Temporal Multiple Correspondence Analysis for Big Data Mining in Soccer Videos

Yimin Yang, Shu-Ching Chen
School of Computing and Information Sciences
Florida International University
Miami, FL 33199, USA
Email: {yyang010,chens}@cs.fiu.edu

Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL 33124, USA
Email: shyu@miami.edu

***Abstract*—A multimedia big data mining framework consisting of two phases for interesting event detection in soccer videos has been proposed in this paper. In the pre-processing phase, it utilizes the multi-modal multi-filtering content analysis techniques for shot boundary detection and feature extraction. A pre-filtering process based on domain knowledge analysis is then applied to clean the noise and obtain a candidate set. In the event detection phase, a temporal multiple correspondence analysis (TMCA) algorithm that adopts an indicator weighting scheme is proposed to efficiently and effectively incorporate the temporal semantic information for improving the detection results. Furthermore, another enhanced MCA (EN-MCA) approach is presented to better capture the correspondence between feature items and classes by thoroughly utilizing the pair-wise principal components. Finally, a re-ranking procedure is performed to retrieve the missed interesting event. Our proposed semantic re-ranking framework is evaluated on a large collection of soccer videos for interesting event detection. The experimental results demonstrate the effectiveness of the proposed framework.**

***Keywords*-Big data; multimedia big data mining; event detection; temporal MCA; re-ranking**

I. Introduction

As we are stepping into the big data era, the volume of multimedia data, especially videos, has been growing enormously, from private digital video to broadcasting programs. How to efficiently and effectively process such great amounts of video data to meet users' interest in near real time is a big challenge. Researchers in both academia and industry have been seeking solutions to conquer this challenge. Some initial attempts include the development of Apache Mahout [1], a scalable machine learning and data mining open source software based mainly on Hadoop. However, the question still remains as to how to establish an integrated and automatic framework to process and analyze videos so that the users can efficiently search and browse the interested content. Hence, interesting event detection has attracted a lot of attention with the usage of high-level indexing and selective video browsing [2].

Videos contain rich multi-modal information such as visual, audio, and textual. Multi-modal approaches become more and more popular since different modalities contribute to interesting event detection from various aspects [3, 4, 5]. In [3], a multi-modal framework is utilized to leverage the audio/visual/text features for the purpose of goal detection. However, due to the limitation of text availability, the framework does not always benefit from text semantic information. In [4], visual clues are extracted for the usage of shot segmentation, shot classification, and goal detection. Then the audience's cheering and the commentator's excited speech are extracted as the audio clues. At the end, both visual and audio values are combined with the domain knowledge of soccer videos to define goal event detection rules.

In addition to the multi-modal features, temporal information is also a critical clue for analyzing potential interesting events. For example, a typical goal shot in a soccer game is usually followed by one or multiple close-up shot, multi-player shot, and audience shot. However, there is no strict order for these temporal patterns. In other words, the temporal information has a loose structure. The well representation and utilization of these temporal semantic features will greatly facilitate the detection of interesting events in sports videos. In this study, an indicator weighting method is proposed to incorporate the extracted temporal semantic features to improve the interesting event detection performance.

In the multimedia information retrieval society, a re-ranking process is usually deployed to improve the retrieval results by utilizing auxiliary information, such as new features or additional models. This idea also applies to the interesting event detection task, as it is known that no one single model always performs well for all types of data sets or even a different set of features. This observation motivates us to introduce a re-ranking framework for interesting event detection, which takes advantages of different types of features and multiple models to improve the detection results. Specifically, the contributions of this work include:

1) A TMCA algorithm is proposed to incorporate the well designed temporal semantic features by using an indicator weighting scheme.
2) An EN-MCA method is presented to explore feature item association in more details, thus capturing more semantic information.
3) An integrated multimedia big data mining framework is developed to effectively detect and retrieve interest-

978-1-4799-8688-0/15 $31.00 © 2015 IEEE
DOI 10.1109/BigMM.2015.88

ing events from soccer videos.

The remainder of this paper is organized as follows. Section II introduces the existing work. Section III discusses the proposed semantic re-ranking framework in details, including the pre-processing, indicator weight generation, enhanced MCA weight calculation, and the re-ranking process. Finally, the experimental analysis is presented in section IV, and section V concludes the paper.

II. Related work

Based on different types of features used for video event detection, the related work can be classified into the following categories: (1) Audio-based methods [6, 7]: in some early approaches, only audio features are analyzed for video event detection. For example, in [6], Xu *et al.* developed the mid-level audio keywords for event detection in soccer videos. In [7], Rui *et al.* used audio features alone for detecting hits and generating baseball highlights. (2) Visual-based methods [8, 9]: visual information is one of the most important clues for video content analysis and is usually the first choice for event detection. In [8], a group of mid-level visual features were proposed to present the characteristics of a view, such as view label, motion descriptor and shot descriptor. In another work [9], wang *et al.* developed a set of descriptors based on low-level visual features for soccer highlight extraction, namely field color descriptor, player size descriptor, goal area descriptor, and midfield descriptor. (3) Multi-modal fusion methods [10, 11, 12, 13, 14, 15]: as mentioned before, it is a good strategy to integrate multi-modal features for better performance. Most of the existing frameworks fall into this category. Audio and visual data are usually combined for event detection in multiple genres of field sports including soccer, rugby, hockey, and Gaelic football [10, 11, 12]. In [13], Xu *et al.* exploited web-casting text crawled from famous sports websites to assist soccer video event detection. There are also studies conducting event detection by applying collaborative analyses of the textual, visual, and audio modalities [14, 15].

