A Hybrid Method for Soccer Video Events Retrieval Using Fuzzy Systems

Ehsan Lotfi[1], M.-R. Akbarzadeh-T[2], H.R. Pourreza[3], Mona Yaghoubi[1] and F. Dehghan[1]

[1]Department of Computer Engineering, Islamic Azad University,
Mashhad Branch, Iran.
[2]Departments of Electrical Engineering and Computer
Engineering, Ferdowsi University of Mashhad, Iran and IEEE Senior Member.
[3]Departments of Computer Engineering, Ferdowsi University of Mashhad.
esilotf@gmail.com, Akbarzadeh@ieee.org, hpourreza@um.ac.ir, mona.yaghoobi@gmail.com

Abstract— *Most information retrieval systems make indirect use of human knowledge in their retrieval process. The new method we present here aims to efficiently use human knowledge directly in combination with support vector machines for clustering. As illustrated in this paper, this approach is particularly applicable to concept retrieval from soccer-related videos. The first phase consists of extracting suitable features from video shots. Then, using a fuzzy rule base containing the experiences of experts, shots that do not include significant events are removed. Finally, the last phase uses SVM to classify results coming from the fuzzy system. The results of the classification phase are accompanied by a textual description and enables retrieval through text based query. Experimental result show good classification and satisfying retrieval process.*

Keywords— Event Retrieval, Semantic Indexing, Hybrid Method, Fuzzy System, SVM, Soccer Video.

1 Introduction

Videos are used for different purposes such as education, mass media, entertainment, and surveillance systems. In each of these applications a video conveys a significant message to the audience. For example, in a football match video events such as penalties, corners, goals, and the very concept of victory or loss are conveyed when fans watch the video. The meanings contained in a video are quite obvious for men, but the same does not go for computers. A computer cannot understand the scoring of a goal in a football match video. This fact is usually referred to as the semantic gap between man and computers. Imagine if computers were capable of detecting and understanding the meaning of videos then we could use them to summarize sport videos automatically and retrieving meanings from a multimedia database. On the other hand the dramatic increase of multimedia data volumes in recent years will eventually force us to use some sort of retrieval system so that the indexing and retrieval of information is handled automatically. As noted earlier there is a semantic gap between the user and retrieval systems. The user queries the system with his/her ideas and the system presents it findings for such ideas as an answer. Such queries are done in different manners: 1. Query by keywords[11] 2. Sketch-based queries 3. Example-based queries[6] 4. Semantic-based queries[1]. Methods 2 and 3 are usually used in Content-Based Information Retrieval (CBIR). Recent studies have focused on Context instead of Content resulting in the presentation of a range of Semantic-Based Video Retrieval methods which are discussed in the published works section. The system presented here is for assessing significant events included in a database containing shots of football videos. Queries with the user are based on text. For instance, the user can query the system with items such as this: shots of goals scored by the yellow team. The user interface is designed in such a way that the user can ask for the summary of any given video. Published works include several major methods and techniques for retrieval of information-which are common in some features- as summarized here: 1. Context Based Information Retrieval (CBIR); In such methods the media is modelled by features such as colour, texture, etc. and special relationships between objects and movements. In CBIR suitable features are extracted and related to high level concepts and meanings, without the system knowing such meanings. In fact the system displays and models the video contents in such a way that is efficient for content retrieval. For example there is a statistical model of motion in [3]. After that there is a training session on dynamic contents and then the data are recognized and classified. The result is a database of different video shots (football, basketball, meeting, and highways) each divided into minor events. The provided structure classifies shots in various groups based on similarity of their contents. Also there is fuzzy presentation of video contents in [12]. In this paper suitable features of a multidimensional fuzzy histogram are made for each video frame. One feature will be suitable for studying the similarities between frames. Consecutive frames are classified based on the fuzzy histogram. In fact the paper divides the videos into shots by this method, and then it chooses a keyframe for each of these frames. Then the shot related to the keyframe is provided as the answer.

2. Semantic Based Retrieval; One of the most essential phases of designing a Semantic-based retrieval system is defining a meaningful list of concept-oriented meanings based on human's available knowledge. Thus each single

ISBN: 978-989-95079-6-8

concept in the list should be accompanied by correct and true descriptions on the video collection. A new perspective introduced in recent years includes the use of Semantic Description of Object Motion for the retrieval process, which has been studied in [1],[2],[8],[7], and [6]. The method offered in [1] is as follows: First objects included in video frames are extracted and their trajectories are tracked. Once the trajectories are extracted they are classified for the training of object motion models. Finally, a meaningful description is added to these motion models. It should be added that such a meaningful description was added manually.

