Flexible Soccer Video Summarization in Compressed Domain

Vahid Kiani
Computer Engineering Department
Ferdowsi University of Mashhad
Mashhad, Iran
vahid.kiani@rocketmail.com

Hamid Reza Pourreza
Computer Engineering Department
Ferdowsi University of Mashhad
Mashhad, Iran
hpourreza@um.ac.ir

Abstract—**Due to vast amount of archived video content around the world, flexible and efficient soccer video summarization systems are required. In this paper, we present an efficient yet accurate framework for soccer video summarization in encoded MPEG videos. A novel logo transition detection method based on spatio-temporal template is proposed. Semantic features along with replays and audio activity are exploited for summary generation based on user defined parameters and rank based video summarization scheme. We conducted objective and subjective evaluations on ten hours of broadcast soccer videos. Rank correlation measures are used for objective video summary evaluation. Our system achieved satisfactory accuracy and speed of nearly two times faster than real-time.**

Keywords- video signal processing, compressed-domain analysis, soccer video summarization, replay detection, rank-based summarization.

I. INTRODUCTION

Video is a rich media in modern computer era due to presenting synchronized visual, acoustic, and textual information. Video summarization systems are inevitable part of any video browsing/retrieval system. A video summarization system provides a compact representation of the input video stream by preserving only most important segments [1]. Two main approaches for soccer video summarization are: highlight detection, and event detection.

In highlight detection, important segments of the video stream are detected without understanding underlying events occurred in the game. In this approach, most researchers tried to extract highlights by detecting special views like replays [2, 3], goal-mouth scenes, and exciting scenes [4]. Then several summaries are generated using a special subset of different highlight types. Another approach is to formulate an importance curve where an importance score/level is assigned to each segment of the input video [5]. Then, video segments with importance higher than a constant or adaptive threshold are included in the output summary [6, 7]. The idea of rank based summarization is also used in highlight detection. In this approach, an integer value called rank is assigned to each segment of the video based on its importance. Then, video segments with highest rank are concatenated to generate game summary [8-10].

In event detection approaches, common events of soccer such as goal, shot-on-goal, yellow/red card, corner kick, free kick, and penalty kick are identified and important events of the game are incorporated in the final summary. Users can generate personalized summaries by choosing a subset of several event types for inclusion in the summary [11, 12].

Modern video compression standards try to remove spatial and temporal redundancies in the image sequence to provide a compact representation. Large amount of compressed videos around the world provide a low cost information source for video analysis [13]. Consequently, some researchers focused on compressed domain analysis to make current video indexing, summarization, and retrieval systems faster. Leonardi [14] presented a goal and kick event detection framework for compressed soccer videos with low accuracy in event detection. Sadlier [15] proposed a more accurate goal detection method for compressed soccer videos. In our previous work, we presented DC sequence extraction and motion extraction methods for compressed MPEG-1 bit-stream [16]. Motion vector information, macro-block type information, grass color modeling module and camera motion estimation modules are described in [16, 17] and used here.

In this paper, we propose an internal soccer video summarization system based on content-aware audio-visual features. An overview of the proposed system is shown in Fig. 1. The rest of the paper is organized as follows. Video segmentation is explained in section II. Shot classification and replay detection described in section III. Section IV is concentrated on details of summary unit segmentation. Rank based video summarization is presented in section V. Section VI presents experimental results of our method. Finally, Section VII concludes the paper and suggests future works.

II. VIDEO SEGMENTATION

Shot boundary detection is a preliminary part of all video processing systems. Abrupt and dissolve transitions are detected based on visual features by using feed forward neural networks. Several researchers proposed promising methods to detect abrupt and dissolve transitions [11, 18-21]; therefore we do not describe details of these modules.

A. Logo Transition Detection

To detect logo transitions, a spatio-temporal template for is constructed by temporal averaging discriminative visual features. Next, logo transitions are detected in test

978-1-4799-2093-8/13/$31.00 ©2013 IEEE

videos by computing two matching scores and a simple classifier.

