An Effective Multi-Concept Classifier for Video Streams

Shu-Ching Chen[1], Mei-Ling Shyu[2], Min Chen[3]
[1]*School of Computing and Information Sciences*
Florida International University, Miami, FL 33199, USA
chens@cs.fiu.edu
[2]*Department of Electrical & Computer Engineering*
University of Miami, Coral Gables, FL 33124, USA
shyu@miami.edu
[3]*Department of Computer Science, University of Montana*
Missoula, MT 59812, USA
chen@cs.umt.edu

Abstract

In this paper, an effective multi-concept classifier is proposed for video semantic concept detection. The core of the proposed classifier is a supervised classification approach called C-RSPM (Collateral Representative Subspace Projection Modeling) which is applied to a set of multimodal video features for knowledge discovery. It adaptively selects non-consecutive principal dimensions to form an accurate modeling of a representative subspace based on the statistical information analysis and thus achieves both promising classification accuracy and operational merits. Its effectiveness is demonstrated by the comparative experiment, as opposed to several well-known supervised classification approaches including SVM, Decision Trees, Neural Network, Multinomial Logistic Regression Model, and One Rule Classifier, on goal/corner event detection and sports/commercials concepts extraction from soccer videos and TRECVID news collections.

1. Introduction

The advanced development in electronic imaging, video devices, storage, networking, and computer power has made it possible and affordable for generating, sharing, and analyzing huge amounts of multimedia data across large-scale distributed data sources (e.g., video and audio databases). For video databases, semantic analysis such as concept detection is vital for effective video data management. Here, the concepts include both important activities that capture users' attentions (soccer goals, traffic accidents, etc.) and high-level semantic features (sports, commercials, etc.) [12].

One of the main challenges in this research area is how to refer high-level semantic concepts automatically (or at least semi-automatically) from the low-level video features, which is the so-called semantic gap. Many research efforts have been devoted toward the following three main processes.

- *Syntactic analysis*. Many studies have been conducted to partition the video clips into appropriate analysis units (shot-level [4], story unit [14], etc.) and to explore more representative features for the targeted video concepts. It is often further classified into two broad categories, i.e., unimodal approaches [20] that study the respective role of visual, audio, and texture mode in the corresponding domain, and multimodal approaches [6][15] that combine the strength of various modalities to capture the video content in a more comprehensive manner [3]. Nevertheless, the studies are mainly at syntax-level which are relatively less domain dependent and capture low-level features directly from video streams.
- *Decision-making process*. This process aims at extracting the semantic index from the feature descriptors to improve the framework robustness. For instance, the Markov-model-based techniques have been extensively studied, including the hidden Markov model (HMM) [18] and controlled Markov chain (CMC) [6], to model the temporal relations among the frames or shots for a certain event. Recently, data mining techniques, such as SVM [9], Neural Network [7], and Decision Tree [4], have been increasingly adopted owing to their

strong capability of uncovering useful and/or nontrivial information from large volumes of data. Though such techniques have been long proven as effective data classification mechanisms and have been successfully applied to many different applications, they still fail to bridge the semantic gap in video concept detection by applying only to low-level features. Instead, most current researches rely heavily on certain artifacts such as domain-knowledge and *a priori* models, leading to the so-called domain-related modeling process.

- *Domain-related modeling*. Though some generalized video concept detection approaches have been conducted, their detection capability is largely limited [11]. Therefore, the domain-related modeling process is widely adopted in the literature to derive domain specific mid-level representations [5] or heuristic rules [4]. They can largely boost the framework accuracy by either improving the feature representations or pruning the data set with the facilitation of domain knowledge, which unfortunately greatly limits their extensibility in handling other application domains and/or video sources.

In this paper, we target at relaxing the dependency on domain knowledge and automating the concept detection process with the adoption of multimodal content analysis and the C-RSPM (Collateral Representative Subspace Projection Modeling) supervised classification approach. In addition, we also address another critical yet less studied challenge, called data imbalance (or rare concept detection) issue. That is, generally the concepts of interests are often infrequent, and thus a large number of negative instances overshadow a small percentage of positive counterparts and dominate the detection model training process. This issue usually results in an undesirable degradation of the detection performance as demonstrated in our previous study [10].