Different levels of features (i.e., low-level, mid-level, and high-level) created from multiple modalities are usually coupled with various machine learning and data mining models for event detection. Specifically, A two-layer hierarchical SVM classifier was proposed to perform mid-level audio classification in [11]. The fixed temporal structure of views was used in exploring an SVM-based incremental method to improve the extensibility of view classification and event detection [8]. The temporal pattern of mid-level keyword sequences was analyzed by the HMM classifer to detect high-level semantics [12]. In [16], Assfalg *et al.* proposed two approaches for soccer highlight detection based on HMMs using only motion information or the combination of player location information. Wang *et al.* [17] presented a three-level framework that employs Conditional Random Fields (CRFs) to fuse temporal multi-modal cues for event detection. Chen *et al.* [18] extended the traditional association rule mining algorithm and presented a hierarchical temporal association mining approach to adapt video event analysis. In other studies, the subspace-based multimedia data mining framework using decision trees was proposed for rare event detection [19, 20].

Despite all these studies on video event detection, there is limited work analyzing and utilizing temporal semantic information. Some initial attempts were described in [21], where a temporal pattern analysis step was conducted to systematically search for the optimal temporal patterns that are significant for characterizing the events. In addition, there is also lack of research on how to incorporate re-ranking or post-processing technique(s) for interesting event detection, which motivates us to develop the proposed framework.

III. Semantic Re-ranking Framework

Depicted in Fig. 1 is the proposed framework composed of 2 phases. In phase I, pre-processing is performed, which includes three sub-routines, namely automatic shot boundary detection, low-level multi-modal feature extraction, and pre-filtering. The output of phase I is the remained candidate instances which contain potential interesting events. In phase II, the candidate set is passed through a classification model, obtaining initial classification results, and then a set of ranking models are applied to retrieve the basic ranking scores. At the same time, the semantic features are extracted for generating semantic scores using the proposed indicator weighting algorithm, which will be combined with the basic ranking score and utilized to determine the final interesting events.

A. Pre-processing

In this work, a video shot is considered as the basic unit for interesting event detection. Therefore, the first step of pre-processing is shot boundary detection, which provides the shot boundaries used for video feature extraction. In our previous work [22], an effective and unsupervised multi-filtering method was proposed, which includes three filters: pixel-level filter, histogram filter and segmentation filter. In this multi-filtering architecture, the histogram filter can be incorporated into the traditional pixel-level comparison to compensate each other and to reduce the number of false positives. Furthermore, because object segmentation and tracking techniques perform especially well on detecting luminance changes and object motion, they are placed as the last filter for determining the actual shot boundaries when both pixel-level and histogram comparisons fail.

Multi-modal features have been proved to be effective for video content analysis. In [23], a total of 17 features were extracted for each shot, including 12 audio features and 5 visual features. The audio features can be classified into three groups: volume-based, energy-based, and spectrum-flux-based features. As for the visual features, they could

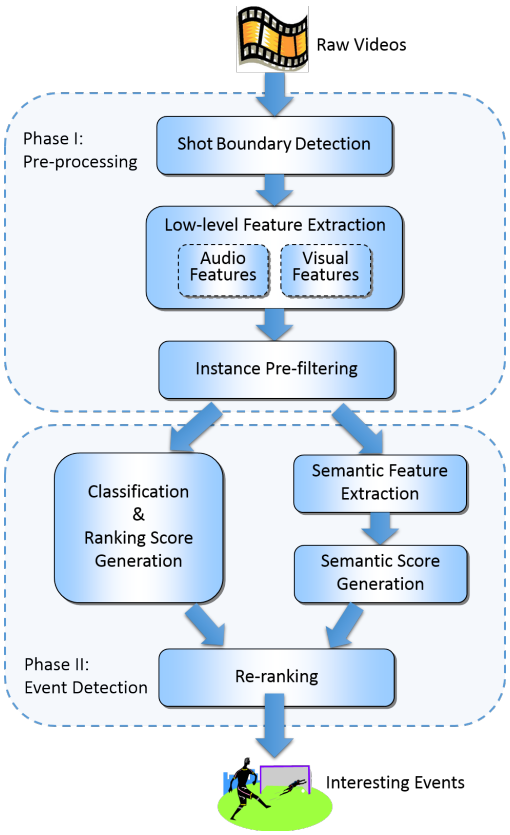

Figure 1: Semantic re-ranking framework.

