popular topics in both real life applications and multimedia research. There are vast amount of video archives including broadcast news, documentary videos, meeting videos, sports, movies etc. The video sharing on the web is also growing with a tremendous speed which creates perhaps the most heterogeneous and the largest publicly available video archive. Finding the desired videos is becoming harder and harder everyday for the users. Research on video retrieval is aiming at the facilitation of this task. In a video there are three main type of information which can be used for video retrieval: visual content, text information and audio information. Even though there are some studies on the use of audio information, it is the least used source for retrieval of videos. Mostly audio information is converted into text using automatic speech recognition (ASR) engines and used as text information. Most of the current effective retrieval methods rely on the noisy text information attached to the videos. This text information can be ASR results, optical character recognition (OCR) texts, social tags or surrounding hypertext information. Nowadays most of the active research is conducted on the utilization of the visual content. Perhaps it is the richest source of information, however analyzing visual content is much harder than analyzing the other two. There are two main frameworks for video retrieval: text-based and content-based. The text-based methods achieve retrieval by using the text information attached to the video. Content-based

approaches the visual features such as color, texture, shape and motion.

II. SEVERAL EVENT RETRIEVAL METHODS ON SPORTS VIDEO

The soccer video event retrieval based on context cues is discussed in [1]. As to the soccer video, the event is defined as the medium-level spatiotemporal entity interesting to users, having certain context cues corresponding to the specific domain knowledge model. As a medium-level entity, the inference of soccer event is based on the fusion of context cues and domain knowledge model. The shooting event is chosen as research target and the event analysis method is expected to be reusable or other soccer events. According to the analysis of shooting event, the following seven kinds of context cues are extracted, respectively including one kind of caption detection, two kinds of face detection(single and multiple), one kind of audience detection, one kind of goal detection, and two kinds of motion estimation.

The content-based TV sports video retrieval based on audio visual features and text information is discussed in [2]. Because video data is composed of multimodal information streams, this paper use multimodal information such as visual information, auditory information and textual information to detect the events of sports video and realize content-based retrieval by quickly browsing tree-like video clips and inputting keywords within predefined domain. This paper particularly concerned with the events that maybe change the score and are interesting for fans and coaches or kinematics researchers: *i*) penalty kicks (PK), *ii*) free kicks next to goal box (FK), and *iii*) corner kicks (CK).

An enhanced query model used for soccer video retrieval using temporal relationships is discussed in [3]. This paper develops a general framework which can automatically analyze the sports video, detect the sports events, and finally offer an efficient and user-friendly system for sports video retrieval. Here, a temporal query model is designed to satisfy the comprehensive temporal query requirements, and the corresponding graphical query language is developed. The advanced characteristics make our model particularly well suited for searching events in a large scale video database. A soccer video retrieval system named *SoccerQ* is developed to support both the basic queries and the relative temporal queries. In *SoccerQ*, video shots are stored, managed, and retrieved along with their corresponding features and meta-data.

A fuzzy logic based approach is introduced in [4], which concurrently detects transition and shot boundaries. This paper introduces a unified algorithm for shot detection in sports video using Fuzzy Logic as a powerful inference mechanism. Fuzzy logic overcomes the problems of hard cut thresholds and the need to train large data. The proposed algorithm integrates many features like color histogram, edgeness, intensity variance, etc. Membership functions to represent different features and transitions between shots have been developed to detect different shot boundary and transition types. Here, algorithm addresses the detection of cut; fade, dissolve, and wipe shot transitions.

