A NEW METHOD FOR SHOT CLASSIFICATION IN SOCCER SPORTS VIDEO BASED ON SVM CLASSIFIER

A. Bagheri-Khaligh, R. Raziperchikolaei
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
{abagheri, razi}@ce.sharif.edu

M. Ebrahimi Moghaddam
Department of Electrical & Computer Engineering
Shahid Beheshti University
Tehran, Iran
m_moghadam@sbu.ac.ir

Abstract— **Sport video shot classification is a basic step in the sport video processing. For many purposes such as event detection and summarization, shot classification is needed for content filtering. In this paper, we present a new method for soccer video shot classification. At first, in-field and out-of-field frames are separated. In in-field frames three features based on number of connected components and shirt color percent in vertical and horizontal strips are extracted. The features are all new and showed excellent discrimination in the feature space. These features are given to SVM for classifying long, medium and close-up shots. One of the advantages of our method is that, close-ups can be detected in both in-field and out-of-field views. For detecting close-ups in out-of-field shots, the mean of shirt color in horizontal strips is used. Since the features are easy to extract and input frames are downsampled, the method works in real-time. The experimental results demonstrated the effectiveness of proposed method.**

Keywords-component; Shot classification; SVM Classifier; Feature extraction; Connected components

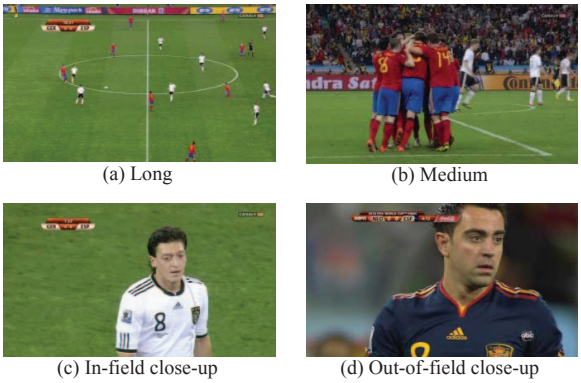
I. INTRODUCTION

Nowadays according to large number of audiences of sport events and broadcasting most of them in various multimedia networks, sport video processing has become an important part of video processing. In most sport video processing, the general aims are detection of significant events and summarization of game, to attain them, some intermediate processing is needed. Shot boundary detection and classification are examples of intermediate processing which are used for content filtering and redundancy reduction.

The important key in sport video processing is speed, because the value of sport video drops significantly after a relatively short period of time [1]. To reach such a speed for extracting game events an approach which does not process all the frames and only needs key frames, should be considered. The best way for extracting these frames is classification of shots into classes such as long, medium and close-up. In general shot classification methods are divided to two categories. In the first category, proposed approaches are independent from sport type such as [2], but in the second one a specific sport is considered such as tennis [3] and soccer [4]. Since in the second approach, more appropriate features can be extracted for each particular sport, the results are usually better in comparison with the first one.

In [5] a simple method for shot classification has been proposed which only uses field percent. The field percent less than a low-threshold considered as close-up, greater than a high-threshold considered as long and between these two values considered as medium. It is clear that the accuracy of this method is too low. In [6] a method has been proposed that its foundation is like [5], only when the field percent is greater than high-threshold, with using minimum bounding rectangle (MBR) and golden section spatial composition, two features are extracted and given to the Bayesian classifier. In addition to requirement of large amounts of training data, such a classification leads to failure in detection of many close-up frames with field background. Another method for classification of shots based on SVM has been presented in [7]. This method uses color distribution, edge distribution and shot length as features that, are given to SVM classifier. Not being real-time is the main problem of this method because shot classifying is not possible until the next shot appears and shot length would be computed, moreover errors of shot boundary detection affect directly the results of classification. In [8] a hierarchical classification has been presented which in the first level, according to audio features, important scenes are extracted. In the second level, based on field percent, field shots and out-of-field shots are separated and then each of these types is divided into subcategories. For example close-up is a subset of out-of-field shots. This method has the same problem in close-ups with field background that was told about [6]. In this classification medium shots have no distinct class but corner shots are separated from straights.