be further categorized into two groups: pixel-level and mid-level features. The mid-level features consist of *grass_ratio*, *back_var*, and *next_grass*. The *grass_ratio* feature denotes the average percent of grass areas in a video shot, which is a very critical feature for identifying goal shots since most of them are of global or mid-view where the green field occupies large areas of a screen. The *back_var* feature is generated by the segmentation filter. A low *back_var* value means a relatively smooth background, such as the play field in soccer games. Finally, the *next_grass* feature represents the average percent of grass areas for the successive shot of the one being processed. It is worth noting that the mid-level features imply certain semantics. For example, most interesting events are followed by one or two close-up view shots, which are often of low grass ratio. Therefore, *next_grass* should be a low value.

Once the video features and audio features have been properly extracted, the data mining techniques can be applied to retrieve the interesting events. However, the data amount is typically huge and the ratio of the interesting event to non-interesting events is less than 1:100 in our study. As the first attempt to solve the class-imbalance issue, the same major observation rules in [23] are carried out to effectively pre-filter the data and enhance the precision of

(a) Interesting event (b) Successive shot 1 (c) Successive shot 2

(d) Successive shot 3 (e) Successive shot 4 (f) Successive shot 5

Figure 2: Examples of semantics.

mining interesting events.

B. Interesting Event Detection

As mentioned before, the semantic information could be useful for identifying interesting events. The problem is how to appropriately represent the semantics and effectively utilize it. In the proposed framework, the semantics are represented by binary features and used as additional information for improving the basic detection results.

Without loss of generality, the interesting event in soccer games is used as an example. Fig. 2 shows the key frames of an interesting event (goal shot) and the following five consecutive shots. As can be seen from the figure, a typical interesting event is usually followed by one (or more) close-up shot (usually the shooter), multi-player shot, and audience shot, which can be characterized as a temporal pattern. In addition, the goal shot should have a high grass ratio and high volume because of the excitement from both audience and commentator. Therefore, a set of binary semantic features are defined in Table I, where each feature is denoted as $F_j,\ j = 1, \ldots, J$, and J is the total number of features. The next problem is how to evaluate the significance of each semantic feature and calculate the total impact for assisting video event detection.

Table I: Semantic features

Feature Id	Semantics	Example
F_1	Football field	
F_2	Close-up shot	
F_3	Multi-player shot	
F_4	Audience shot	
F_5	Excitement from audience and commentator	N/A

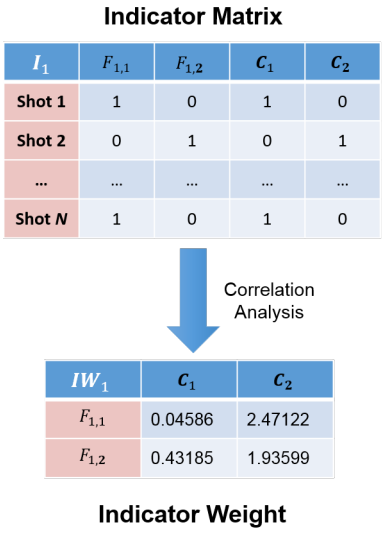

Figure 3: Indicator weight generation

1) TMCA: MCA (Multiple Correspondence Analysis) has been successfully applied to various multimedia analysis tasks such as feature selection [24], discretization [25], data pruning [26], classification [27] and video semantic concept detection [20]. MCA analyzes the correlation between feature value pairs (also called feature items in this paper) and evaluates the contribution of each feature (attribute) in a finer granularity, hence assisting the targeted analysis task. Generally speaking, MCA is performed on the attribute level and correspondence analysis is carried out to project the original feature items to a new space for better representation. However, there is inevitable information loss during the projection and each new component in the projected space does not hold specific physical meaning. MCA has demonstrated its efficiency and effectiveness over numerical features, where each feature item after routine discretization does not carry semantic in the first place. However, in our scenario, each semantic feature attribute is already a bit vector (with the nominal value 0 or 1), which carries specific semantics. It is desirable to retain the original semantic information as much as possible while exploring the feature item level associations. To solve this problem, a TMCA (Temporal MCA) algorithm is proposed to analyze feature item correspondence and seamlessly integrate temporal information for semantic ranking.