The spatio-temporal template is created by averaging of inra-coded MB ratio and histogram features as follows:

$$\hat{H}_c^m(t) = \frac{1}{N}\sum_{i=1}^{N} H_c^m(S_i + t) \quad (1)$$

$$\hat{R}_I(t) = \frac{1}{N}\sum_{i=1}^{N} R_I(S_i + t) \quad (2)$$

Where $H_c^m(t)$ denotes value of m'th bin in the histogram of color component c (Y,Cb,Cr), R_I denotes ratio of intra-coded MBs, S_i denotes start frame of i'th logo transition, and N is the number of logo transitions in training data. Temporal length of the template is considered to be 15 pictures for our dataset.

Logo transitions are detected in test videos by matching extracted color and motion feature sequence with temporal logo template. The color and motion similarity features are defined as follows:

$$S_H(t) = 1 - \frac{1}{45}\sum_{c\in(Y,Cb,Cr)}\sum_{i=0}^{14}\sum_{m=0}^{63}\left|H_c^m(t+i) - \hat{H}_c^m(i)\right| \quad (3)$$

$$S_I(t) = 1 - \frac{1}{15}\sum_{i=0}^{14}\left|R_I(t+i) - \hat{R}_I(i)\right| \quad (4)$$

A feed forward neural network classifier is used to detect logo transitions based on defined similarity features. This network consists of only one hidden layer with three neurons. Back Propagation learning algorithm is used to train the classifier.

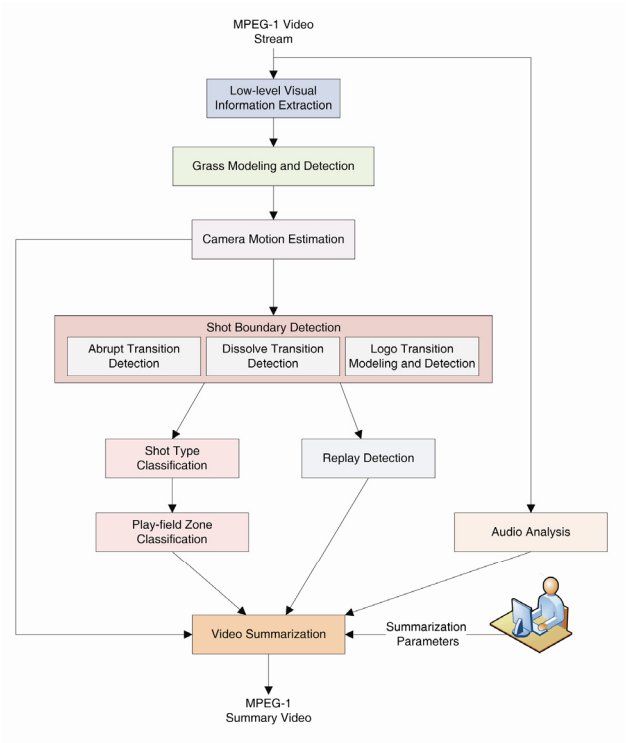

Figure 1. Rank based soccer video summarization framework

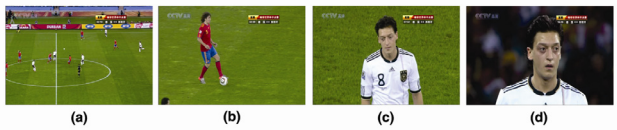

Figure 2. Four shot classes in soccer video; (a) Long shot; (b) Medium shot; (c) Close-up shot; (d) Out-of-field shot

R_1	R_2	R_3
R_4	R_5	R_6
R_7	R_8	R_9

Figure 3. Decomposition of each picture to nine local regions

III. Semantic Shot Classification

Classification of video shots into several semantic shots brings a higher level of intelligence to the whole summarization system. We try to detect long, medium, close-up, out-of-field, and replay shots. In addition, we divide long shots into smaller temporal units based on playing field-zone displayed on the screen.