In this paper, we propose a new method based on SVM which can classify main shots in soccer video analysis to close-up, medium, and long. Definition of these shots have been presented in [9], also Fig. 1 shows some examples of them. In addition to being real-time, high accuracy in classification is another advantage of this method. Fig. 2 shows general structure of proposed method. At first out-of-field shots, based on a threshold are separated from in-field ones, then for classifying long, medium, and close-up shots among in-field views, SVM classifier is used. Three features which we used are: 1) Number of connected components which are acceptable as player, 2) Maximum shirt color percent in four overlapped vertical strips in middle rectangle,



(a) Long (b) Medium

(c) In-field close-up (d) Out-of-field close-up

Figure 1. Different shot types.

3) mean of shirt color percent in two horizontal strips. They are all new and presented here for the first time. For detection of close-ups in out-of-field shots, a novel approach based on color of shirts in two horizontal strips is presented.

The rest of this paper is organized as follows. In section II the proposed method is introduced in detail. In section III experimental results are presented. Section IV gives the conclusion of this paper.

II. PROPOSED METHOD

Like most of methods, in the first step dominant color (field color) is extracted from a frame which has enough grass. For extracting dominant color, the method presented in [6] is used, in this method at first, frame is converted to the HSI format then the color mean is computed from histogram in each component. In each input frame, pixels which their cylindrical distance with this color mean is less than a threshold are considered as field pixel. If grass ratio for a frame is greater than a threshold T_{grass}, then it is given to the classifier, otherwise it is considered as out-of-field and checked if it is close-up or not. In our implementation, we set T_{grass} to 5.

A. Feature Extraction

In this section three new features which are used for classifying long, medium, and close-up shots among in-field views are introduced. For extracting the first feature minimum bounding rectangle (MBR) of grass region is obtained from a binary frame. In Fig. 3-a an original frame is shown, Fig. 3-b is its binary frame in which ones are grass pixels and zeros are non-grass pixels and Fig. 3-c shows the MBR. Now we are interested in connected components (CCs) of non-grass pixels which can be considered as $player$. Such CCs have distinct height over width ratio and rational number of pixels which is shown in (1), the number of CCs satisfying these conditions is given by (2)

$$player = \{c \epsilon C : R_{min} < \frac{H_c}{W_c} < R_{max} \quad and\ T_{min} < size_c < T_{max} \} \tag{1}$$

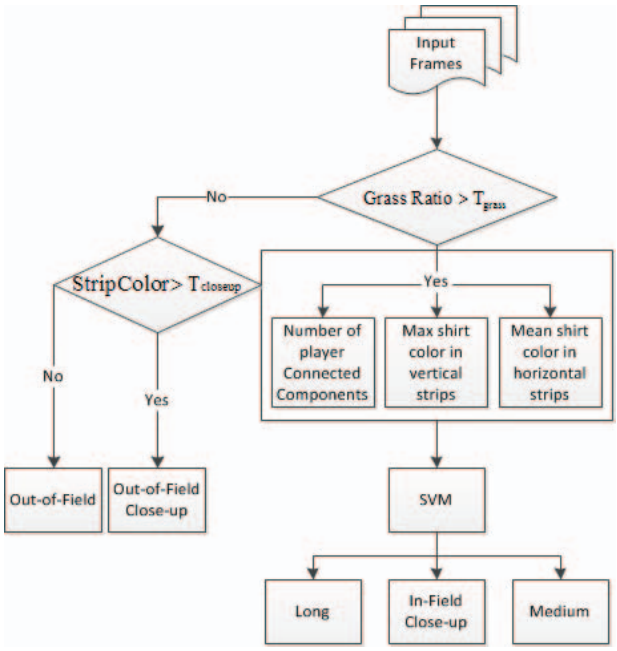


Figure 2. General structure of proposed method.

