http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 11 (10): 1381-1390, 2012
ISSN 1812-5638 / DOI: 10.3923/itj.2012.1381.1390

A Framework Based on Multi-models and Multi-features for Sports Video Semantic Analysis

Jiaqi Fu, Hongping Hu, Richao Chen and Heng Ren
School of Information Sciences and Engineering, Hunan University, No. 252, Lushan South, Road, Changsha, 410082, China

Abstract: The proliferation of video posed a challenging problem for the automatic analysis, interpretation and indexing of video data. Among them, sports video analysis has attracted the most attention because of the appeal of sports to large audience. This study presented an effective sports video semantic analysis algorithm based on the fusion and interaction of multi-models and multi-features. By utilizing the semantic color ratio, the video shot was classified into global shot, in-field shot and out-of-field shot which facilitated the HMM-based classification. For shot corresponding to a specific scene, by introducing image registration, the artifacts of noise and camera movement were reduced and accurate local motion features were obtained. Then, Hidden Markov Models (HMMs) were exploited to associate every video shot with a particular semantic class. Experimental results on Football and Tennis sequence showed that the proposed approach can achieve a relatively high ratio of correct semantic recognition.

Key words: Sports video, semantic analysis, multi-models, multi-features, local motion, image registration

INTRODUCTION

Sports video plays a great role in our daily life and has a wide range of audiences which have resulted in enormous growth in the amount of sports video generated everyday. On the other side, the content of sports video is usually redundant so that the highlight points in them distributed sparsely. Therefore, efficient browse and retrieval tools are needed to facilitate the management and visit of large-scale sports video data (Haering *et al*., 2008). However, digital video is a special visual data with specific spatial-temporal structure and its data amount is massive. The conventional analysis, management and retrieval methods for text data are not suitable for digital video (Renuga and Sadasivam, 2009). The need for automatic analysis techniques regarding the description of sports video content at the level of semantic emerges as a demanding and instant issue. To this end, semantic analysis for sports video becomes one of hot research topics of video processing.

In recent years, a wide variety of content-based sports video analysis approaches have been proposed. They utilize those low-level features, such as color, texture, shape and motion, to achieve event detection, shot classification and video annotation. The most typical works can be stated as follows. The player, line marks and motion features are utilized to identify specific events in broadcast soccer video (Gong *et al*., 1995). Speech pitch of judges and baseball batting sound are employed to detect exciting segments in baseball matches (Rui *et al*., 2000). Cinematic features such as shot type and moving object features are used to detect target event in broadcast soccer video by Bayesian Network classifier (Ekin *et al*., 2003). Cheng and Hsu (2006) exploits a novel likelihood-model based representation method, to measure the "likeliness" of low-level audio features and motion features to a set of predefined audio types and motion categories, respectively. Moreover, it utilizes a Hidden Markov Model (HMM) to model and extract baseball game highlights based on audio-motion integrated cues.

Attempting to bridge the so-called semantic gap which exists between low-level features and high-level semantic concept (Lavee *et al*., 2009), some semantic analysis approaches for sports video have been proposed. They are mainly divided into three categories: probability statistics based, statistical learning based and rule reasoning based approaches. However, there are two prominent drawbacks for them. First, the knowledge and rules specific to certain domain are overemphasized which limits their usages and constrains their generality. Second,

Corresponding Author: Jiaqi Fu, School of Information Sciences and Engineering, Hunan University, No. 252, Lushan South, Road, Changsha, 410082, China