An unsupervised content-based indexing for sports video retrieval is discussed in [5]. This paper uses the concept of a grounded language model to motivate a framework in which video is searched using natural language with no reliance on predetermined concepts or hand labeled events. This grounded language model is learned from an unlabeled corpus of baseball games and the paired closed-caption transcripts of the game. Three important features in learning grounded language model are Feature extraction, temporal pattern mining and Linguistic mapping. In Feature extraction two types of features are used visual context, and camera motion. Visual context features encode general properties of the visual scene in a video segment. The first step in extracting such features is to split the raw video into "shots" based on changes in the visual scene due to editing. After it is segmented into shots, each shot is automatically categorized into one of three categories: *pitching-scene*, *field-scene*, or *other*. Detecting camera motion (i.e., pan/tilt/zoom) is a well-studied problem in video analysis. Here a algorithm is used which computes the pan, tilt, and zoom motions using the parameters of a two-dimensional affine model fit to every pair of sequential frames in a video segment. Once feature streams have been extracted, we use temporal data mining to discover a codebook of temporal patterns that are used to represent the video events. The algorithm used here is fully unsupervised. It processes feature streams by examining the relations that occur between individual features within a moving time window. The last step in learning the grounded language model is to map words onto the representations of events mined from the raw video.

A content-oriented video retrieval system is presented in [6], which is capable of handling high volumes of content as well as various functionality requirements. It allows the audience to access the video contents based on their different interests of selected video program. The retrieval system consists of a content-based scalable access platform for supporting content-based scalable multifunctional video retrieval (MFVR). The retrieval methodology is based on different content semantic layers. Based upon their semantic significance video content is divided into video clip layer, object class layer, action class layer and conclusion layer. Each frame in each layer (except the video clip frame) has pointers to the related frames in the lower and the upper layer. Each frame in the upper three layers has a pointer to the related video clip frame in the bottom layer. This paper not only contains the content-

scalable scheme that is adaptive for different subjective client user but also supports different video scalable content retrieval.

An Event-based Soccer Video Retrieval with Interactive Genetic Algorithm is introduced in [7]. This paper presents a system to segment a soccer video into four different kinds of events (i.e. "play-event", "replay-event", "goal-event", and "break-event") and implements retrieval for different purposes to meet the different needs for the users with interactive genetic algorithm. Once soccer video is segmented firstly in this paper soccer game videos are divided into audio and video. Then, eight audio-visual features (i.e. average shot duration, standard deviation of shot duration, average motion activity, standard deviation of motion activity, average sound energy, standard deviation of sound energy, average speech rate and standard deviation of speech rate) are extracted from video and the corresponding audio track. These features are encoded as a chromosome and indexed into the database. For all the database videos, the above procedure is repeated. For retrieval, IGA is exploited that performs optimization with the human evaluation. System displays 15 videos, obtains a relevance feedback of the videos from a human, and selects the candidates based on the relevance. A genetic crossover operator is applied to the selected candidates. To find the next 15 videos, the stored video information is evaluated by each criterion. Fifteen videos that have higher similarity to the candidates are provided as a result of the retrieval.

An advanced sports video browsing and retrieval system based on multimodal analysis, SportsBR, is proposed in [8]. Here, system analyzes an inputting sports video by dividing it into video and audio streams respectively. In video streams, it processes the visual features and extracts the textual transcripts to detect the shots that probably contain the events. Visual features are not sufficient for detecting events, so the textual transcripts detection can improve the accuracy and it also use to generate the textual indices for users to query the video events clips. In audio streams the speech recognition of special words such as "goal" or "penalty kick" and use these words to generate the textual indices, too. In audio signal processing module, several parameters of audio signal are computed to find the interesting parts of sports video more accurately. After the video events clips is found, it is organized for content-based browsing and retrieval. Combining audio-visual features and caption text information, the system can automatically selects the interesting events. Then using automatically extracted text caption and results of speech recognition as index files, SportsBR supports keyword-based sports event retrieval. The system also supports event-based sports video clips browsing and generates key-frame-based video abstract for each clip.

An advanced content-based broadcasted sports video retrieval system, SportBR, is proposed in [9]. Its main features include event-based sports video browsing and keyword-based sports video retrieval. The algorithm introduces a novel approach that integrates multimodal analysis, such as visual streams analysis, speech recognition, and speech signal processing and text extraction to realize event-based video clips selection. The algorithm first

selects video clips that may contain events. Then it describes how to get keywords such as "goal" or "penalty kick" by speech recognition and detect interesting segments by computing the short time average energy and other parameters of audio. At last, a method of extracting textual transcript within video images is introduced to detect events and use these textual words to generate the indexing keywords. The system is proved to be helpful and effective for understanding the sports video content.