$$N_{player} = n(player). \tag{2}$$

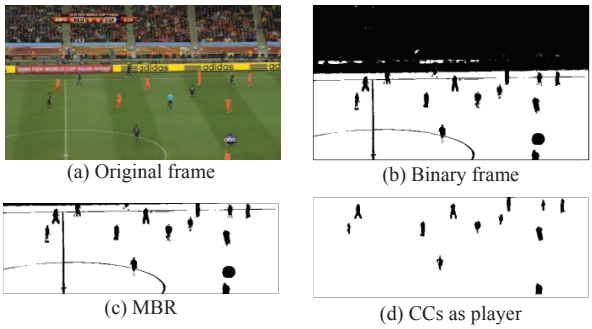
C is the set of all CCs and c is a member of it. H_c, W_c, and $size_c$ are height, width, and number of pixels of c respectively. R_{min} and R_{max} are the minimum and maximum of rational ratios for a player in long views. T_{min} and T_{max} are the minimum and maximum of acceptable sizes for a player, $player$ is a subset of CCs which can be considered as player in long view, and also $n(player)$ is the number of members of player set. According to player properties in the long shots, for the R_{min}, R_{max}, T_{min}, and T_{max} ,the values 1.5, 3.5 , 20, and 600 are assigned respectively. Fig. 3-d shows an example of acceptable CCs.

For extracting second feature, middle rectangle which contains 0.7 of whole frame is considered, and then it is divided to four overlapping vertical strips. Overlapped section of each strip is 1/12 of the original frame width. Fig. 4-a and Fig. 4-b show middle rectangle and vertical strips, respectively. Maximum percent of shirt colors in strips is computed as:

$$S = max(s_i) \quad , i = 1, \dots, 4 \tag{3}$$

where s_i is shirt color percent in i^{th} vertical strip, the shirt colors should be given to system in the beginning of the match. We gave it to our system by computing dominant color [6] from a piece of each team's shirt. S is maximum of them and the second feature.

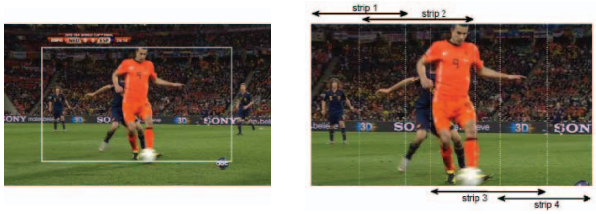
For third feature, two horizontal strips with predefined height and distance from the bottom of frame are considered; Fig. 5 shows an example of these strips. Shirt color percent of both these strips computed and the average is obtained as:



(a) Original frame (b) Binary frame

(c) MBR (d) CCs as player

Figure 3. Four steps for extracting CCs corresponding to players.



(a) Middle rectangle (b) Vertical strips

Figure 4. Representation of middle rectangle, and vertical overlapped strips.

$$H = (h_1 + h_2)/2 \tag{4}$$

where h_i is shirt color percent in i^{th} horizontal strip and H is the mean of them. In some cases, bottom of frame is filled with the labels that broadcasters add to the game such as game result, advertising, etc. using two strips reduces the effect of these cases in H. Height of strips and distance between them are obtained from the height of the original frame, in our implementation, we set them 2 and 5 percent of frame height, respectively.

B. Classifying Using SVM

Main idea of SVM is finding the hyperplane that maximizes margin for separating two classes and also is extendable for separating non-separable classes [10]. In many cases, samples which cannot be classified appropriately in the input space, in another space with higher dimensions can be classified correctly. Transformation to a space with higher dimensions is accomplished with kernels.

After extracting three features N_{player}, S, and H from train data, all of them are given to SVM classifier. In SVM, the polynomial kernel with degree 3 is used as follows:

$$k(x, z) = (x^T z + 1)^3 \tag{5}$$

where x and z are two vectors in the original space and $k(x, z)$ is dot product of them in the new space. SVM is fundamentally a two-class classifier, for classifying three classes, pair-wise approach was used. In pair-wise approach for classifying l classes, $l(l-1)/2$ different two-class SVM is needed, in [11] more details about this approach and a method for resolving unclassified regions are presented.