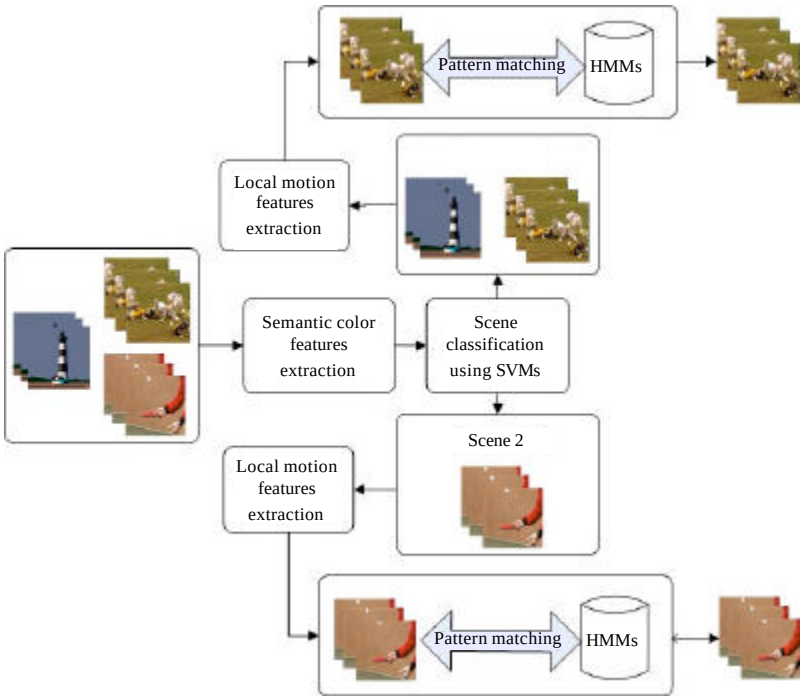

Fig. 1: Proposed system framework for sports video analysis

the feature extraction, processing and analysis are performed globally for most algorithms which increase their computational complexity. For example, the motion feature is extracted from the whole motion field and the visual feature is extracted from the whole frame. To address the above-mentioned issues, a few video analysis approaches which extract domain-specific local features, are proposed. A semantic video analysis approach which is based on the statistical estimation and representation of motion signal, is proposed by Papadopoulos *et al*. (2009). To identify which motion originates from true motion rather than the measurement noise, the kurtosis of the motion energy estimated using optical flow method, is calculated (Papadopoulos *et al*., 2009). In addition, a new representation is also presented to feed local-level motion information into HMM. The computational complexity is significantly reduced due to the processing of local-level motion information. However, it neglects some other important visual features which limits its performance to some extent. In fact, different from single-modal text and image data, digital video is more complicated multi-modal data, including visual features, texture features and motion features (Zhang *et al*., 2010). Apparently, the fusion and interaction of multi-models will play important roles in bridging the semantic gap between low-level features and high-level semantic concept (Lavee *et al*., 2009).

From the above analysis, we observe that the motion feature and corresponding motion region will be extracted more accurately if the external interferences are reduced. Moreover, due to the multi-modal properties of video data, multiple models should be connected and fully utilized to recognize the semantic concept (Jin *et al*., 2011). To this end, a framework combining both motion statistics and multi-models for semantic classification of sports video shot is proposed in this study.

The block diagram of the proposed sports video semantic analysis approach is presented in Fig. 1. It does not depend on any domain-specific knowledge and rules. The scene classification is conducted by SVMs using semantic color feature. For shot corresponding to a specific scene, the motion energy is filtered by image registration and the true motion is obtained by further by morphological operation. This will effectively reduce the effects due to noises and camera movement. Then, a local feature represented using the method (Papadopoulos *et al*., 2009) is put into HMMs to associate each shot with one of the semantic classes that of interest.

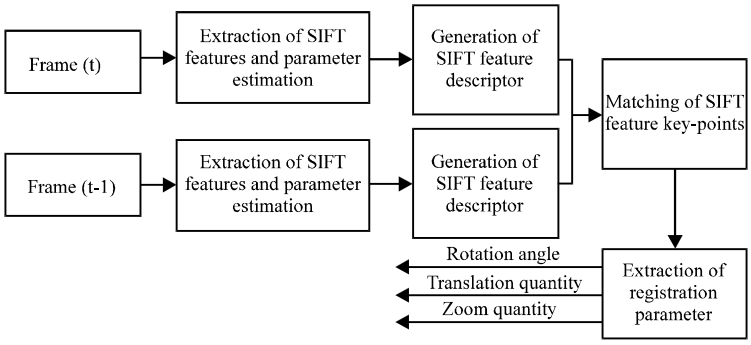

Fig. 2: Flow chart of image registration

EXTRACTION AND STATISTICAL ANALYSIS OF MOTION FEATURE

The motion features of digital video contain rich semantic information. For example, event detection in the domain of video semantic analysis usually utilizes the motion features. The effective extraction of statistical motion information and its concise representation is a key technique for video semantic analysis. In this section, the motion energy for all pixels is calculated from the motion vector which is extracted based on Scale Invariant Feature Transform (SIFT) and differential multiplication. And the true active pixels are filtered out by setting threshold and mathematical morphology. They are usually distributed in some local areas which facilitate local processing.