This paper is organized as follows. Section 2 describes our proposed framework in details. The empirical study and results are presented and analyzed in Section 3. We conclude our study in Section 4.

2. The proposed framework

The proposed framework consists of three major components, namely video syntactic analysis, subspace-based data pruning, and C-RSPM data classification (as shown in Figure 1). As will be discussed later, both data pruning and C-RSPM components are subspace-based. Therefore, the advantages of adopting subspace-based data pruning followed by the C-RSPM classification approach are twofold. First, the removal of large portions of negative instances greatly alleviates the data imbalance issue and results in better class distribution. Second, a large portion of calculation results from the data pruning component can be re-used in the C-RSPM component.

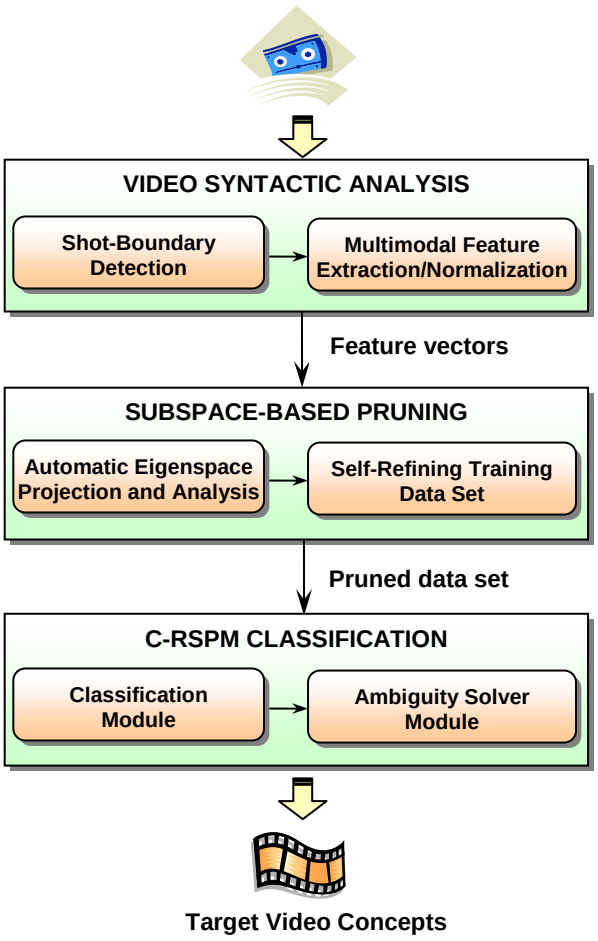

Figure 1. Overview of the Proposed Framework

2.1. Video syntactic analysis

In this component, a shot-boundary detection algorithm proposed in our previous work [1] is adopted to partition the video sequence into basic syntactic units (i.e., shots), which serve as the basis for feature extraction and semantic analysis. Multimodal features including five visual features and twelve audio features are then extracted for each shot. The visual features are *pixel_change*, *histo_change*, *dominant_color_ratio*, *background_mean*, and *background_var*. The first two features denote the average pixel/histogram changes

between the consecutive frames within a shot, and *dominant_color_ratio* represents the ratio of dominant color in the frames. *background_mean* and *background_var* are based on region-level analysis where the SPCPE segmentation algorithm [2] is used to identify background/foreground regions of video frames. They then capture the shot-level standard deviation and mean color values in the background regions. Audio features are exploited in both time-domain and frequency-domain, which include ten shot-level generic features (one volume, five energy, and four spectrum flux features) and two "around-boundary" audio features, that is, due to the reason that an audio stream can be continuous even around the shot boundary, the volume statistics information (i.e., mean and max) is captured for the duration of 3 seconds around the shot boundary to explore the audio track information.

More details about the visual-audio features are presented in our previous work [4]. The feature set is normalized to minimize the feature scale effects for multivariate data. One important fact is that the feature set is not domain specific and is used for the detection of all the concepts in our study.