Let $I_1 \in \mathbb{R}^{N\times 4}$ be an indicator matrix for a particular semantic feature (F_1) as shown in Fig. 3, where each column represents a feature item ($F_{1,1}$ or $F_{1,2}$) or a class label (C_1 or C_2), and each line is an instance (or some analysis unit such as a video shot in this paper), with a total number of N shots. The semantic meaning embedded in the indicator matrix is as follows. For example, the values for $F_{1,1}$ and C_1 for shot 1 are 1, which means shot 1 shows a football field and it is an interesting event. On the contrary, shot 2 has $F_{1,2}$ and C_2 with the value 1, which means it does not show a football field and it is not an interesting event. Without loss of generality, we use 1-D subscripts to represent a feature attribute (e.g., F_1) and 2-D subscripts to represent a feature item (e.g., $F_{1,1}$) throughout the paper. To calculate the correlation between a feature item ($F_{j,k},\ k = 1,\ldots,K$) and a class label ($C_l,\ l = 1,\ldots,L$), an indicator weighting method is illustrated in Eq. 1, where K is the total number of feature items for attribute F_1, L is the number of classes, and $\lambda \in [0,1]$ is a tuning parameter to accommodate the effect of the number of features. This indicator weight calculation approach takes advantages of both the traditional cosine similarity and Tanimoto coefficient [28]. .

$$IW_{j,k}^{l} = \frac{\vec{F}_{j,k}\cdot\vec{C}_l}{\left\|\vec{F}_{j,k}\right\|_2\cdot\left\|\vec{C}_l\right\|_2-\lambda\cdot\vec{F}_{j,k}\cdot\vec{C}_l} = \frac{\sum_{i=1}^{N}(f_{j,k}^{i}\cdot c_l^i)}{\sqrt{\sum_{i=1}^{N}(f_{j,k}^{i})^2}\cdot\sqrt{\sum_{i=1}^{N}(c_l^i)^2}-\lambda\cdot\sum_{i=1}^{N}(f_{j,k}^{i}\cdot c_l^i)} \quad (1)$$

Algorithm 1 Indicator Weight for Ranking

Input: Training data set Tr, testing data set Te
Output: Ranking score for Te based on indicator weights

1: **procedure** GENIW(Tr) ▷ Training
2: **for** each F_j $(j = 1,\cdots,J)$ **do**
3: Construct indicator matrix I_j;
4: **for** each $F_{j,k}$ $(k = 1,\cdots,K)$ **do**
5: **for** each C_l $(l = 1,\cdots,L)$ **do**
6: calculate $IW_{j,k}^{l}$ using Eq. 1;
7: **end for**
8: **end for**
9: **end for**
10: **return** IW ▷ IW is a 3-D matrix.
11: **end procedure**

12: **procedure** CALCSCORE(Tr, Te) ▷ Testing
13: $IW \leftarrow$ GENIW(Tr);
14: **for** each X_i in Te $(i = 1,\cdots,N)$ **do**
15: **for** each F_j $(j = 1,\cdots,J)$ **do**
16: Look up iw_j from IW;
17: $S_i \leftarrow S_i + (1 - iw_j)^2$;
18: **end for**
19: $S_i \leftarrow S_i/J$;
20: Add S_i to RS_1;
21: **end for**
22: **return** RS_1 ▷ Ranking score for Te
23: **end procedure**

The above indicate weight generation procedure is considered as a training process (as described in Algorithm 1 lines 1 to 11). Intuitively, to calculate the overall effect

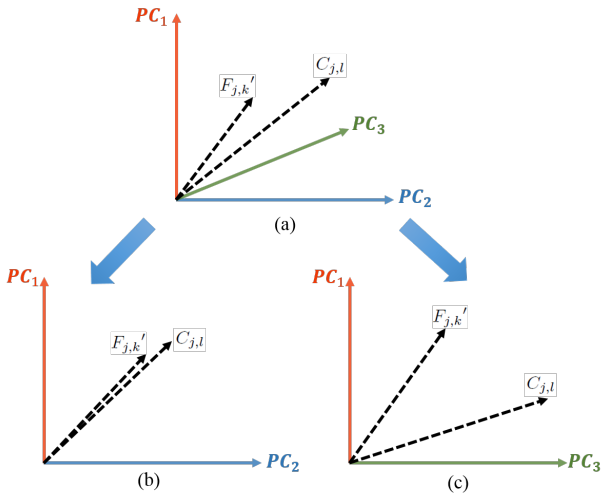

Figure 4: Feature item and class projection

of all the feature items for a specific instance towards a particular class, a summarization over all the feature attributes is required. The summarized value is known as an instance score. Eq. 2 shows the weighting scheme based on the trained indicator matrix IW, where the final score is normalized by the total number of attributes. Algorithm 1 lines 12 to 23 describe the procedure for calculating the scores, known as the testing process. For ease of illustration, the number of testing instances is also denoted as N.

$$S_i = \frac{\sum_{j=1}^{J}(1 - iw_j)^2}{J} \tag{2}$$

2) EN-MCA: As discussed earlier, the traditional MCA algorithm analyzes feature item correlations by projecting them into another space by keeping the first two principal components (PCs) in the transformed space. However, the projection process will unavoidably lose certain information. Moreover, it is not guaranteed that the first two PCs are the best representatives. For example, Fig. 4 illustrates the feature item vector $F_{j,k}{}'$ (different from $F_{j,k}$ since it is in the projected space) and the class vector $C_{j,l}$ in a different projection space. Fig. 4 (a) shows the two vectors in the 3-D space projected by the first three PCs, i.e., PC_1, PC_2 and PC_3. As can be seen from the figure, $F_{j,k}{}'$ and $C_{j,l}$ are pretty close in the space expanded by PC_1 and PC_2 (Fig. 4 (b)), while they are relatively distant from each other (with a larger angle) in the space composed of PC_1 and PC_3 (Fig. 4 (c)). However, the projection in (b) does not necessarily perform better than (c). By taking into consideration of all potential valuable PCs, we propose an enhanced MCA (EN-MCA) algorithm by fully utilizing all critical PCs.