The pre-processing of motion analysis: In this study, video shot is chosen as the basic unit for semantic analysis. The examined video sequence is initially segmented into shots and the resulted shots set is denoted by $S = \{s_i, i = 1,2,...,I\}$. Each shot will be mapped to a class of semantic set which is denoted by $E = \{e_j, i = 1,2,...,J\}$. For the ith shot s_i, it is further divided into a set of video segment W_i with an non-overlapping time interval of TW, $W_I = \{\omega_{ir}, r = 1,2,...R_i\}$. Frames in each time interval ω_{ir} constitute the basic unit for the following feature process.

Local motion extraction: As one of the main motion detection algorithms, frame difference algorithm is simple and easy to implement but the detection results are not always accurate enough when the video camera moves (Elgammal *et al*., 2002). Therefore, the image SIFT-based registration method which is applied to calculate transformation parameters, including translation factors, rotation angles and scaling coefficients, is introduced to improve the detection accuracy. Furthermore, the fact that sports video is characteristic of camera that usually moves simply or even never, makes the registration more feasible.

Image registration based on SIFT features: The movement of camera can be classified into translation, rotation and zoom which can be modeled by affine transformation (Yedjour *et al*., 2011). Simultaneously, in order to make a good compromise between the need for time efficiency and accuracy of motion description, the 6-parameter affine model is adopted. Assume that the locations of corresponding key-points in frame (t-1) and the frame t are denoted by (f_x^{t-1}, f_y^{t-1}) and (f_x^t, f_y^t), respectively. Then, the affine transformation can be represented as follows:

$$\begin{bmatrix} f_x^t \\ f_y^t \end{bmatrix} = \begin{bmatrix} a_0 f_x^{t-1} + a_1 f_y^{t-1} + a_2 \\ a_3 f_x^{t-1} + a_4 f_y^{t-1} + a_5 \end{bmatrix} \quad (1)$$

where, a_0, a_1, a_2, a_4 and a_5 represents the rotation and zoom of image while, a_2 and a_5 represents the translation.

In this study, the matching algorithm proposed by Lowe (1999) is utilized to estimate the SIFT matched pairs in two video frames. Additionally, because of the tiny differences between two frames in one shot, there are redundant SIFT matched pairs to estimate the transformation parameters using the least square parameters identification. The flow chart of image registration is shown in Fig. 2 (Mei *et al*., 2011).

The extraction of activity area: After registration, suppose $I_{k-1}(x,y)$ and $I_k(x,y)$ represent two adjacent frames, the motion vector can be defined by:

$$\vec{V}(x,y) = I_k(x,y) - I_{k-1}(x,y) \quad (2)$$

Compared with motion vector, the motion energy can provide richer information about motion-based semantic