2.2. Subspace-based data pruning

It is often challenging for a typical detection process to be able to capture a small portion of targeted instances from the huge amount of video data (e.g., less than 1:100 in our goal event detection empirical study), especially with the existence of noisy and irrelevant information introduced during video production and feature extraction processes. To address this data imbalance (or rare concept detection) issue, many existing studies seek the help from domain knowledge through domain-related modeling; whereas in contrast, we adopt the subspace-based data pruning scheme proposed in our previous study [10] to eliminate a great portion of negative instances without the dependency of domain knowledge.

Let $X = \{x_{ij}\}$ $(i = 1,2,..., p$ and $j = 1,2,..., N)$ be the feature matrix resulted from the first component (i.e., video syntactic analysis), containing N p-dimensional column vectors $X_j = (x_{1j}, x_{2j},..., x_{pj})'$, where p and N indicate the number of features (e.g., seventeen in this study) and analytical units (i.e., shots in this study) in the original data set. For supervised data classification, X is divided into a training data set X^A (i.e., class labels are given with $N1$ labeled positive instances X^{Ae} and $N2$ labeled negative instances X^{An}) and a testing data set X^B (with unknown class labels).

2.2.1. Step 1: Automatic eigenspace projection and analysis. The main idea in this step is to utilize the correlation matrix of the trimmed training data set to acquire the statistics on the transformed principal component space. Let $\overline{X} = (1/N)\sum_{j=1}^{N} X_j$, the robust correlation matrix is defined as $S = (1/(N-1))\sum_{j=1}^{N}(X_j - \overline{X})(X_j - \overline{X})'$.

We randomly select $T1$ data instances from X^{Ae} for K times and pick the best group as X^e. The statistical properties of this group, as will be defined in Eq. (1), can be used to recognize 100% data instances in X^{Ae} (i.e., they are considered normal data instances to X^e) and at the same time to reject the maximal percentage of data instances (as abnormal data instances) in X^{An}. Note that we can always find the group(s) with 100% recognizing rate for data instances in X^{Ae} when K and $T1$ are big enough (an extreme case would be to set $T1 = N1$). Similarly, X^n can be defined as the selected $T2$ ($T2 < N2$) negative data instances which reject 100% data instances in X^{Ae} and at the same time recognize the maximal percentage of data instances in X^{An}. In our proposed framework, X^e and X^n are called typical positive data instances and typical negative data instances, respectively, which facilitate further data analysis and better exploration of the statistical information present in the data set.

Now, the key challenge is how to locate X^e and X^n, which in turns is converted into the issue of how to differentiate the normal and anomalous data instances in the view of a set of positive (or negative) data instances.

Assume that (λ_1, E_1), (λ_2, E_2), …, (λ_p, E_p) are the p eigenvalue-eigenvector pairs of the robust correlation matrix S for the selected $T1$ positive data instances from X^{Ae} or $T2$ negative data instances from X^n. E_i is the i^{th} typical eigenvector and $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$. Let $Y = \{y_{ij}\}$ $(i = 1,2,..., p$ and $j = 1,2,..., T1$ for the positive data instances or $j = 1,2,..., T2$ for the negative data instances) be the instance score matrix, and R_i $(i = 1,2,..., p)$ be the row vector in Y. We define the class-deviation measure as shown in Eq. (1).

$$c_j = \sum_{m \in M} (y_{mj})^2 / \lambda_m . \qquad (1)$$

Here, $m \in M$ is the selected positive (or negative) component space which satisfies the following condition:

$$STD(R_m) < \alpha . \qquad (2)$$

α is the arithmetic mean of the standard deviation values of all R_i. The maximum value $c_{\max} = \max(c_j)$ is selected as a threshold to determine if an incoming data instance is statistically normal to the selected $T1$ positive (or $T2$ negative) data instances, i.e., the data instance k is abnormal if $c_k > c_{\max}$ and normal otherwise.

Once X^e and X^n are identified, X^e is used to reject negative data instances in X, that is, to locate anomalous data instances in the view of X^e following the above stated approach, and the recognized normal data instances are then checked by X^n for further pruning. Consequently, a large portion of negative data instances are successfully removed, which greatly alleviates the aforementioned issue. In addition, all the remaining data instances are projected onto the typical negative eigenspace and the score values are extracted to replace the original feature set with the extra benefits of feature reduction.