A. Shot Classification

As shown in Fig. 2, there are four semantic shot types in soccer namely Long, Medium, Close-up, and Out-of-field [11]. Shot class determined by averaging following features during each shot:

1) Normalized luminance histogram difference.
2) Mean of MVs magnitude in each P-Picture.
3) Grass ratio of each picture.
4) Grass ratio of central region: to compute this feature, each picture is divided to nine regions with 3:5:3 proportion in both directions [11] as shown in Fig. 3.
5) Grass ratio difference: The mean value of grass ratio difference between region R_5 and R_4, and between R_5 and R_6, defined in [11] as:

$$G_{Diff}(t) = \frac{1}{2}\left(\left|G_5(t) - G_4(t)\right| + \left|G_5(t) - G_6(t)\right|\right) \quad (5)$$

Where $G_k(t)$ denotes grass ratio in region R_k of t'th picture.

6) Grass density: An algorithm is applied on DC image of each picture to extract the field region. Then, grass ratio of field region is computed.
7) Object ratio: This feature examines biggest object ratio in each picture of the sequencce. We try to estimate the size of the biggest object on the screen by extracting the biggest smooth region in I-Pictures of each shot. Macro-blocks in DC image of each I-Picture are segmented into smooth regions by using a simple horizontal scanning based segmentation method in YCbCr color space. All coincident regions with grass are removed from segmentation results.

A classification method based on a hierarchy of SVM classifiers is proposed in this section [22]. Feature of each shot are fed into a SVM hierarchy shown in Fig. 4. In each SVM classifier a linear kernel is used. Training and

978-1-4799-2093-8/13/$31.00 ©2013 IEEE

test data are normalized in both training and test phase to avoid classification errors caused by wide range of values in some features.

B. Replay Detection

Replay segments in sport videos cover most important events of the game. A replay segment is composed from one or more shots representing an event in slow-motion speed. In our system, candidate replay segments are determined by pairing consecutive detected logo transitions. Due to inaccuracy in logo transition detection, each segment between two consequent detected logo transitions can be a live or replay segment. We propose following cinematic features for each candidate replay segment:

1) Length: The length of segment.
2) Length ratio: The ratio between length of current segment and maximum length of its neighboring segments.
3) Number of transitions: Number of abrupt transitions in each segment.

These features are fed into a feed forward neural network classifier similar to logo transition detection section with a hidden layer of three neurons.

C. Play-field Zone Classification

To classify play-field zone, we use a method similar to our previous work [16] on DC image of each picture. Each picture is divided to bottom region (RB), upper left region (RUL), and upper right region (RUR). Then, ratio of grass in each of these regions is computed using DC image and denoted by G_{RB}, G_{RUL}, G_{RUR} respectively. Finally, a rule-based system shown in Fig. 5 is used on each picture of the video stream. We used $Th_{FieldSide}$=0.1 and Th_{Long}=0.7 in our experiments.

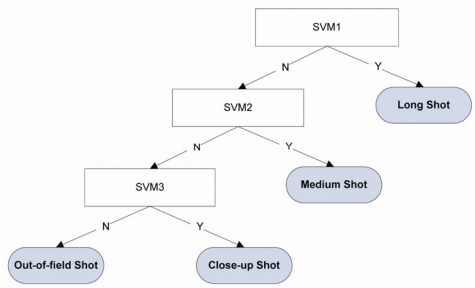

Figure 4. Hierarchy of SVM classifiers for shot classification

---

IF $G_{RB}>Th_{Long}$ AND $|G_{RUL}-G_{RUR}|>Th_{FieldSide}$ AND $G_{RUR}>G_{RUL}$ THEN
FieldZone is Left-Side
ELSE IF $G_{RB}>Th_{Long}$ AND $|G_{RUL}-G_{RUR}|>Th_{FieldSide}$ AND $G_{RUR}<G_{RUL}$ THEN
FieldZone is Right-Side
ELSE IF $G_{RB}>Th_{Long}$ AND $|G_{RUL}-G_{RUR}|<=Th_{FieldSide}$ THEN
FieldZone is Mid-Field
ELSE
FieldZone is None

Figure 5. Rule-based algorithm for play-field zone classification in soccer videos

IV. Summary Unit Segmentation and Description

The shot unit is the most common temporal unit for video summarization applications. In soccer videos, important medium and close-up shots represent at most one event in the game. But, long shots have longer duration and can cover multiple events. An important event often lasts for a short interval during a lengthy long shot. Therefore, Considering shot unit as the basic summary unit could result in a lengthy or poor summary.