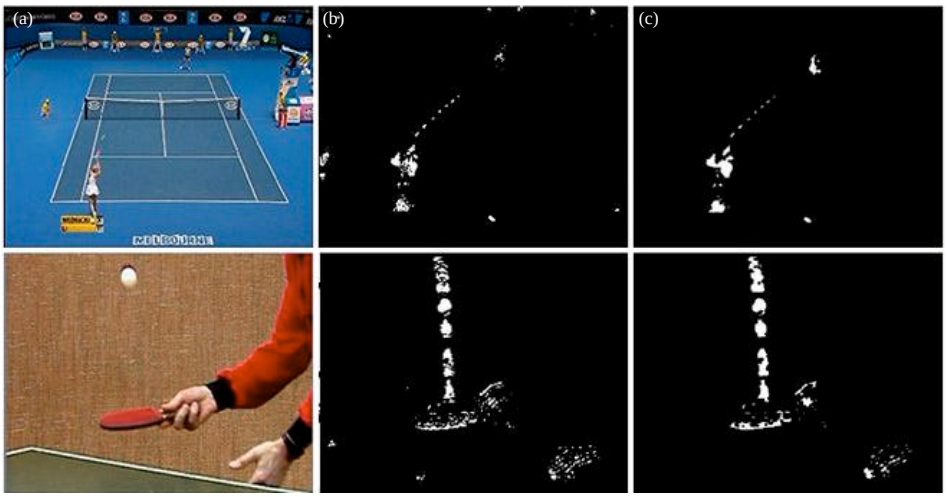

Fig. 3(a-c): Comparison of extraction of activity area in two conditions: using differential multiplication or not, (a) First frame of sequence, (b) Binary image without different multiplication and (c) Binary image using differential multiplication

(Xu *et al*., 2005). Therefore, the corresponding motion energy field is calculated as follows:

$$M(x,y) = \left\| \bar{V}(x,y) \right\| \tag{3}$$

where, $\|\cdot\|$ denotes the norm of a vector and M(x,y) is the resulting motion energy field.

Then, the activity area can be segmented by frame difference and setting a threshold as follows:

$$A_{k-1,k}(x,y) = \begin{cases} 1 & M(x,y) > TH \\ 0 & others \end{cases} \tag{4}$$

In the experiment, there are many noisy pixels in binary image, due to high-frequency noise, illumination variance and the small changes of background. Additionally, the image registration may also introduce errors. However, it is observed that there must be stable motion overlap area between adjacent binary images. Therefore, supposing the noisy pixels distribute randomly, the accurate motion area can be extracted by differential multiplication of multiple binary images. In order to make a good compromise between the need for time and efficiency and accuracy of motion extraction, the differential multiplication for 4 consecutive frames are utilized which can be represented as follows:

$$A_{k-1,k+2} = A_{k-1,k+1}(x,y) + A_{k,k+2}(x,y) \tag{5}$$

To illustrate the effects of noise suppression, experimental results on Tennis and Table-tennis sequence with additive Gaussian noise are shown in Fig. 3. Additionally, after differential multiplication, the binary image is further processed by morphological operation such as erosion, filling and the effect is shown in Fig. 3c.

Motion representation: HMM has good adaptability and performs well in predicting random time-series data. However, majority of the HMM-based semantic analysis methods in the literature are focusing only on global-level motion representation. In fact, if local-level analysis of motion information is suitably exploited, it can provide significant cues for semantic analysis. Moreover, it is also beneficial for the reduction of computational complexity.

According to the binary image A(x,y) obtained, we define a Minimum Bounding Rectangle (MBR) $A^L(x_l,y_l)$ to surround all active pixels, where:

$$x_l \in [x^{L0}, x^{L1}]\ (1 \le x^{L0} \le x^{L1} \le V_{dim}, y_l \in [y^{L0}, y^{L1}](1 \le y^{L0} \le y^{L1} H_{dim})$$

And the corresponding local energy field of MBR is defined as $M^L(x_l,y_l)$, as shown in Fig. 4 (Papadopoulos *et al*., 2009).

The estimated local energy field is usually of high dimensionality. It is down-sampled and denoted as $\hat{M}(x,y)$. Meanwhile, according to the HMM theory (Rabiner, 1989), the set of sequential observation vector that constitute an observation sequence need to be of fixed length and low dimensionality. To this end, the aforementioned $\hat{M}(x_l,y_l)$ field is approximated by a 2-D polynomial function based on generalized least square criteria and the coefficients of each order polynomial constitute the motion feature which will be served as the observation data for HMMs.