An Event Retrieval in Soccer Video from coarse to fine based on Multi-modal Approach is introduced in [10]. A soccer video clip contains both visual data and audio data. If both types of data are available video clip could be understood more comprehensively. This paper imitates the way human beings understand videos by multimodal retrieval model. It uses the multimodal method in video retrieval taking advantages of two types of data. Event could be described by shot sequence context containing interest objects, transcript text. Retrieval strategy of this algorithm is searching from coarse to fine. At first stage, the video sequence is segmented into shots and classified them into 4 basic types including long view, medium view, and close up and audience. This stage helps gathering event candidates in the pre-filter stage at high speed. In second stage, refining the results is done based on interest objects and transcript text. The soccer objects that we suggest using are referee, penalty area, text box, excitement (shout), whistle and audio keyword (translated from speech). Excitement sound here may stem from two sources; shouts from the audience and the report of the commentators that appears when there is an attractive situation. Whistle can be detected based on frequency domain with the appearance of strong signal in the specific prior range. In this paper, new query language is introduced to describe an event as sequence of components in visual and audio data. It is used to retrieve relevant sequences automatically. It can be performed quickly and flexibly. More than that, system can be easily developed with new visual and audio components.

Semantic analysis and retrieval of sports video is discussed in [11]. A fully automatic and computationally efficient for the indexing, browsing and retrieval of sports video is explained. A hierarchical organizing model based-on MPEG-7 is introduced to effectively represent high-level and low-level information in sports video. The model's attributes have been instantiated through a series of multi-modal processing approaches for soccer video. To effectively retrieve highlight clip of sport video, adaptive abstraction is designed based on the excitement time curve, and an XML-based query scheme for semantic search and retrieval is based on the hierarchical model. A robust detection algorithm is developed to automatically locate the field region of sports video based on low-level color features. An effective mid-level shot classification algorithm is proposed based on spatial features and the coarseness of human figures. For retrieval, XQuery, a XML query language, is employed in this paper.

Semantic Segmentation and Event Detection in Sports Video using Rule Based Approach is discussed in [12]. The paper addresses two main problems of sports video processing: semantic segmentation and event detection.

This paper proposes a novel hybrid multilayered approach for semantic segmentation of cricket video and major cricket events detection. The algorithm uses low level features and high level semantics with the rule based approach. The top layer uses the DLER tool to extract and recognize the super imposed text and the bottom layer applies the game rules to detect the boundaries of the video segments and major cricket events.

Semantic Event Detection in Soccer video by Integrating Multi-features using Bayesian Network is discussed in [13]. In this paper Bayesian network is used to statistically model the scoring event detection based on the recording and editing rules of soccer video. The Bayesian network fuses the five low-level video content cues with the graphical model and probability theory. Thus the problem of event detection is converted to one of the statistical pattern classification. The learning and inference of Bayesian network are given in this paper.

Goal Event Detection in Broadcast Soccer Videos by Combining Heuristic Rules with Unsupervised Fuzzy C-Means Algorithm is discussed in [14]. In this paper, based on three generally defined shot types, goal events are detected by combining heuristic rules with unsupervised fuzzy c-means algorithm. First heuristic rule based primary selection/filtering for potential goals is carried out in the shot layer which composed of three generally defined shot types, together with the number of frames within each shot is recorded as representative feature. Then to further classify goal events from other potential goals unsupervised fuzzy c-means (FCM) algorithm is adopted. The main contribution of this work is the combination of heuristic rule which is based on three generally defined shot types with unsupervised fuzzy c-means algorithm. And when defuzzified with prior knowledge of the number of goals within each match, accurate and robust results can be achieved over five half matches from different series produced by different broadcast stations.

A Hybrid Method for Soccer Video Events Retrieval Using Fuzzy Systems is discussed in [15]. The new method here uses human knowledge directly and in a very efficient way by a fuzzy rule base. The presented structure allows the system to process based on soccer video shots available in the database. The first phase is devoted to extracting shots from each video and making a list of features extracted from each shot. Then a fuzzy system is used to eliminate shots including insignificant events. Finally shots are classified and associated with predefined classes using a SVM. Then shots related to the class associated with the user query are provided as an answer to that query. The user may make queries on different events and concepts such as occurrence of penalties, corners or goals or team attacks throughout the data base.