2.2.2. Step 2: Self-refining training data set. The quality of the training data set has a great influence on the final classification performance, especially in the case of semantic concept detection, where the number of positive data instances is so limited that several mislabeled training data instances may greatly degrade the training model. To overcome this problem, a training data set self-refining process is proposed based on the first dimension of the typical negative eigenspace (denoted as R_1^{Ae} and R_1^{An} for the data instances remained in X^{Ae} and X^{An}, respectively) since it presents the most data information. Normally, R_1^{Ae} (the relative anomalous ones) would have a higher value than that of R_1^{An}. Therefore, the following two self-learning rules are proposed to refine the training data set.

- Any data instance whose corresponding value in R_1^{An} is larger than the average value of R_1^{An} is removed;
- Any data instance whose corresponding value in R_1^{Ae} is smaller than half of the average value of R_1^{Ae} is removed.

2.3. C-RSPM classification

Given the pruned data set, the C-RSPM classification algorithm is applied for final concept detection. In our previous study [8], we have developed C-RSPM as a multi-class supervised classification framework for intrusion detection application, whose main idea is that the training data instances belong to different classes can in fact be considered anomalous to one another. Intuitively, C-RSPM can be extended in concept detection applications, since the concept units can be considered as anomalous to the non-concept ones and vice versa. C-RSPM consists of a Classification Module and an Ambiguity Solver Module.

2.3.1. Classification module. The classification module contains an array of component classifiers, where the number of classifiers is determined by the number of classes required by a specific application (e.g., two classifiers in our study as there are two classes, event/concept vs. non-event/concept). An unknown data instance is input to each of the classifiers, and each classifier classifies this unknown data instance as normal (i.e., belonging to the concept of its training data instances) or anomalous (i.e., non-concept). The basic idea in this module is to generate a predictive model that learns the similarities among training data instances of a particular class (concept), after which the attained similarity information can be used in recognizing the testing data instances that are normal to the corresponding concept. Meanwhile, all other data instances belonging to other classes (concepts) are recognized as statistically anomalous to the classifier via a computed class threshold measure which is derived from our proposed instance class-deviation measure.

Specifically, given R_m ($m \in M$) obtained in Eq. (2), a refined principal component space V is determined using parameter TH as defined in Eq. (3).

$$TH = \frac{\lambda_m}{[STD(R_m)]^2}. \qquad (3)$$

Consequently, we get $|M|$ number of TH values and the principal components with the top ranked TH values. Based on our empirical studies, the ones rank top 20% are selected to form the space V in this paper. It is noted that the selected principal components are possibly non-consecutive and may not be only major principal components, breaking the assumption widely used in previous PCA (Principal Component Analysis)-based algorithms that only the principal components with larger eigenvalues are important to the representation of the original data set. It is not always true since it ignores that principal components with larger eigenvalues only have a significant effect on the representation of the information embracing both the "similarity" and "dissimilarity" information, instead of only the "similarity" of the original data set. However, in C-RSPM, the most important aspect is the representation of the similarity of a training class.

Similarly, Eq. (1) is applied as the class-deviation equation, with the only change that the subspace *V* is used instead of the space *M*. This results in an array $C = \{c_j\}$. From a geometrical point of view, this can be considered as an ellipsoid modeling function in the refined eigenspace and *C* is an array of values corresponding to the possible ellipsoidal borders that can be used to enclose the projected training data set of the class. Thus a threshold (denoted as C_{th}) is generated based on *C*, the desired pre-set false alarm rate β, and the Cumulative Distribution Function (*CDF*) as given in Eq. (4).

$$CDF_C(C_{th}) = 1 - \beta. \quad (4)$$

That is, C_{th} is defined by finding the *CDF* of the array *C*, and the Parzen window non-parametric fitting method [19] is adopted to select the threshold value based on β which is an adjustable input parameter in C-RSPM. In order to coalesce both aspects of a high true detection rate and a low false alarm rate, a typical low value ($\beta = 0.1\%$) that has been employed in several research areas and applications [17] is chosen as the default value for each classifier component. Then for each component classifier, the decision rules to classify each of the testing data instances $X_j^B, j = 1,2,..., N'$ (assuming N' testing data instances) can be established naturally based on C_{th}, i.e., the data instance j is abnormal if $c_j^B > C_{th}$ and normal otherwise.