We divide long shots to smaller temporal units called view. Each medium, close-up, or out-of-field shot is considered as one view unit as a whole. Each long shot is divided into several view units when field-zone changes. Fig. 6 demonstrates this segmentation process.

A multi-modal set of features is used to describe summary units and discriminate important scenes from normal scenes. Specifically, view-type reflects event type occurred in the summary unit. Motion and audio features determine exciting degree of the summary unit.

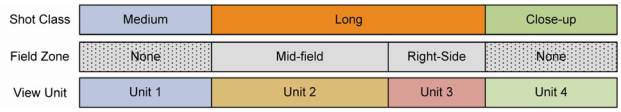

Figure 6. View unit segmentation based on shot class and field view

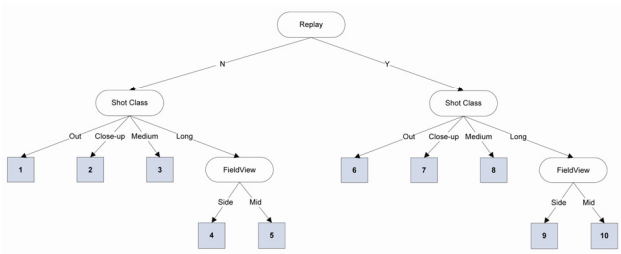

Figure 7. Decision tree for view-type classification

A. View-type Classification

Some types of events occur only in special play-field zones. In addition, replays cover most important events of the game. Combining these assumptions, we define ten views as follows:

1) Live out-of-field view: often shows spectators and coaches.
2) Live close-up view: often shows players and coaches.
3) Live medium view: often shows players fighting and actions.
4) Live field-side view: somehow shows near-goal events like goal, shot-on-goal, and offside.
5) Live mid-field view: somehow shows foul events.
6) Replay out-of-field view: often shows spectators and coaches.
7) Replay close-up view: often shows players who had main role in last event.
8) Replay medium view: often shows an event from closer view.
9) Replay field-side view: often shows near-goal events like goal, shot-on-goal, and offside.
10) Replay mid-field view: rarely used, and shows fouls occurred in the mid-field.

978-1-4799-2093-8/13/$31.00 ©2013 IEEE

The view-type of each summary unit is denoted by V(k) and computed using a decision tree shown in Fig. 7.

B. Audio Activity

Audio energy level in large portion of a summary unit could be low. When an important event occurs, audio energy level suddenly increases and remains high for a short time. Therefore, we consider the maximum value of normalized audio energy in each summary unit as second descriptor and define it as:

$$f_1(k) = \max_{t \in unit_k} \left(A(t) \right) \tag{6}$$

Where A(t) denotes normalized audio energy feature, k denotes index of summary unit, and t denotes index of pictures contained in the summary unit.

C. Motion Activity

During an important event, motion activity increases due to fast camera and objects movements. A common representative feature for motion activity is mean of motion magnitude. To make our method robust against director's production style, the mean of motion magnitude feature is normalized for each video stream and considered as third summary unit descriptor as follows:

$$f_2(k) = \frac{M(k)}{\max_j \left(M(j) \right)} \tag{7}$$

Where M(k) denotes mean motion magnitude feature, and k denotes index of summary unit.

D. Camera Motion Activity

Important events are accompanied with fast pan, tilt and fast zoom-in or zoom-out camera motions. The mean of camera pan, and tilt factors magnitude is computed for each summary unit as $C_p(k)$ and $C_t(k)$ respectively. The mean of camera zoom in/out factor magnitude is also computed for each summary unit as:

$$C_z(k) = \frac{1}{L_k} \sum_{t \in unit_k} \left| zf(t) - 1 \right| \tag{8}$$

Where L(k) denotes length of summary unit, and zf(t) denotes camera zoom factor.