III. PROPOSED MODEL

In most of the information retrieval systems attempt is made to embed the human knowledge in their retrieval process in one or the other form. The proposed method here aims to efficiently use human knowledge directly for soccer video events retrieval by fuzzy systems.

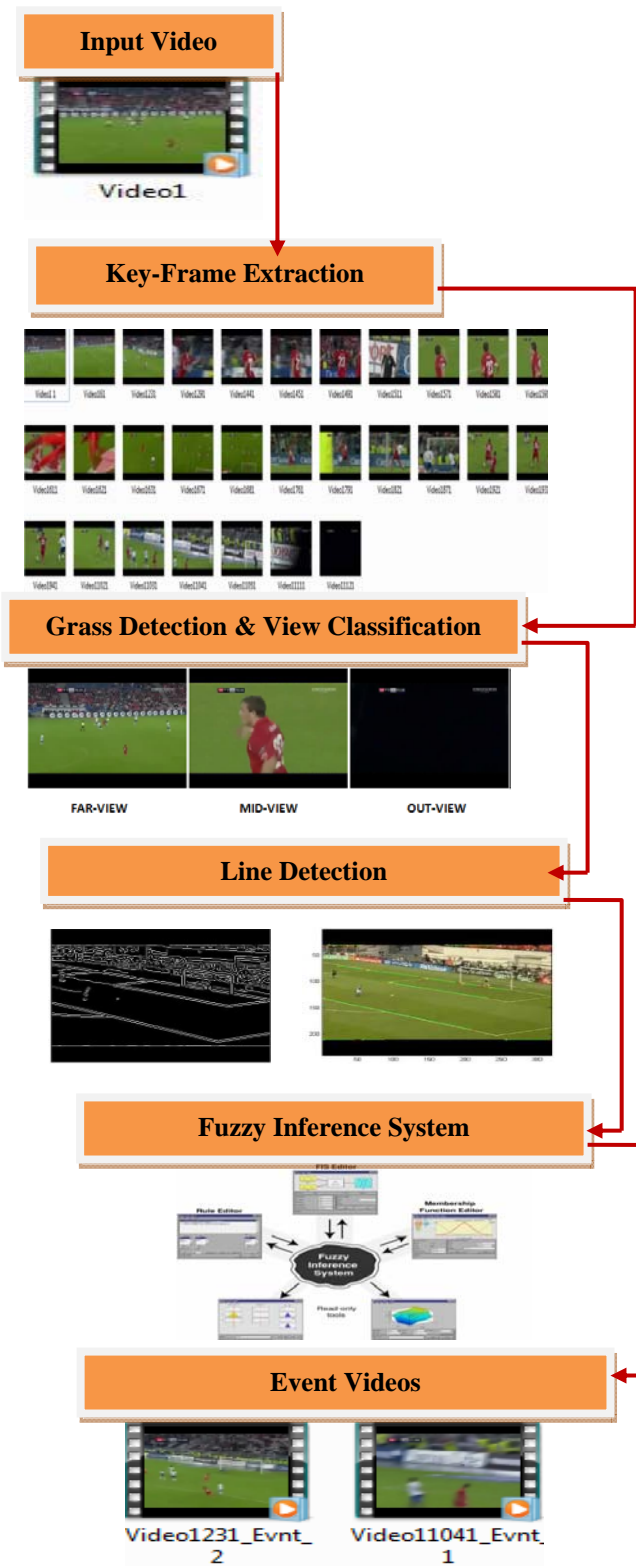


Fig. 1 Methodology for event retrieval.