2.3.2. Ambiguity solver module. Ideally, a data instance normal to a particular classifier should be rejected by the remaining classifiers. However, in real applications, it is possible that an instance either (i) may be classified as normal by multiple classifiers, or (ii) may not be recognized as normal by any classifier. Such an ambiguous situation is addressed by the ambiguity solver module in C-RSPM.

The first issue arises from the fact that hardly any classifier can ensure 100% classification accuracy and the quality of data sources is rarely perfect. The second issue usually translates into an incoming testing data instance belonging to an unknown class which has not been modeled by C-RSPM or originates from the pre-set false alarm rate. To solve the ambiguity issue, both the Cumulative Distribution Function (*CDF*) and Probability Density Function (*PDF*) are first applied, and the one with the stronger differentiation power for a certain ambiguous testing instance among the classifiers is used. The testing data instance is classified to be the class (concept) of the classifier with the smallest *CDF* value (if using *CDF*) or the largest *PDF* value (if using *PDF*), and it is considered to be abnormal to the rest of the classifiers.

Formally, in the Classification Module for concept detection, we obtain C^e / C_{th}^e for the concept class and C^n / C_{th}^n for the non-concept class, respectively, following the procedures discussed earlier. For the j^{th} data instance in X^B (i.e., X_j^B), we get c_j^e and c_j^n accordingly. It is considered as ambiguous if (i) $c_j^e \le C_{th}^e$ and $c_j^n \le C_{th}^n$, or (ii) $c_j^e > C_{th}^e$ and $c_j^n > C_{th}^n$. The Ambiguity Solving Module uses the following steps to classify whether the X_j^B data instance belongs to the concept class.

STEP 1:
Calculate $CDF_{c^e}(c_j^e)$, $PDF_{c^e}(c_j^e)$, $CDF_{c^n}(c_j^n)$, $PDF_{c^n}(c_j^n)$.

STEP 2:
if $\frac{|CDF_{c^e}(c_j^e) - CDF_{c^n}(c_j^n)|}{CDF_{c^e}(c_j^e)} \ge \frac{|PDF_{c^e}(c_j^e) - PDF_{c^n}(c_j^n)|}{PDF_{c^e}(c_j^e)}$
  if $CDF_{c^e}(c_j^e) \le CDF_{c^n}(c_j^n)$, X_j^B is a concept;
  else X_j^B is a non-concept;
  endif
else
  if $PDF_{c^e}(c_j^e) \ge PDF_{c^n}(c_j^n)$, X_j^B is a concept;
  else X_j^B is a non-concept;
  endif
endif

Note that in the rare case that $CDF_{c^e}(c_j^e) == CDF_{c^n}(c_j^n)$ or $PDF_{c^e}(c_j^e) == PDF_{c^n}(c_j^n)$, the data instance is assigned to be a concept. This is because in concept detection, the recall metric is normally considered as more important than the precision metric. In other word, we would like to be able to classify as many data instances to the correct concepts as possible even at the cost of including a small number of false positives.

3. Empirical study

The proposed framework was rigorously tested upon a large experimental data set with 27 soccer videos and 6 TRECVID videos [13]. The total duration of the video clips is about 800 minutes, and these

videos were obtained from a variety of sources with various production styles. Soccer videos are one of the most widely adopted testbeds for concept detection due to their popularity and loose structures. TRECVID videos are used and promoted by the National Institute of Standards and Technology to boost the researches in semantic media analysis by offering a common video corpus [13]. In the experiment, we target to detect goal and corner concepts from the soccer videos, where they account for only 41 and 95 shots, respectively, out of 4,885 shots. As discussed earlier, it is quite challenging to detect such rare concepts. In addition, two concepts, sports and commercial, are selected as the target concepts from the TRECVID videos since they differ from each other in terms of production styles and occurrence frequencies (63 and 898 out of 2,304 shots, respectively). By applying a single framework on different video sources and concepts, we intend to testify the effectiveness and generalization of our proposed framework.