The EN-MCA algorithm is described in Algorithm 2. Specifically, the training data set Tr is first discretized into nominal values using the well known Minimum Description

Algorithm 2 EN-MCA

Input: Training data set Tr
Output: EN-MCA weight matrix

1: **procedure** GENMW(Tr)
2: Discretize Tr into nominal intervals;
3: **for** each F_j $(j = 1, \cdots, J)$ **do**
4: Construct indicator matrix I_j;
5: Calculate burt matrix B_j;
6: $\{Z_j, V_j, E_j\} \leftarrow MCA(B_j)$;
7: Determine the number of PCs, Q_j;
8: $count \leftarrow 1$;
9: **for** $m \leftarrow 1, \ldots, Q_j - 1$ **do**
10: **for** $n \leftarrow m + 1, \ldots, Q_j$ **do**
11: $VP \leftarrow [\ ]$;
12: Calculate $F_j{}'$ and C_j;
13: **for** each $F_{j,k}{}'$ $(k = 1, \cdots, K)$ **do**
14: **for** each $C_{j,l}$ $(l = 1, \cdots, L)$ **do**
15: calculate $W^l_{j,k}(count)$;
16: **end for**
17: **end for**
18: $VP[count] \leftarrow V_j[m] * V_j[n]$;
19: $count \leftarrow count + 1$;
20: **end for**
21: **end for**
22: **for** $q \leftarrow 1, \ldots, count$ **do**
23: $w_q \leftarrow VP[q]/sum(VP)$;
24: **for** each $F_{j,k}{}'$ $(k = 1, \cdots, K)$ **do**
25: **for** each $C_{j,l}$ $(l = 1, \cdots, L)$ **do**
26: $MW^l_{j,k} \leftarrow MW^l_{j,k} + W^l_{j,k}(q) * w_q$;
27: **end for**
28: **end for**
29: **end for**
30: **end for**
31: **return** MW ▷ MW is a 3-D matrix.
32: **end procedure**

Length (MDL) algorithm [29]. Then, for each feature attribute, an indicator matrix (I_j, similar to Fig. 3) is built, followed by the generation of burt matrix B_j. Subsequently, the traditional MCA is performed, obtaining three matrices, i.e., Z_j, V_j, and E_j, where Z_j is the centralized and normalized burt matrix, V_j is sorted eigen vectors, and E_j is the corresponding eigen values. The number of PCs to be retained is determined by the accumulated variance calculated from V_j [30]. For each pair of PCs, the projected vectors $F_j{}'$ and C_j are generated based on Z_j and V_j. Then the MCA weight is calculated for each feature item $F_{j,k}{}'$ and class label $C_{j,l}$ as shown in Algorithm 1 lines 13 to 17. For details about how to perform the conventional MCA and calculate the weight, please refer to [31]. At the same time, the significance of each PCs pair is evaluated in Algorithm 2

line 18. Finally, the final MCA weight for each pair of ${F_{j,k}}'$ and $C_{j,l}$ is calculated using the linear combination of each $W^l_{j,k}$ based on the normalized weight factor w_q. This is the training stage for the EN-MCA algorithm. For calculating the testing score for Te, a similar procedure as in Algorithm 1 (lines 12 to 23) is followed. A valuable conclusion drawn from the study and analysis of the EN-MCA algorithm is that it is more effective for a larger number of feature items. In other words, the feature attribute with more discretized intervals will benefit more from the EN-MCA algorithm.

3) Re-Ranking: In this paper, the binary classification problem is taken as an example to illustrate the proposed re-ranking procedure. As shown in Algorithm 3, the input of the re-ranking algorithm is the initial classification results, denoted as CM, which contains the classified positive and negative instances represented as G_1 and G_2 respectively. Another input is the ranking score from different ranking models, e.g., RS_1 is the ranking score obtained from the indicating weight ranking procedure. Finally, we have the significance factor for each ranking model. The re-ranking procedure starts by normalizing the ranking score for each model (Algorithm 3 line 5) based on training data using Z-score normalization method. Then the normalized ranking scores are linearly combined by using the corresponding significance factor ρ_t, generating the final ranking score Φ. Finally, the refined positive instances ${G_1}'$ is generated by excluding the instances below a preset threshold θ_1 in G_1 (Algorithm 3 line 10), and including the instances above θ_2 in G_2 (Algorithm 3 line 16), vise versa for ${G_2}'$.