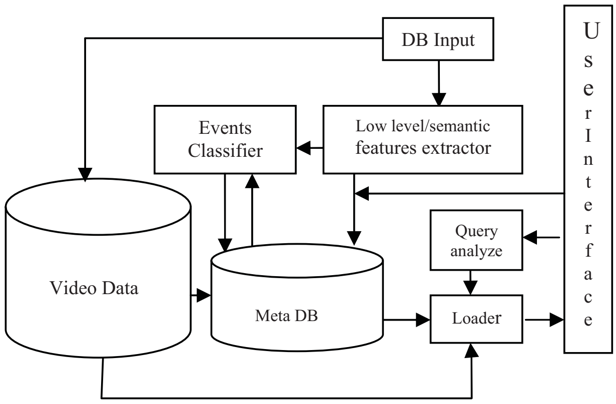

Fig. 1 The overall proposed structure

2 Proposed Structure

Most information retrieval systems make indirect use of human knowledge in their retrieval process(for example in the query by example, all of instance are a priori knowledge and in Semantic Based retrieval a meaningful description as priori knowledge should be added to the visual models manually.) The new method we are presenting in 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 classed 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 DB. The overall structure is provided in Figure 1. As can been seen each video is processed before entering the DB and its low level features and conceptual properties are extracted, and once the events are classified all collected information is stored along with all data provided by the user (such as names of teams, time and place of the match, etc.) inside the Meta DB, while the video itself is stored in another location. Indeed the assumed database is of Multimedia linked meta database type. Figure 2 demonstrates feature extraction, Shot filtering using fuzzy system and event classification stages.

2.1 Extracting Low-Level and Semantic-Based Features

Used low-level features includes: Gi percent of pixels including grass of the field in keyframe of shot i for which we used the method provided in [5]. Once the grass region was recognized we divided it into 3 sections according to the Fig.3 and calculated Gi(E2) , percent of pixels covering the section E2 of the grass field, for keyframe. These features(Gi,Gi(E2)) may be used for Shot boundary detection and Shot classification and may be used as input to the fuzzy system not only in the Shot Rejection but also in the next phase of Event Classification along with low-level features.

The first step of proposed structure is devoted to extracting shots from each video (Fig. 2) 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 classed using SVM The results of the classification phase are accompanied by a textual description and enables retrieval through text based query. The proposed method is based on the idea of using a fuzzy system for shot rejection and textual description for Soccer Events retrieval. In this section first the general structure is studied and we proceed to describe the extraction of low level features. Afterwards we go to high level features and then development of visual models. The fuzzy system including fuzzy rule base is presented in the section 2.2.