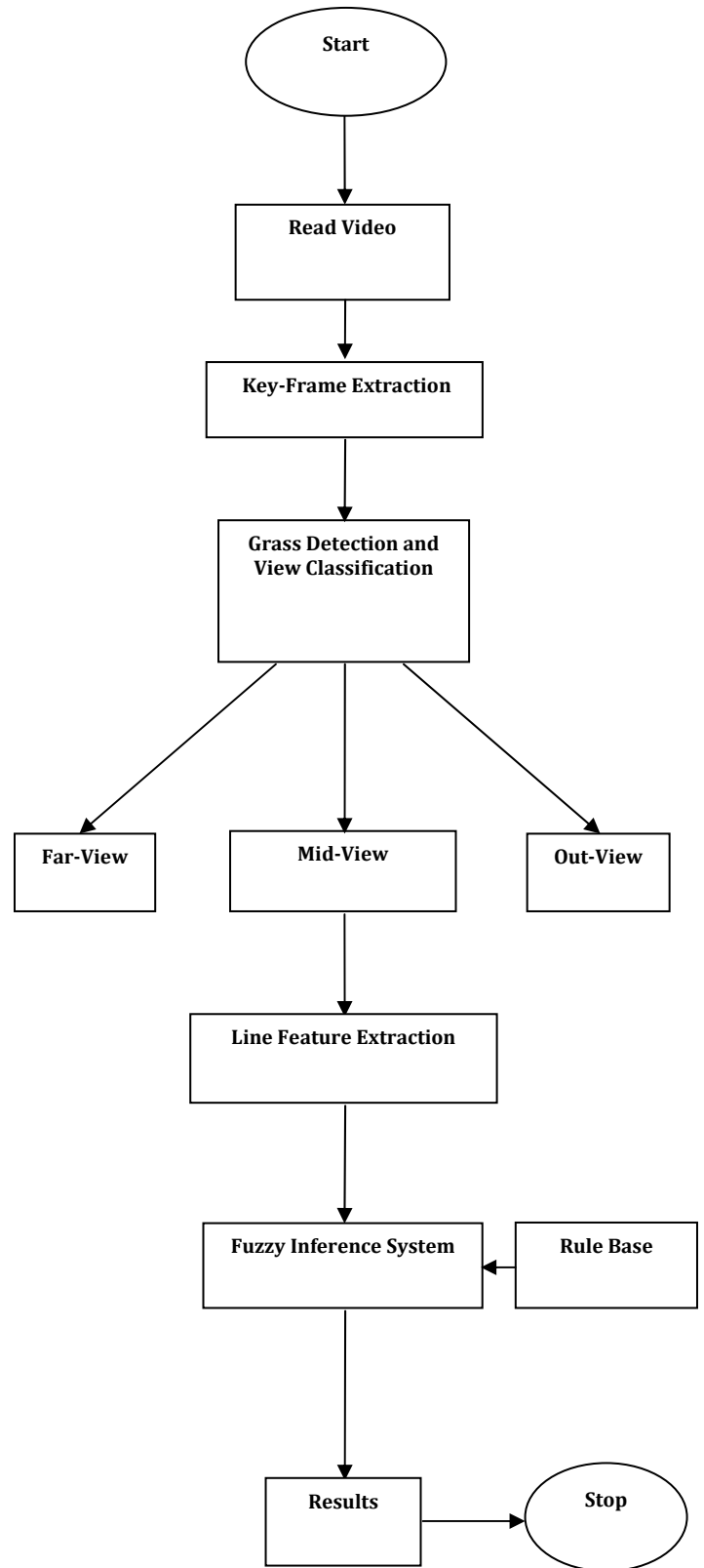


Fig. 2 Overall structure of the proposed plan.

Fuzzy logic is relatively young theory. Major advantage of this theory is that it allows the natural description, in linguistic terms, of problems that should be solved rather than in terms of relationships between precise numerical values. This advantage, dealing with the complicated systems in simple way, is the main reason why fuzzy logic theory is widely applied in technique. The first phase consists of extracting suitable features from video shots. Then, using a fuzzy rule shots that include significant events are extracted. The overall structure of the proposed plan is shown in Figure 1 and the diagrammatic representation of proposed algorithm can be shown as above figure 2.

The proposed algorithm can be explained by following steps.

Step 1: The input for this algorithm is soccer videos, hence initially Avi soccer clip is been read.

Step 2: Key-frames are been extracted..