Algorithm 3 Re-Ranking

Input: Classification results $CM = \{G_1, G_2\}$, ranking score for different models $\{RS_t \mid t = 1, \ldots, T\}$, significance factor for each model $\{\rho_t \mid \sum_{t=1}^{T} \rho_t = 1\}$
Output: Refined classification results based on re-ranking CM'

1: **procedure** RERANKING(CM, RS, ρ)
2: $\quad \Phi \leftarrow [\ ]$;
3: $\quad CM' \leftarrow CM$;
4: $\quad$ **for** each RS_t **do**
5: $\quad\quad$ Perform normalization;
6: $\quad\quad \Phi \leftarrow \Phi + RS_t * \rho_t$;
7: $\quad$ **end for**
8: $\quad$ **for** each X_i in G_1 $(i = 1, \cdots, size(G_1))$ **do**
9: $\quad\quad$ **if** $\Phi(X_i) < \theta_1$ **then**
10: $\quad\quad\quad {G_1}' \leftarrow {G_1}' - X_i$;
11: $\quad\quad\quad {G_2}' \leftarrow {G_2}' + X_i$;
12: $\quad\quad$ **end if**
13: $\quad$ **end for**
14: $\quad$ **for** each X_i in G_2 $(i = 1, \cdots, size(G_2))$ **do**
15: $\quad\quad$ **if** $\Phi(X_i) > \theta_2$ **then**
16: $\quad\quad\quad {G_1}' \leftarrow {G_1}' + X_i$;
17: $\quad\quad\quad {G_2}' \leftarrow {G_2}' - X_i$;
18: $\quad\quad$ **end if**
19: $\quad$ **end for**
20: $\quad$ **return** CM'
21: **end procedure**

IV. Experimental Results

The proposed framework was rigorously tested upon a large experimental data set, which contains 23 soccer videos collected from the FIFA World Cup of 2010 and 2014. The total number of frames is about 2.8 millions and the total duration of the videos is over 31 hours. Among the total 20k video shots, only 91 of them contain the interesting events, which contributes only 0.5% to the total number of shots. A summary of the data set is shown in Table II. Within the scope of this study, the interesting events in any soccer game include both goal shot and goal attempt, since it is not uncommon that the users are sometimes interested in a certain goal attempt event.

The experimental settings are as follows. The decision tree classifier in Weka [32] is used to generate the basic classification results (CM). Then the proposed indicator weighting scheme is applied to generate the semantic score (RS_1). Another two sets of ranking scores, RS_2 and RS_3, are produced by the proposed EN-MCA algorithm and the LibSVM model [33] using the multi-modal features [23]. Intuitively, the more ranking models used, the better the performance. However, the computational complexity is higher. Therefore, there is a trade-off between performance and complexity. If another feature set with a larger number of features is used, the Hidden Coherent Feature Groups (HCFGs) [34] analysis method could be used to select the exemplar features and greatly reduce the overall complexity. All the ranking models are considered equally important in our experimental analysis, i.e., $\rho_t = 1/T$ with $T = 3$. A more advanced approach is to determine the factor based on training performance [34]. θ_1 and θ_2 are set to be empirical values, i.e., 0 and 2 in the experiments. A more flexible solution is to determine them based on training statistics.

Those semantic features shown in Table I could be extracted with different manners. For example, F_1 could be analyzed by the dominant color in a video frame (i.e., the *grass ratio* should be over a certain threshold); F_2 to F_4 can be determined by the number of faces (either frontal or profile, which is used in the experiments) [35] in a frame or by an object and crowd detection [10, 36]. Lastly, F_5 is decided by the volume value in the audio modality. Based on our experimental observation, these features are relatively easy to extract and reasonable in the sense of conveying preliminary semantics. In addition, more semantic features could be added to boost the performance.

To better evaluate our proposed framework, three-fold

cross-validation is used. That is, two thirds of the data set are randomly selected for training and the rest one third is for testing. The precision (Pre), recall (Rec), and F1 are calculated as the performance measurements.

Table III and IV show the interesting event detection results before and after applying the proposed re-ranking framework. Specifically, Table III presents the base classification results from the decision tree, and Table IV is the outcome of the re-ranking process. As can be seen from the table, the proposed framework improved the detection results by reducing the numbers of FN (missed interesting events) and FP (mis-identified interesting events) to 5 and 7, respectively. The overall precision, recall, and F1 are all increased by 2% to 4%. Considering the rareness of the interesting events and the skewed nature of the data set, the improvement is promising. Based on our observation, most of the mis-identified interesting events are foul or those shots being close to the real event, which have similar patterns with the real event and are difficult to be identified. On the other hand, a number of interesting events are resulted from corner kick and penalty kick, which have irregular characteristics compared with the normal goal shots and are easily identified as non-interesting event. From the experimental results, it can be concluded that the incorporation of temporal semantic information by using the indicator weighting method together with the EN-MCA ranking mechanism have improved the interesting event detection results. Finally, it is worth noting that although our indicator weighting algorithm is designed for nominal features, it also applies to numerical features when they are properly discretized.