These features are normalized for each video stream as:

$$f_3(k) = \frac{C_p(k)}{\max_j \left(C_p(j) \right)} \tag{9}$$

$$f_4(k) = \frac{C_t(k)}{\max_j \left(C_t(j) \right)} \tag{10}$$

$$f_5(k) = \frac{C_z(k)}{\max_j \left(C_z(j) \right)} \tag{11}$$

V. Video Summarization

In our audiovisual soccer video summarization system, a summary is created by computing an importance score for each summary unit and ranking summary units based on their scores. Then, highest ranked summary units are concatenated to produce a video summary with desired length constraint.

The importance score of each summary unit is computed by weighted fusion of summary unit descriptors as follows:

$$ImportanceScore(k) = \alpha \bullet \lambda\left(V(k)\right) + \beta \bullet \sum_{j=1}^{5} w_j \cdot f_j(k) \tag{12}$$

Where λ denotes importance value of each view-type and lies between zero and one. The weight w_j denotes contribution coefficient of corresponding feature f_j in computation of the importance score. Finally, α and β are two constants.

The ordinal ranking policy used to rank summary units by their importance score values. Then, highest ranked summary units are inserted into the summary, until the desired length constraint of the summary is satisfied. For two summary units with the same importance score, precedence is given to summary unit which came earlier in the video stream.

TABLE I. Specification of FUM-BSVD Dataset

No.	Video Name	Video Length
1	Germany vs. Argentina	92 mins.
2*	Germany vs. England	93 mins.
3	Germany vs. Spain	94 mins.
4*	Germany vs. Uruguay	94 mins.
5*	Greece vs. Argentina	94 mins.
6	Slovakia vs. Italy	99 mins.

* Star symbols in first column indicate manually ranked videos

VI. Experimental Results

A collection of twelve videos from six games of World Cup 2010 are gathered and annotated to examine our method. Table 1 shows details of this dataset called FUM-BSVD. Shot boundary, shot class, and replay segments are tagged for all videos in the dataset by hand. In addition, all summary units are ranked manually by five subjects for six videos in the dataset indicated with star mark in Table I. All videos in the dataset are compressed by FFMPEG encoder into MPEG-1 format. Compression parameters are shown in Table II.

For logo and replay detection, training samples are gathered from first sequence of the dataset. Results of the proposed method are compared to manually tagged boundaries in terms of precision and recall. Table III shows performance of the proposed method in logo and replay detection on the overall dataset. Comparing to [23-25] our results are promising.

Similarly, training data for shot classification gathered from first video of dataset. Performance of the proposed method is measured in terms of accuracy [11]. Table IV shows performance of the proposed method in shot classification. In this experiment, most classification errors

978-1-4799-2093-8/13/$31.00 ©2013 IEEE

are due to classification of a close-up shot as medium and vice-versa.

TABLE II. VIDEO COMPRESSION PARAMETERS FOR FUM-BSVD DATASET

Parameter	Value
Compression Standard	MPEG-1
GOP Size	16
Num of B-Frame in each GOP	0
Frame-rate	25 fps
Bit-rate	10240 kbps
Picture Size	720x480
Motion Vector Search Radius	100 pixels

TABLE III. LOGO AND REPLAY DETECTION RESULTS

Type	TP	FP	FN	Precision	Recall
Logo	669	12	0	98%	100%
Replay	335	3	2	99%	99%

TABLE IV. PERFORMANCE OF SHOT CLASSIFICATION

Dataset	Correct	False	Accuracy
Training	463	29	94%
Test	3626	305	92%

In this paper, we use quantitative and qualitative evaluations to examine performance of the video summarization module. In our experiments, parameters α and β are set to 5 and 2 respectively. A graphical user interface is designed to let the user define other parameter values manually. Summarization parameters could also determined by an automatic learning system based on user behaviour in a video retrieval system.

Video summarization is a tasty process and strongly depends on user preferences and point of view. Therefore, comparison of generated summaries with fixed ground truth summaries in terms of precision and recall could not exhibit accuracy of the system in the general sense [5, 11]. In contrast, we use a rank based evaluation approach for quantitative evaluation of proposed system. In this approach importance score of all game scenes involves in the comparison. We use two rank correlation coefficient measures namely Spearman and Tau-B. Six videos of FUM-BSVD dataset marked with star symbol in Table I are given to five human subjects. Each subject assigned a rank from one to ten to each summary unit of these six videos. Rank of one indicates most important summary units, and rank of 10 indicates lowest ranked units. For each summary unit, the mean of ranks assigned by five subjects is computed and considered as mean manual score. The proposed method is applied on each of these videos and automatic ranking results are compared to mean manual ranks. Table V shows Spearman and Tau-b correlation coefficient values.