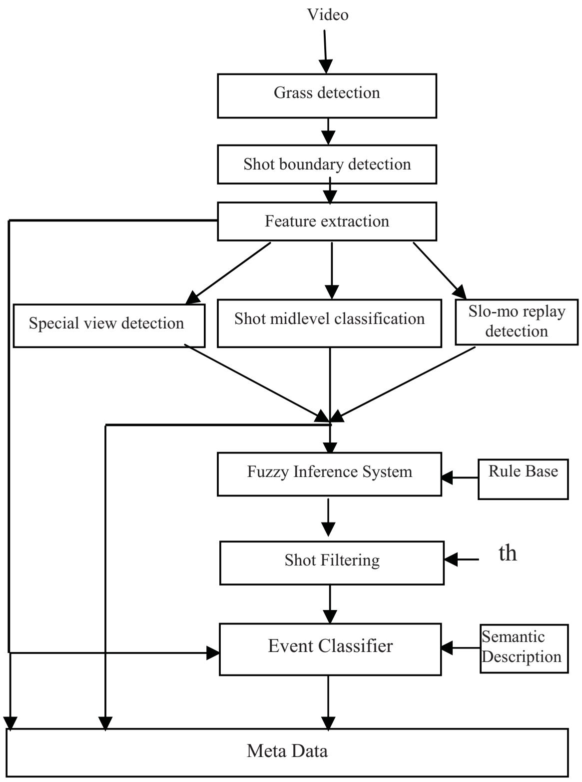

Fig. 2 Proposed structure in detail

ISBN: 978-989-95079-6-8

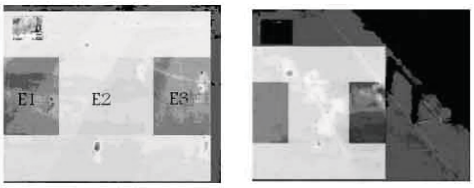

Fig. 3 Segmenting the area including the grass field

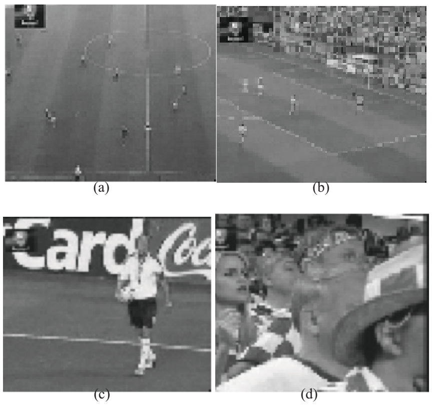

Fig.4 Types of views: a) Far-center, b) Far-side, c) Medium view, d) Out of field

Used semantic based features includes: 1. Far views from midfield, Fig. 4.a. 2. Far views from field sides, Fig. 4.b. 3. Medium views from inside the field 4.c, 4. View of the area outside of the field / closed view, showing the spectators or the upper body of a player, Fig. 4.d 5. Slow-Motion Replay Detection[9]. we have proposed a structural classification method for extracting this views in [5].

2.2 Using a fuzzy system to eliminate shots containing no significant events

Data retrievable from each shot by the above features as input of fuzzy system up to this phase are included in Table 1 along with abbreviations used for each one of them. The goal of designing a fuzzy system is to use the collected data in defining a degree of significance for each shot. This fuzzy system shows how important are the contents of that shot in terms of events occurring in it. For example, the system is expected to give a higher degree of significance to a shot containing a goal scoring than the one containing images of fans and no other important event. The rulebase and 3 triangle membership functions (low, Medium, High) are presented in Figure 5,6. The inference method used here is Mamdani product, for example for one execution of fuzzy inference result see Fig 7. Once each shot is given a degree of significance by the fuzzy system, a thresholding phase is run to eliminate the fuzzy system outputs for some shots. In this phase all shots with a significance degree below th(our system parameter) are eliminated from the classification. by defining a threshold(th) on the output of the fuzzy system we can separate the useful and none-useful shots for the purpose of classification. The threshold must be between 0 and 1. The lowest amount of the threshold is zero. By defining the threshold as 0 the all of shots in the database will be seen in the output of rejection phase as well. the higher the degree of significance the more sensitive the system vice versa.

Table 1 The input/output of fuzzy system with their abbreviations

Shot Degree	Out	Sd
Percent of Far-center view in a shot	In	Fc
Percent of Far-side view in a shot	In	Fs
Percent of Medium view	In	Mv
Percent of Out of field	In	Of
Percent of Frame includes Slow-Motion	In	Sm

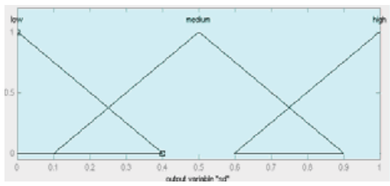

Fig. 5 triangle membership functions (low, Medium, High) for most In/Out variables

1. If (fc is high) then (sd is low) (1)
2. If (fs is high) then (sd is very high) (1)
3. If (fs is medium) then (sd is high) (1)
4. If (mv is high) then (sd is high) (1)
5. If (of is high) then (sd is low) (1)
6. If (cu is high) then (sd is low) (1)
7. If (cu is medium) then (sd is medium) (1)
8. If (fs is low) then (sd is medium) (1)
9. If (sm is high) then (sd is high) (1)
10. If (sm is medium) then (sd is high) (1)

Fig. 6 Fuzzy rule base for none-useful shot rejection

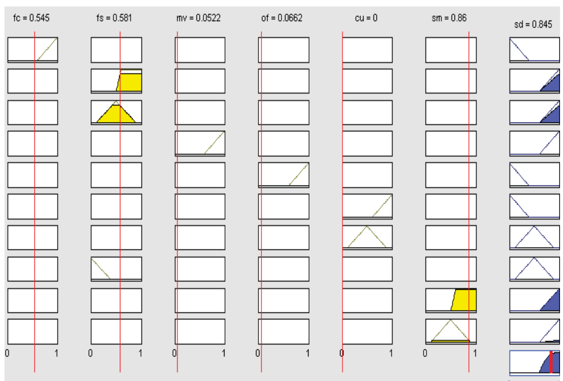

Fig. 7 if(fc=0.545 & fs=0.581 & mv=0.052 & of=0.0682 & cu=0 & sm=0.86) then Shot Degree = 0.845

ISBN: 978-989-95079-6-8

3 Event Retrieval

Up to the previous level we rejected most video shots imported into the DB with high probability of being irrelevant to significant events. So classification of shots in this phase may be done based on the events included in them. We have listed the set of features that may be used to classify the events , features include: G, percent of pixels covering the grass field in shot keyframes, G2, percent of pixels in section E2 on the grass field in shot keyframes and the Sd (the output of fuzzy system) and Fs, Percent of Far-side view in a shot. After normalizing the above 4 features the value of each of them for each shot will range between 0 and 1, so the feature vector for each shot is shown by Fv (I,V) and expressed in the following form:

Fv(i,V)={G,G2,Fs,Sd}
Where: i is the shot number from video V.