Table II: Data set summary.

No. Files	Total Frame	Total Time	Total Shots	No. Events
23	2,084,102	31 h 20 m	20,082	91

Table III: Performance evaluation before re-ranking.

Fold Number	No. Events	TP	FN	FP	Pre	Rec	F1
Fold 1	31	30	1	2	93.7%	96.8%	95.2%
Fold 2	31	26	5	3	89.7%	83.9%	86.7%
Fold 3	29	27	2	4	87.1%	93.1%	90.0%
Summary	91	83	8	9	90.2%	91.2%	90.7%

Table IV: Performance evaluation after re-ranking.

Fold Number	No. Events	TP	FN	FP	Pre	Rec	F1
Fold 1	31	31	0	3	91.2%	100%	95.4%
Fold 2	31	28	3	1	96.6%	90.3%	93.3%
Fold 3	29	27	2	3	90.0%	93.1%	91.5%
Summary	91	86	5	7	92.5%	94.5%	93.5%

V. Conclusions and future work

This paper proposes a two-phase multimedia big data mining framework for interesting event detection in soccer videos. In the pre-processing phase, the multi-filtering content analysis techniques are used for shot boundary detection and the multi-modal features are extracted. A candidate set is produced by the pre-filtering process based on domain knowledge analysis. In the event detection phase, TMCA algorithm is used to generate semantic re-ranking score by incorporating temporal and semantic features using the indicator weighting strategy. Moreover, the EN-MCA algorithm is applied to better capture feature item association for final decision-making. The whole interesting event detection framework is successfully evaluated on a soccer game data set. The experimental results are satisfactory considering the rareness of the interesting event. It is worth mentioning that our proposed framework could be easily adapted to other video types. There are several future study directions. First, both indicator weight and MCA weight try to capture the correlation between feature items and class labels, but from different levels. While the indicator weight keeps all the original information and carry more semantics, the MCA weight provides more detailed analysis within each feature item. It is promising to effectively integrate these two types of weights for various semantic analysis tasks. Second, since the processing of each feature attribute is independent, it is feasible and desirable to parallel the calculation by introducing the MapReduce framework on the Hadoop platform for distributed computing. It will greatly accommodate the big data requirement, considering the ever-increasing amount of multimedia data. Finally, the temporal information is loosely incorporated into our framework, and therefore, it is another potential direction for better utilizing the embedded temporal characteristics. Finally, more data sets should be used and more detailed evaluation should be carried out to further demonstrate the effectiveness and efficiency of our proposed algorithms and framework. In addition, more work has to be done to relieve the effect of domain knowledge and provide a more universal framework for the general use purpose.

Acknowledgment

This research was supported by NSF HRD-0833093 and CNS-1126619. The authors would like to thank Haiman Tian for her assistance in conducting the experiment and preparing the previous work.

References

[1] "Apache mahout," http://mahout.apache.org.

[2] V. Tovinkere and R. J. Qian, "Detecting semantic events in soccer games: Towards a complete solution." in *IEEE International Conference on Multimedia and Expo (ICME)*, 2001, pp. 1040–1043.

[3] S. Dagtas and M. Abdel-Mottaleb, "Extraction of tv highlights using multimedia features," in *IEEE Workshop on Multimedia Signal Processing*, 2001, pp. 91–96.

[4] P. Shi and Y. Xiao-qing, "Goal event detection in soccer videos using multi-clues detection rules," in *IEEE International Conference on Management and Service Science (MASS)*, 2009, pp. 1–4.

[5] W. Zhu, C. Toklu, and S.-P. Liou, "Automatic news video segmentation and categorization based on closed-captioned text," *Urbana*, vol. 51, p. 61801, 2001.

[6] M. Xu, N. C. Maddage, C. Xu, M. Kankanhalli, and Q. Tian, "Creating audio keywords for event detection in soccer video," in *IEEE International Conference on Multimedia and Expo (ICME)*, vol. 2, 2003, pp. II–281.

[7] Y. Rui, A. Gupta, and A. Acero, "Automatically extracting highlights for tv baseball programs," in *Proceedings of the ACM international conference on Multimedia*, 2000, pp. 105–115.

[8] Q. Ye, Q. Huang, W. Gao, and S. Jiang, "Exciting event detection in broadcast soccer video with mid-level description and incremental learning," in *Proceedings of the ACM international conference on Multimedia*, 2005, pp. 455–458.

[9] F. Wang, Y.-F. Ma, H.-J. Zhang, and J.-T. Li, "Dynamic bayesian network based event detection for soccer highlight extraction," in *IEEE International Conference On Image Processing (ICIP)*, vol. 1, 2004, pp. 633–636.

[10] 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, no. 10, pp. 1225–1233, 2005.

[11] M. Xu, L.-Y. Duan, C.-S. Xu, and Q. Tian, "A fusion scheme of visual and auditory modalities for event detection in sports video," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, 2003, pp. III–189.