For qualitative evaluation, a summary set with length of 2, 3, 5, and 10 minutes are generated for each video stream in the FUM-BSVD dataset. Twenty subjects aging from 22 to 28 are invited to participate in this evaluation. In each evaluation case, one summary set is evaluated by one subject. A total of 46 evaluation cases are done in this experiment. Each subject evaluated summary set of at least one video stream in the FUM-BSVD. Subjects are asked to evaluate each summary video in the summary set using three attributes: completeness, smoothness [26], and conciseness. Completeness measures whether summary video contains all important events of the game. Smoothness measures whether consecutive shots in the summary video are smoothly and rationally changed. Conciseness measures whether summary video lacks unimportant scenes of the game. Each subject gives a score of five (strongly accept), four (accept), three (marginally accept), two (reject), and one (strongly reject) to each attribute. Table VI shows collected average scores. According to experiment results, the completeness attribute of summary videos is strongly satisfied for all summary durations.

TABLE V. COMPARISON OF MANUAL AND AUTOMATIC RANKS

Video Name	Half	Spearman	Tau-b
Germany vs. England	1st	0.7543	0.5869
Germany vs. England	2nd	0.8161	0.6452
Germany vs. Uruguay	1st	0.8218	0.6418
Germany vs. Uruguay	2nd	0.8644	0.6751
Greece vs. Argentina	1st	0.8302	0.6565
Greece vs. Argentina	2nd	0.8134	0.6306
Average		**0.8167**	**0.6394**

TABLE VI. QUALITATIVE EVALUATION RESULTS

Summary Length	Completeness	Smoothness	Conciseness
2 minutes	4.43	4.28	4.32
3 minutes	4.60	4.10	4.06
5 minutes	4.73	3.82	3.39
10 minutes	4.73	3.26	2.52

VII. CONCLUSION

In this paper, we presented a multi-modal content-based video summarization framework for compressed soccer videos. A complete set of robust mid-level feature extraction methods are also proposed to attain satisfactory accuracy. Rank based summarization in our framework made it possible to generate flexible summaries and carry out reliable objective evaluations. While the MPEG-1 is an old compression standard, it shares common compression techniques like motion compensation with MPEG-2, H.264, and H.265. Therefore, our framework could be extended to other compression formats easily.

REFERENCES

[1] A. G. Money and H. Agius, "Video summarisation: A conceptual framework and survey of the state of the art," *Journal of Visual Communication and Image Representation,* vol. 19, pp. 121-143, 2008.

[2] B. Li, H. Pan, and I. Sezan, "A general framework for sports video summarization with its application to soccer," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003, pp. III-169-72 vol.3.

[3] H. Pan, P. van Beek, and M. I. Sezan, "Detection of slow-motion replay segments in sports video for highlights generation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001, pp. 1649-1652 vol.3.

[4] D. Tjondronegoro, Y.-P. P. Chen, and B. Pham, "Integrating Highlights for more complete sports video summarization," *IEEE Multimedia Magazine,* vol. 11, pp. 22-37, 2004.

[5] B. T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification," *ACM Transactions on Multimedia Computing, Communications, and Applications,* vol. 3, p. 3, 2007.

[6] Z. Zhao, S. Jiang, Q. Huang, and Q. Ye, "Highlight Summarization in Soccer Video based on Goalmouth Detection," in *Asia-Pacific Workshop on Visual Information Processing (VIP06)*, Beijing, China, 2006.

[7] Z. Zhao, S. Jiang, Q. Huang, and G. Zhu, "Highlight Summarization in Sports Video Based on Replay Detection," in *IEEE International Conference on Multimedia and Expo*, 2006, pp. 1613-1616.

[8] X. Tong, Q. Liu, Y. Zhang, and H. Lu, "Highlight ranking for sports video browsing," in 13th annual ACM international conference on Multimedia, Hilton, Singapore, 2005.