We used SVM to classify the events. In this phase events must be associated with 5 predefined classes. All classes have their special keywords which are presented here:

EventClass = { Goal, Penalty, Kroner, Free kick, other}

After the learning phase the videos stored in the DB for making up the Meta Data are processed and then the events are classified. Afterwards the above keywords are associated with the created classes manually and then each new event automatically inherits its keyword and is recorded in the Meta Data for that event. A table is formed in the Meta Data for each single video processed (Figure 8). During implementation the name given to each table of the Meta Data is taken to be the same as the file name recorded in the database, so relations with the raw video are maintained. Also we can save the information entered into the system for each video in Title of main video file to be used in the retrieval process. Such information may include the name of competing teams, time or place of the match or other details.

Shot Index	Start of shot	End of Shot	Type of Shot	Slow-Motion	Class of Event
. . .	. .	. . .	. .	. . . . . .	

Fig 8 A table is formed in the Meta Data for each single video processed (The field related to Type of Shot shows the view that has been most frequent in the shot.)

The query model we used for this project was the same proposed in [2]. The base expression used in the model included the following:

select video from search_space [where condition];
select shot from search_space [where condition];

Expression SELECT specifies the SEARCH TARGET which may be Videos or Shots. The search space comes after FROM which may be all videos available in the DB or a list of chosen videos from any given video specified with a "where" condition which may only constitute one restriction or a combination of and/or phrases used to create conditions for time dependencies, which we ignore here.The provided system will be able to answer questions such as these:

Q1: "Find all the goal shots from all the soccer videos"
Q2: "Find all the corner kick shots from all the soccer videos where the goal accurse by England team"
Remember that the prerequisite to doing the condition contained in the second question is that the required information are added to the Meta Data by the system manager while the video is being entered into the database. To find the answer it suffices to extract the defined terms from the query expression and search for them in the Meta Data.

4 Evaluation

We have used more than 5.5 hours of football Videos in the database includes: 1 match of the World Cup 2006, 2 matches from the UEFA Champions League 2005, 1 match from the FA Premier League 2004, and more than 5 clips from Euro 2004(Table 2). The file format of the input films was non-compressed avi and the size of output films is 88×79.

Table 2. Names and the length (min:sec) of the clips in the database

Euro001(46:40),Euro002(47:00),Euro003(47:30), Euro004(46:00), Euro005(46:45), FIFA-div01f(12:38), FIFA-div02f(18:23),FIFA-div03f(15:38),FIFA-div04f(17:40),FIFA-div05f(18:22),Eng1(42:7) ,Eng2(20:51)

In the training phase we used 119 shots to learning for SVM and the rate of correct class in training data is 87%. As noted earlier the output of fuzzy system is a degree for each shot and the system rejects no useful shots using a threshold(th) as system input parameter. The input margin of the system for evaluation is 0.2 and then inputs include 0.2, 0.4, 0.6, 0.8 and 1. The results of changing the input amount of the system and the accuracy of the events classification are shown in tables 3,4,5,6. In this evaluation scenario we used 187 shots includes 5 Goals, 5 Penalties, 6 Free kicks, 9 Corners and 162 shots containing no significant events. Table 3 where the input th is 0.2 then system rejected 23

ISBN: 978-989-95079-6-8

shots before classification, table 3 shows the result of classification after rejection phase with input 0.2, Table 4 where the input th is 0.4 then system rejected 86 shots. Table 5 where th is 0.6 and the system rejected 114 shots and Table 6 where th is 0.8 and the system rejected 172 shots.
When th is 0.6, 4 shots containing corner and free kicks are rejected incorrectly(see the '# of rejected' row of table 5) and When th is 0.8, 10 shots containing very important events are rejected incorrectly (see the '# of rejected row' of table 6).
In the tables accuracy formula is:
Accuracy =100* Correct/(Total + False)
The Tables indicate that the higher the threshold, the higher the number of rejected shots will be. Also, the higher the threshold, the lower the number of incorrect shot detection (see 'False' row of tables) will be. See Table 6 where th = 0.8 and False detection for all events is zero but 1 Goal, 5 Free kicks and 4 corner are rejected incorrectly.