3.1. Experimental setup

In order to better evaluate our proposed framework, the five-fold cross-validation scheme is used. That is, the 2/3 of the video data are randomly selected for training and the rest are used for testing. Accordingly, for each empirical study, totally five decision models are constructed and tested with the corresponding testing data sets. The performance is then compared with a set of well-known classification methods, such as SVM, Decision Trees (C4.5), Neural Network (NN), Multinomial Logistic Regression Model (MLR), and One Rule Classifier (OR), which are enclosed in the WEKA package [16]. Three evaluation metrics, recall (R), precision (P), and F1 measure (F), are adopted. In the literature, the pair of recall and precision is generally used. However, as it is always possible to sacrifice one metric value in order to boost the other, the F1 measure, which is a combination of recall and precision and is defined as 2RP/(R+P), is deemed as a better performance metrics.

3.2. Performance comparison

As presented in Section 2, the video sources are processed via syntactic analysis. Then in the subspace data pruning component, after 50 times random selection and comparison (i.e., K=50), the selected typical goal, corner, sports, and commercial instances can recognize 100% corresponding concept instances and reject about 85%, 72%, 74%, and 36% non-concept instances, respectively. On the other hand, the selected typical negative instances can reject 100% concept instances and recognize about 80%, 72%, 83%, and 67% non-concept ones. As can be seen, a large portion of non-concept instances can be pruned and thus to alleviate the data imbalance issue. The cleaned data are then passed to the C-RSPM classification component.

Table 1. Performance comparison

		C-RSPM (%)	SVM (%)	C4.5 (%)	NN (%)	MLR (%)	OR (%)
goal	R	**85.0**	35.1	70.3	67.6	48.6	48.6
	P	69.1	**100.0**	81.3	75.8	81.8	72.0
	F	**75.7**	52.0	75.4	71.4	61.0	58.1
corner	R	**66.7**	0.0	28.7	23.0	0.0	26.4
	P	32.4	0.0	**78.1**	60.6	0.0	67.6
	F	**43.6**	0.0	42.0	33.3	0.0	38.0
sports	R	**81.7**	39.7	61.9	55.6	49.2	58.7
	P	58.5	86.2	70.9	85.4	**100.0**	80.4
	F	**68.2**	54.3	66.1	67.3	66.0	67.9
commercial	R	**94.8**	86.2	83.5	79.0	84.4	72.4
	P	75.2	75.9	**77.3**	76.1	76.3	66.7
	F	**83.9**	80.7	80.3	77.5	80.1	69.4

Table 1 shows the performance comparison. The best performance (P, R, and F) for each concept across all the classification methods are shown in bold fonts. From this table, we have the following observations. First, C-RSPM always achieves the best recall values in all the test cases. As we discussed before, recall is generally considered to be more important than precision in concept detection. Second, though some other classification approaches can yield better precision than C-RSPM, our F1 measure is always the

best in all the cases, which captures the system overall performance in a more complete manner. In summary, C-RSPM outperforms all the other classification approaches used in the experiments for concept detection. The proposed framework is very promising in the sense that it works automatically for concept detection by using only seventeen low-level features and without or with limited dependency on domain knowledge.

4. Conclusion

In this paper, an effective multi-concept detection framework is presented, which consists of the syntactic video analysis, subspace-based data pruning, and C-RSPM classification components. Our proposed framework has the following unique characteristics. First, it intelligently integrates the strengths of multi-modal video analysis and data mining method and seeks to bridge the semantic gap without or with limited dependency on domain knowledge. Second, it effectively addresses the data imbalance issue by the subspace-based data pruning component. Third, it efficiently enables a large portion of calculation results from the data pruning component to be re-used in the C-RSPM component since both components are subspace-based. The comparative experiments on various concept detections from a large set of video clips demonstrate the effectiveness and generalization of our proposed framework.

5. Acknowledgement

For Shu-Ching Chen, this work was supported in part by NSF HRD-0317692, NSF OISE-0730065, and Florida Hurricane Alliance Research Program sponsored by the National Oceanic and Atmospheric Administration. For Mei-Ling Shyu, this research was supported in part by National Oceanic and Atmospheric Administration (NOAA). This research was carried out in part under the auspices of the Cooperative Institute for Marine and Atmospheric Studies (CIMAS), a Joint Institute of the University of Miami and NOAA, cooperative agreement #NA17RJ1226. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the funding agency.