Figure 5. Representation of two horizontal strips, filled with black in bottom of frame.

C. Close-up Shots in Out-of-field

The frames which have low grass ratio are considered as out-of-field, among these frames we are going to separate close-ups. The approach that used for extracting third feature can be used here, after computing H from (4), close-ups and other shots are separated as follows:

$$\begin{cases} close-ups & H \geq T_{close-up} \\ other\ shots & H < T_{close-up} \end{cases} \tag{6}$$

Where $T_{close-up}$ is a threshold and we set it to 7.

III. Experimental Results

Two matches from FIFA World Cup 2010 were used in our experiment and all the input shots were downsampled in rows and columns with rate 2. The first match is between Spain and Germany with resolution 640*352 and 25 frames per second, the results of proposed method and method [6] on this match demonstrated in Table I. Since the code of method [6] was not available we implemented it ourselves. The second match is between Spain and Netherlands with resolution 624*352 and 30 frames per second, results of our method and Ekin's method [6] on this match, reported in Table II.

For training SVM, 40 shots of each class were used. Fig. 6 shows distribution of 40 training data of each class in three dimensional feature space for second match. In all the classes our accuracy is better than [6], also we separate in-field close-ups and out-of-field close-ups but in Ekin's method all the close-ups and out-of-field shots are considered as a same class, so we tested all the out-of-field and close-up shots as one class to both of the methods and the results added to Table I and Table II. Since many of the in-field close-ups have field background, the grass ratio of them is relatively high and almost in all these cases Ekin's method fails to detect shot's class correctly.

For showing our robustness, we tested our method on the match from FIFA World Cup 2002 between England and Sweden, as reported in [9]. The resolution and frame rate were not mentioned in [9], but we used a video of 640*480 resolution and 30 frames per second. Our results and the reported results in [9] both are shown in Table III. In long, medium, and close-up shots our method has sensible improvement. In out-of-field shots we are less accurate, but

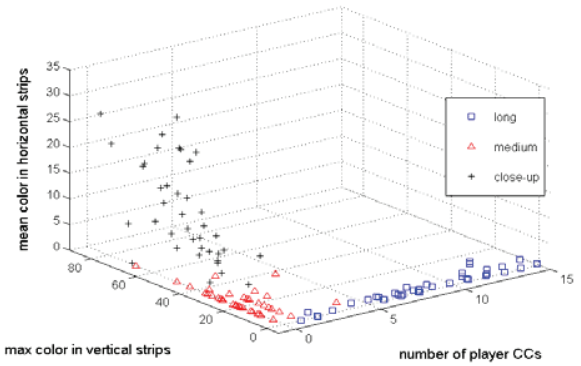


Figure 6. Representation of forty shots of each class in feature space.

TABLE I. RESULTS OF PROPOSED METHOD (PM) AND EKIN'S METHOD [6] FOR THE FIRST MATCH (SPAIN-GERMANY).

Shot Type	#of Shots	Correct		False		Recall (%)		Precision (%)	
		PM	[6]	PM	[6]	PM	[6]	PM	[6]
Long	156	151	136	0	19	96.7	87.1	100	87.7
Medium	124	121	106	9	77	97.5	85.4	93.0	57.9
Close-up (in-field)	64	60	-	3	-	93.7	-	95.2	-
Close-up (out-field)	82	77	-	1	-	93.9	-	98.7	-
All close-ups & out-fields	157	153	100	3	0	97.4	63.6	98.0	100

TABLE II. RESULTS OF PROPOSED METHOD (PM) AND EKIN'S METHOD [6] FOR THE SECOND MATCH (SPAIN-NETHERLANDS).