[12] J. Wang, C. Xu, E. Chng, and Q. Tian, "Sports highlight detection from keyword sequences using HMM," in *IEEE International Conference on Multimedia and Expo (ICME)*, vol. 1, 2004, pp. 599–602.

[13] 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, no. 3, pp. 421–436, 2008.

[14] H. Xu and T.-S. Chua, "The fusion of audio-visual features and external knowledge for event detection in team sports video," in *Proceedings of the ACM SIGMM international workshop on Multimedia information retrieval*, 2004, pp. 127–134.

[15] A. A. Halin, M. Rajeswari, and M. Abbasnejad, "Soccer event detection via collaborative multimodal feature analysis and candidate ranking," *Int. Arab J. Inf. Technol.*, vol. 10, no. 5, pp. 493–502, 2013.

[16] J. Assfalg, M. Bertini, A. Del Bimbo, W. Nunziati, and P. Pala, "Soccer highlights detection and recognition using HMMs," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2002, pp. 825–828.

[17] T. Wang, J. Li, Q. Diao, W. Hu, Y. Zhang, and C. Dulong, "Semantic event detection using conditional random fields," in *IEEE International Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, 2006, pp. 109–109.

[18] M. Chen, S.-C. Chen, and M.-L. Shyu, "Hierarchical temporal association mining for video event detection in video databases," in *Proceedings of the IEEE International Workshop on Multimedia Databases and Data Management (MDDM), in conjunction with IEEE International Conference on Data Engineering (ICDE)*, 2007, pp. 137–145.

[19] Z. Xie, M.-L. Shyu, and S.-C. Chen, "Video event detection with combined distance-based and rule-based data mining techniques," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2007, pp. 2026–2029.

[20] M.-L. Shyu, Z. Xie, M. Chen, and S.-C. Chen, "Video semantic event/concept detection using a subspace-based multimedia data mining framework," *IEEE Transactions on Multimedia*, vol. 10, no. 2, pp. 252–259, 2008.

[21] M. Chen, S.-C. Chen, M.-L. Shyu, and K. Wickramaratna, "Semantic event detection via multimodal data mining," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 38–46, 2006.

[22] S.-C. Chen, M.-L. Shyu, and C. Zhang, "Innovative shot boundary detection for video indexing," *Video data management and information retrieval*, pp. 217–236, 2005.

[23] S.-C. Chen, M.-L. Shyu, M. Chen, and C. Zhang, "A decision tree-based multimodal data mining framework for soccer goal detection," in *IEEE International Conference on Multimedia and Expo (ICME)*, vol. 1, 2004, pp. 265–268.

[24] Q. Zhu, L. Lin, M.-L. Shyu, and S.-C. Chen, "Feature selection using correlation and reliability based scoring metric for video semantic detection," in *IEEE International Conference on Semantic Computing (ICSC)*, 2010, pp. 462–469.

[25] ——, "Effective supervised discretization for classification based on correlation maximization," in *IEEE International Conference on Information Reuse and Integration (IRI)*, 2011, pp. 390–395.

[26] L. Lin, M.-L. Shyu, and S.-C. Chen, "Enhancing concept detection by pruning data with mca-based transaction weights," in *IEEE International Symposium on Multimedia (ISM)*, 2009, pp. 304–311.

[27] Y. Yang, H.-Y. Ha, F. Fleites, S.-C. Chen, and S. Luis, "Hierarchical disaster image classification for situation report enhancement," in *IEEE International Conference on Information Reuse and Integration (IRI)*, 2011, pp. 181–186.

[28] A. H. Lipkus, "A proof of the triangle inequality for the tanimoto distance," *Journal of Mathematical Chemistry*, vol. 26, no. 1-3, pp. 263–265, 1999.

[29] U. M. Fayyad and K. B. Irani, "On the handling of continuous-valued attributes in decision tree generation," *Machine learning*, vol. 8, no. 1, pp. 87–102, 1992.

[30] P. R. Peres-Neto, D. A. Jackson, and K. M. Somers, "How many principal components? stopping rules for determining the number of non-trivial axes revisited," *Computational Statistics & Data Analysis*, vol. 49, no. 4, pp. 974–997, 2005.

[31] L. Lin, C. Chen, M.-L. Shyu, and S.-C. Chen, "Weighted subspace filtering and ranking algorithms for video concept retrieval," *IEEE MultiMedia*, vol. 18, no. 3, pp. 32–43, 2011.

[32] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[33] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[34] Y. Yang, H.-Y. Ha, F. C. Fleites, and S.-C. Chen, "A multimedia semantic retrieval mobile system based on HCFGs," *IEEE MultiMedia*, vol. 21, no. 1, pp. 36–46, 2014.

[35] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[36] Y. Yang, F. C. Fleites, H. Wang, and S.-C. Chen, "An automatic object retrieval framework for complex background," in *IEEE International Symposium on Multimedia (ISM)*, 2013, pp. 374–377.