Step 3: Input to this step is the key-frames extracted from the above step using which grass detection and 3 types of views, Far-view, mid-view and out-view is been categorized.

Step 4: Considering the frames of mid-view generated above, line features are extracted.

Step 5: Fuzzy Inference System is designed to retrieve significant events from a soccer video. The input to the fuzzy inference system is line features and grass percentage extracted from the above steps.

Step 6: Finally, the significant video is constructed around the event frame generated by fuzzy.

IV. CONCLUSIONS

As reviewed, there are different methods in event retrieval. The basic idea in this proposal is to indicate that other than selection of good shots, provisions should be made for rejection of useless shots. As we all know most of the time the football is in the middle of the field and the major events such as goals, corner kicks, penalties, etc. bound to happen near goal post. Consider a retrieval that includes removal of all probably useless shots, such as shots of fans, close shots of coaches, shots of the midfield, etc. and concentrate near goal post area. Such retrieval will mostly cover the attacks and major events of the match. The

proposed method uses human expert knowledge directly for soccer video events retrieval by fuzzy systems. This advantage of applying fuzzy is that, it deals with the complicated systems in simple way.

REFERENCES

- [1] SUN Xing-hua and YANG Jing-yu, "Inference and retrieval of soccer event", IEEE 2007.
- [2] LIU Huayong, "Content-based Tv sports video retrieval based on audio-visual features and text information", IEEE 2004.
- [3] Shu-Ching Chen, Mei-Ling Shyu and Na Zhao, "An Enhanced query model for soccer video retrieval using temporal relationships", IEEE 2005.
- [4] Mohammed A. Refaey, Khaled M. Elsayed, Sanaa M. Hanaf and Larry S. Davis, "Concurrent transition and shot detection in football videos using fuzzy logic", IEEE 2009.
- [5] Michael Fleischman, Humberto Evans and Deb Roy, "Unsupervised content-based indexing for sports video retrieval", IEEE 2007.
- [6] Huang-Chia Shih and Chung-Lin Huang, "Content-based scalable sports video retrieval system", IEEE 2005.
- [7] Guangsheng Zhao, "Event-based soccer video retrieval with interactive genetic algorithm", IEEE 2008.
- [8] HUA-YONG LIU and HUI ZHANG, "A Sports video browsing and retrieval system based on multimodal analysis: Sportsbr", IEEE 2005.
- [9] Liu Huayong and Zhang Hui, "A Content-based broadcasted sports video retrieval system using multiple modalities: Sportbr" IEEE 2005.
- [10] Tung Thanh Pham, Tuyet Thi Trinh, Viet Hoai Vo, Ngoc Quoc Ly and Duc Anh Duong, "Event retrieval in soccer video from coarse to fine based on multi-modal approach", IEEE 2010.
- [11] YU Jun-qing, HE Yun-feng, SUN Kai, Wang Zhi-fang and WU Xiang-mei, "Semantic analysis and retrieval of sports video", IEEE 2006.
- [12] Dr. Sunitha Abburu, "Semantic segmentation and event detection in sports video using rule based approach", IJCSNS International Journal of Computer Science and Network Security, 2010.
- [13] Chen Jianyun, Li Yunhao, Wu Lingda and Lao Songyang, "Semantic Event Detection in Soccer video by Integrating Multi-features using Bayesian Network". Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.
- [14] Ehsan Lotfi, M.-R. Akbarzadeh-T, H.R. Pourreza, Mona Yaghoubi and F. Dehghan, "Event Detection and Automatic Summarization in Soccer Video", IEEE 2009.
- [15] Ehsan Lotfi, M.-R. Akbarzadeh-T, H.R. Pourreza, Mona Yaghoubi and F. Dehghan, "A Hybrid Method for Soccer Video Events Retrieval Using Fuzzy Systems", IEEE 2009.
- [16] Yina Han, Guizhong Liu and Gerard Chollet., "Goal event detection in broadcast soccer videos by combining heuristic rules with unsupervised fuzzy c-means algorithm" 2008 10th Intl. Conf. on Control, Automation, Robotics and Vision Hanoi, Vietnam.