6. References

[1] S.-C. Chen, M.-L. Shyu, and C. Zhang, "Innovative Shot Boundary Detection for Video Indexing," *Video Data Management and Information Retrieval*, S. Deb, Ed. Hershey, PA: Idea Group Publishing, pp. 217-236, 2005.

[2] S.-C. Chen, M.-L. Shyu, C. Zhang, and R. L. Kashyap, "Identifying Overlapped Objects for Video Indexing and Modeling in Multimedia Database Systems," *International Journal on Artificial Intelligence Tools*, vol. 10, no. 4, pp. 715-734, 2001.

[3] M. Chen, S.-C. Chen, M.-L. Shyu, and K. Wickramaratna, "Semantic Event Detection via Temporal Analysis and Multimodal Data Mining," *IEEE Signal Processing Magazine*, Special Issue on Semantic Retrieval of Multimedia, vol. 23, no. 2, pp. 38-46, March 2006.

[4] S.-C. Chen, M.-L. Shyu, C. Zhang, and M. Chen, "A Multimodal Data Mining Framework for Soccer Goal Detection Based on Decision Tree Logic," *International Journal of Computer Applications in Technology*, vol. 27, no. 4, pp. 312-323, 2006.

[5] L.-Y. Duan, et al., "A Mid-level Representative Framework for Semantic Sports Video Analysis," *Proceedings of ACM Multimedia*, pp. 33-44, 2003.

[6] R. Leonardi, P. Migliorati, and M. Prandini, "Semantic Indexing of Soccer Audiovisual Sequences: A Multimodal Approach Based on Controlled Markov Chains," *IEEE Transactions on Circuit and System for Video Technology*, vol. 14, no. 5, pp. 634–643, 2004.

[7] W.-N. Lie, T.-C. Lin, and S.-H. Hsia, "Motion-Based Event Detection and Semantic Classification for Baseball Sport Videos," *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. 3, pp.1567-1570, 2004.

[8] T. Quirino, Z. Xie, M.-L. Shyu, S.-C. Chen, and L. Chang, "Collateral Representative Subspace Projection Modeling for Supervised Classification," *Proceedings of the IEEE International Conference on Tools with Artificial Intelligence*, pp. 98-105, 2006.

[9] D. 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.

[10] 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*, Special Issue on Multimedia Data Mining, Vol. 10, No, 2, pp. 252-259, February 2008.

[11] E.A. Tekalp and A.M. Tekalp, "Generic Play-Break Event Detection for Summarization and Hierarchical Sports Video Analysis," *Proceedings of IEEE Conference on Multimedia and Expo*, pp. 169-172, 2003.

[12] D. Tjondronegoro, Y.-P. Chen, and B. Pham, "Content-based Video Indexing for Sports Analysis," *Proceedings of ACM International Conference on Multimedia*, pp. 1035-1036, 2005.

[13] http://www-nlpir.nist.gov/projects/trecvid/.

[14] L. Wang, M. Lew, and G. Xu, "Offense Based Temporal Segmentation for Event Detection in Soccer Video," *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 259-266, 2004.

[15] J. Wang, C. Xu, E. Chng, K. Wah, and Q. Tian, "Automatic Replay Generation for Soccer Video

Broadcasting," *Proceedings of ACM Multimedia*, pp. 311-314, 2004.
[16] WEKA: http://www.cs.waikato.ac.nz/ml/weka/.
[17] S. Wold, *Pattern Recognition*, no. 8, pp. 127-139, 1976.
[18] L. Xie, S.-F. Chang, A. Divakaran, and H. Sun, "Unsupervised Discovery of Multilevel Statistical Video Structures Using Hierarchical Hidden Markov Models," *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, vol. 3, pp. 29–32, 2003.
[19] D.-Y. Yeung and C. Chow, "Parzen-Window Network Intrusion Detectors," *Proceedings of International Conference on Pattern Recognition*, vol. 4, pp. 403-405, 2002.
[20] D. Zhang and S.-F. Chang, "Event Detection in Baseball Video Using Superimposed Caption Recognition," *Proceedings of ACM Multimedia*, pp. 315-318, 2002.