[9] C. Liu, Q. Huang, S. Jiang, L. Xing, Q. Ye, and W. Gao, "A framework for flexible summarization of racquet sports video using multiple modalities," *Computer Vision and Image Understanding,* vol. 113, pp. 415-424, 2009.

[10] M. A. Refaey, W. Abd-Almageed, and L. S. Davis, "A Logic Framework for Sports Video Summarization Using Text-Based Semantic Annotation," in *Third International Workshop on Semantic Media Adaptation and Personalization*, 2008, pp. 69-75.

[11] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Transactions on Image Processing,* vol. 12, pp. 796-807, 2003.

[12] H. M. Zawbaa, N. El-Bendary, A. E. Hassanien, and A. Abraham, "SVM-based soccer video summarization system," in *Third World Congress on Nature and Biologically Inspired Computing*, 2011, pp. 7-11.

[13] T. Sikora, "Trends and Perspectives in Image and Video Coding," *Proceedings of the IEEE,* vol. 93, pp. 6-17, 2005.

[14] R. Leonardi and P. Migliorati, "Semantic indexing of multimedia documents," *IEEE Multimedia Magazine,* vol. 9, pp. 44-51, 2002.

[15] D. A. Sadlier and N. E. O'Connor, "Event detection in field sports video using audio-visual features and a support vector Machine," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 15, pp. 1225-1233, 2005.

[16] V. Kiani and H. R. Pourreza, "An Effective Slow-motion Detection Approach for Compressed Soccer Videos," *ISRN Machine Vision,* vol. 2012, p. 8, 2012.

[17] V. Kiani and H. R. Pourreza, "Robust GME in encoded MPEG video," in 9th International Conference on Advances in Mobile Computing and Multimedia, 2011, pp. 147-154.

[18] U. Gargi, R. Kasturi, and S. H. Strayer, "Performance characterization of video-shot-change detection methods," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 10, pp. 1-13, 2000.

[19] B. Han, Y. Hu, G. Wang, W. Wu, and T. Yoshigahara, "Enhanced Sports Video Shot Boundary Detection Based on Middle Level Features and a Unified Model," *IEEE Transactions on Consumer Electronics,* vol. 53, pp. 1168-1176, 2007.

[20] K. Hoashi, M. Sugano, M. Naito, K. Matsumoto, F. Sugaya, and Y. Nakajima, "Shot Boundary Determination on MPEG Compressed Domain and Story Segmentation Experiments for TRECVID 2004," in *TREC Video Retrieval Evaluation (TRECVID)*, 2004.

[21] X. Jiang, T. Sun, J. Li, and X. Chen, "A Novel Shot Edge Detection Algorithm based on Chi-square Histogram and Macro-block Statistics," in *International Symposium on Information Science and Engineering*, 2008, pp. 604-607.

[22] Y.-H. Zhou, Y.-D. Cao, L.-F. Zhang, and H.-X. Zhang, "An SVM-based soccer video shot classification," in *International Conference on Machine Learning and Cybernetics*, 2005, pp. 5398-5403.

[23] Q. Huang, J. Hu, W. Hu, T. Wang, H. Bai, and Y. Zhang, "A Reliable Logo and Replay Detector for Sports Video," in *IEEE International Conference on Multimedia and Expo*, 2007, pp. 1695-1698.

[24] Z. Dang, J. Du, Q. Huang, and S. Jiang, "Replay Detection Based on Semi-automatic Logo Template Sequence Extraction in Sports Video," in *Fourth International Conference on Image and Graphics*, 2007, pp. 839-844.

[25] M. Y. Eldib, B. Zaid, H. M. Zawbaa, M. El-Zahar, and M. El-Saban, "Soccer video summarization using enhanced logo detection," in *16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 4345-4348.

[26] C. Xu, J. Wang, H. Lu, and Y. Zhang, "A Novel Framework for Semantic Annotation and Personalized Retrieval of Sports Video," *IEEE Transactions on Multimedia,* vol. 10, pp. 421-436, 2008.

978-1-4799-2093-8/13/$31.00 ©2013 IEEE