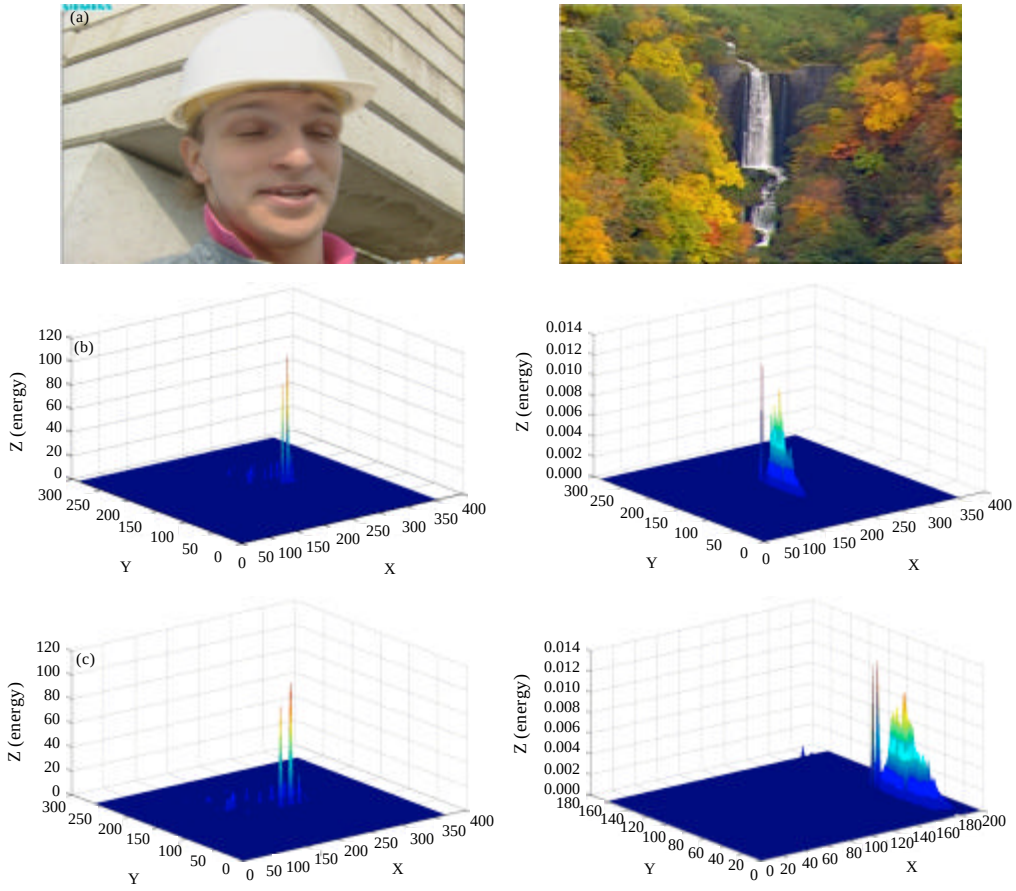

Fig. 4(a-c): Binary image (a) First frame video sequence energy field calculation in global and local area, (b) Global area M(x,y) and (c) Local area $M_L(x_l,y_l)$

EXTRACTION AND STATISTICAL ANALYSIS OF COLOR FEATURE

Generally, there are a few of basic color in sports video (Abdulfetah *et al.*, 2010). Take football video for example, color of grass, background color of audience and color of player's uniform constitute the main part of each video frame. This phenomenon implies some important semantic information which was defined as semantic color.

For sports video, it is observed that the ratio of these semantic colors to the colors of full frame varies greatly with different shot type. For example, in the domain of football video, the ratio of grass color is much higher than any other semantic color in the global-view shot while the ratio reduces sharply in the close-view shot. Therefore, this clue might be helpful to the classification shots. And in this study, we utilize simply the background color ratio to classify the global shot, in-field shot and out-of-field shot.

Supposing H semantic C_i (i = 1,2,...H) colors were pre-defined. For each frame with the size of m×n pixels, the semantic color ratio feature can be expressed as a one-dimensional vector CR with H elements, of which each elements CR(i), (i = 1,2,...H)corresponding to one specific semantic color C_i, (i = 1,2,...H). And the specific ratio CR(i), (i = 1,2,...H) is defined as follows:

$$CR(i) = \frac{NC_i}{m \times n} \quad (6)$$

where, Nc_i represents the number of pixel with the semantic color C_i. Considering the computational complexity, the color ratio is estimated in terms of image blocks in this section. Simultaneously, because of the fact that pixels with the same semantic color distributed continuously in an image, only some eligible ones can be selected which can perform better in robustness. The process is presented as follows:

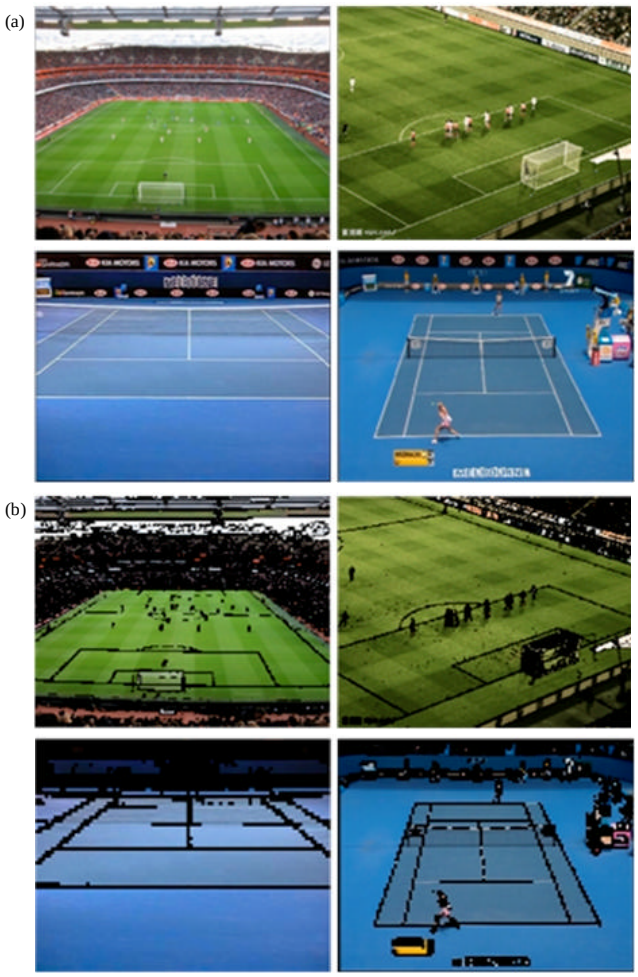

Fig. 5(a-b): Block selection, (a) Original image and (b) Processed image in which the undesirable blocks were painted into black

- (1) The input image will be segmented into a series of blocks with the size of p×q. For each block B(i,j), the mean color mean(i,j) and color covariance cov(i,j) are calculated. In order to reduce the computational complexity, only the blocks whose color covariance is smaller than a given threshold TH will be selected. As shown in Fig. 5, the undesirable blocks are painted into black
- Considering the fact that there exist a non-negligible difference between RGB color space and Human Visual Perception (Ishak *et al*., 2006), RGB space should be transformed into HSV space for the pixels in the selected blocks. Then, each component of H, S and V is quantized using nonlinear quantitative method. Meanwhile, to effectively reduce the computational complexity, the three components are divided into 8, 3 and 3 quantization levels, respectively:

$$H=\begin{cases}0, & h\in[316,20]\\ 1, & h\in[21,40]\\ 2, & h\in[41,75]\\ 3, & h\in[76,155]\\ 4, & h\in[156,190]\\ 5, & h\in[191,270]\\ 6, & h\in[271,295]\\ 7, & h\in[296,315]\end{cases}$$

$$S=\begin{cases}0, & s\in[0,0.2]\\ 1, & s\in[0.2,0.7]\\ 2, & s\in[0.7,1]\end{cases}$$

$$V=\begin{cases}0, & v\in[0,0.2]\\ 1, & v\in[0.2,0.7]\\ 2, & v\in[0.7,1]\end{cases} \quad (7)$$