Table 3. The result of proposed method with th = 0.2

	Goal	Penalties	Free kick	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	0	0	0	0	23	23
Correct	5	5	6	9	86	111
False	13	9	17	14	0	53
Accuracy	27%	36%	26%	39%	-	-

Table 4. The result of proposed method with th = 0.4

	Goal	Penalties	Free kicks	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	0	0	0	0	86	86
Correct	5	5	5	8	51	74
False	6	0	13	6	1	27
Accuracy	45%	100%	32%	53%	-	-

Table 5. The result of proposed method with th = 0.6

	Goal	Penalties	Free kicks	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	0	0	2	2	110	114
Correct	5	5	4	7	42	63
False	2	0	5	3	0	10
Accuracy	71%	100%	36%	58%	-	-

Table 6. The result of proposed method with th = 0.8

	Goal	Penalties	Free kicks	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	1	0	5	4	162	172
Correct	4	5	1	5	0	15
False	0	0	0	0	0	0
Accuracy	80%	100%	16%	55%	-	-

5 Conclusions

The proposed method uses human expert knowledge for soccer video events retrieval by fuzzy systems. After feature extraction, each shot is given a degree of significance by the fuzzy system, thereby significantly reducing time complexity of video processing. By defining a threshold on the output of the fuzzy system, the useful and none-useful shots are separated for the purpose of event classification. Support vector machines are then applied for final classification of video information.

References

[1] Weiming Hu, Dan Xie, Zhouyu Fu, Wenrong Zeng, and Steve Maybank, "Semantic-Based Surveillance Video Retrieval", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 16, NO. 4, APRIL 2007.

[2] Cees G. M. Snoek, Bouke Huurnink, Laura Hollink,"Adding Semantics to Detectors for Video Retrieval",IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 9, NO. 5, AUGUST 2007.

[3] Gwenaëlle Piriou, Patrick Bouthemy, and Jian-Feng Yao,"Recognition of Dynamic Video Contents With Global Probabilistic Models of Visual Motion' , IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 15, NO. 11, NOVEMBER 2006.

[4] Cees G.M. Snoek , "Multimedia Event based Video Indexing using Time Intervals," IEEE TRANSACTIONS ON MULTIMEDIA, 2004.

[5] Ehsan Lotfi, H.R Pourreza, "Event Detection and Automatic Summarization in Soccer Video ," 4th Iranian Conference on Image Processing and Machine Vision, Mashhad, Iran, Feb 2007.

[6] Z. Q. Liu, L. T. Bruton, J. C. Bezdek, J. M. Keller, S. Dance, N. R. Bartley, and C. Zhang, "Dynamic image sequence analysis using fuzzy measures," IEEE Trans. Syst., Man, Cybern., vol. 31, no. 4, pp. 557–571, Aug. 2001.

[7] Haiping Sun, Joo-Hwee Lim, Qi Tian, Mohan S. Kankanhalli, "SEMANTIC LABELING OF SOCCER VIDEO", Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on MultimediaVolume 3, Issue , 15-18 Dec. 2003.

[8] B. Bell and L. F. Pau, "Context knowledge and search control issues in object-oriented prolog-based image understanding," Pattern Recognit. Lett., vol. 13, no. 4, pp. 279–290, 1992

[9] Ahmet Eklin,A.murat Tekalp, and Rajiv Mehrotra,"Automatic Soccer Video Analysis and Summarization", IEEE Transactions on Image processing Volume 12, Issue 7, July 2003.

[10] N.-S. Chang and K.-S. Fu, "Query-by-Pictorial-Example," IEEE Transactions on Software Engineering, vol. 6, no. 6, pp. 5 19-524, November 1980.

[11] Yu-Lin Kang ,"SOCCER VIDEO EVENT DETECTION WITH VISUAL KEYWORDS", ICICS-CM 200315-18 December 2003 Singapore.

[12] Anastasios D. Doulamis*, Nikolaos D. Doulamis, Stefanos D. Kollias "A fuzzy video content representation for video summarization and content-based retrieval" , ELAIVIER, Signal Processing 80 (2000) 1049.

[13] " Inference and retrieval of soccer event", SUN Xing-hua, YANG Jing-yu, Journal of Communication and Computer, Volume 4, No.3 (Serial No.28) , Mar. 2007.

ISBN: 978-989-95079-6-8