Shot type	#of Shots	Correct		False		Recall (%)		Precision(%)	
		PM	[6]	PM	[6]	PM	[6]	PM	[6]
Long	255	254	237	5	15	99.6	92.9	98.0	94.0
Medium	178	163	158	7	112	91.5	88.7	95.8	58.5
Close-up (in-field)	97	90	-	11	-	92.7	-	89.1	-
Close-up (out-field)	144	142	-	9	-	98.6	-	94.0	-
All close-ups & out-fields	262	255	167	11	6	97.3	63.7	95.8	96.5

the number of this type of shots is usually less than 20 per match and actually a little misclassification of this type (considered as close-ups) will not affect later processing.

IV. CONCLUSION

In this paper, we proposed a new method for classifying shots based on SVM. The features are simple and meaningful; the first one comes from connected components which can be considered as player and the other two features are related to the shirt color of players. In addition to high accuracy, the method is also real-time because features are easy to extract and input shots are downsampled. Since two of features are using color, the method is sensitive to poor quality which with using subtle algorithms for obtaining dominant color, the effect of poor quality can be compensated. This method can be used for content filtering and highlights extraction in soccer video analysis. In the future, we will work on event detection and summarization of a game.

TABLE III. RESULTS OF PROPOSED METHOD (PM) AND METHOD IN [9] FOR THE MATCH REPORTED IN [9] (ENGLAND-SWEDEN).

Shot type	Recall (%)		Precision (%)	
	PM	[9]	PM	[9]
Long	95.2	93.6	97.5	97.9
Medium	93.4	91.4	88.0	78.4
Close-up	96.8	90.7	97.8	98.0
Out-of-field	90.0	100	81.0	75.0

REFERENCES

[1] S. –Fu. Chang, "The Holy Grail of Content-Based Media Analysis," *IEEE Multimedia*, vol. 9, no. 2, pp. 6-10, 2002.

[2] L. –Y. Duan, M. Xu, Q. Tian, C. Xu, and J. S. Jin, "A Unified Framework for Semantic Shot Classification in Sports Video," Multimedia, IEEE Transactions on, vol. 7, no.6, pp.1066-1083, 2005.

[3] H. Jiang and M. Zhang, "Tennis Video Shot Classification Based On Support Vector Machine," *Computer Science and Automation Engineering (CSAE), IEEE International Conference on*, vol.2, pp. 757-761, 2011.

[4] L. Li, X. Zhang, W. Hu, W. Li, and P. Zhu, "Soccer Video Shot Classification Based on Color Characterization Using Dominant Sets Clustering," P. Muneesawang, F. Wu, I. Kumazawa, A. Roeksabutr, M. Liao, and X. Tang, Eds., *LNCS*, PCM, Springer, Berlin, vol. 5879, pp. 923-929, 2009.

[5] P. Xu, L. Xie, S. Chang, A. Divakaran, A. Vetro, and H. Sun, "Algorithms and System for Segmentation and Structure Analysis in Soccer Video," *Multimedia and Expo, IEEE International Conference on*, pp. 721-724, 2001.

[6] A. Ekin and A. M. Telkap, "Automatic Soccer Video Analysis and Summarization," *Image Processing, IEEE Transactions on*, vol. 12, no.7, pp. 796-807, 2003.

[7] Y. Zhao, Y. Cao, L. Zhang, and H. Zhang, "An SVM-Based Soccer Video Shot Classification," *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, pp. 5398-5403, 2005.

[8] M. H. Kolekar and K. Palaniappan, "A Hierarchical Framework for Semantic Scene Classification in Soccer Sport Video," *IEEE Region 10 Conference (TENCON)*, pages 1-6, 2008.

[9] X. Tong, Q. Liu, and H. Liu, "Shot Classification in Broadcast Soccer Video," *Electronics Letter On Computer Vision and Image Analysis (ElCIVIA)*, vol. 1, pp. 16-25, 2008.

[10] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, pp. 121-167, 2000.

[11] J. C. Platt, N. Cristianini, and J. Shawe-Taylor, "Large Margin DAGs for Multiclass Classification", S. A. Solla, T. K. Leen, and K. –R. Muller, Eds., *Advanced in Neural Information Processing Systems*, MIT Press, pp. 547-553, 2000.