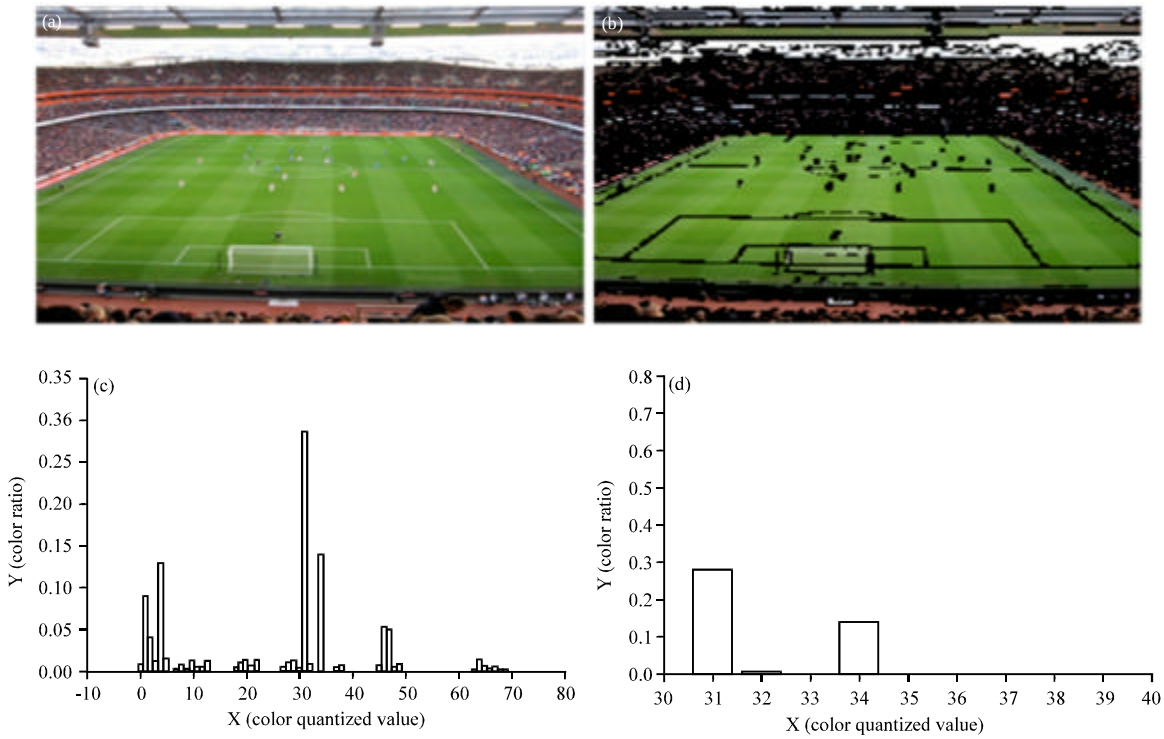

Fig. 6(a-d): The process of color ratio calculation in the quantized HSV space, (a) Original frame, (b) Processed frame, (c) Color ratio of all colors and (d) Color ratio of semantic colors

After quantification, a one-dimensional vector, $G = HQ_SQ_V$, is generated from the three color components, where Q_S and Q_V are quantitative series of S and V, respectively. In this study, $Q_S = 3$, $Q_V = 3$, so $G = 9H+3S+V$. Thus, the color space has been divided into 72 quantized intervals which vary in the interval [0,1,...71], as shown in Fig. 6c. Take football video for example, we predefine the grass color as the semantic color. After estimation and experiment, we found that the quantized HSV sub-space located at the range of [30,31,...40] is corresponding to the semantic colors in football video, as shown in Fig. 6d. Each ratio of these semantic colors constitutes the extracted one-dimensional color feature vector which will be served as input for SVM.

MODELLING ANALYSIS BASED ON COLOUR AND MOTION

In video processing field, video shot is not merely a sequence of images. The temporal context information in videos is a significant cue for content understanding. The purpose of video semantic analysis may be of discovering the hidden states or semantics behind video signals. To this end, HMM is chosen as the statistical tool for its fitness for solving problems that exhibit an inherent temporality (Ping *et al*., 2009). However, a mass of frames are needed to be processed in most HMM-based applications which usually cause magnificent computational complexity. Therefore, a simple scene classification based on semantic color before the HMM-based application is introduced; it can also improve the accuracy of classification to some extent. And SVM is employed to conduct the scene classification, duo to the fact that SVM has been one of the discriminative classification tools which is generally admitted to be more accurate (Lee *et al*., 2010).

The semantic modeling construction is shown in Fig. 7. First, for every shot s_i, the first frame in shot s_i is picked out to extract the semantic color feature using the method. Then, the SVM is employed to decide which scene the shot belongs to. After scene classification, a set of HMMs are utilized to extract specific semantic of the estimated scene.

What should be noted is the HMM-based modeling procedure. More details will be stated as follows: We modeled each pre-defined shot by a HMM, denoted by (λ_j) $(1 \le j \le J)$ and the shot classification is based on the well-known Viterbi algorithm. As shown in Fig. 8, features extracted from the video stream constitute the respective observation interval ω_r of shot s_i, they are used to form a single observation vector. And these vectors for all ω_r of shot s_i, form a respective shot observation sequence.

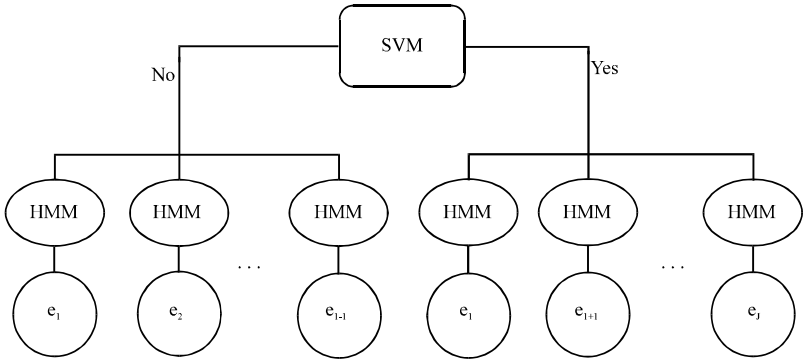

Fig. 7: The basic block diagram of semantic analysis model

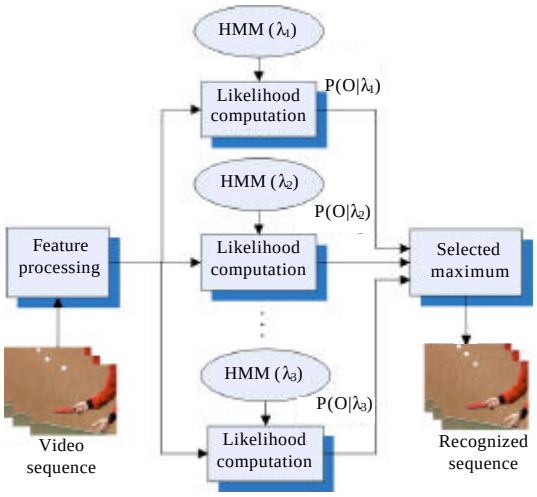

Fig. 8: The general block diagram of modeling by HMMs

Then, in order to perform the association of the examined shot s_i with the defined classes e_j, based on the computed shot observation sequence, a set of J HMMs is employed, where an individual HMM is introduced for every defined class e_j. More specifically, each HMM receives the aforementioned observation sequence O as input and estimates a posterior probability $P(O|\lambda_j)$ which indicates the degree of confidence with which class e_j is associated with shot s_i. And the class e_j corresponding to the maximum of all $P(O|\lambda_i)$ is the estimated shot class.

EXPERIMENTAL RESULTS AND ANALYSIS

Comparisons of experimental results are made between the proposed approaches with MEIs (Bobick and Davis, 2001). The experimental conditions are as follows. First, 4 videos of "Australia Tennis Open" and "America Tennis Open" and 3 videos of "South Africa World Cup" are collected for experimental evaluation. We define 4 classes of semantic events for Football shot: "Break", "Run" and "Shoot", "Slow-motion Replay" and 4 classes for Tennis shot: "Break", "Rally", "Serve" and "Slow-motion Replay". Then, the test videos are segmented into 330 shots manually (Tennis: 122, 46, 27, 19; Football: 17, 46, 32 21). Third, the modeling process and the experimental platform are built as Fig. 9 and 1, respectively. The time interval is set as: ω_r = 0.4 sec. The statistical model is Gaussian of mixture first-order hidden Markov model.

Table 1: Semantic class association results in the football domain

		Associated class			
Method	Actual class	e_1 (%)	e_2 (%)	e_3 (%)	e_4 (%)
Overall accuracy: 87.2%					
Proposed approach	e_1	99.7	0.0	0.3	0.0
	e_2	0.2	85.3	2.4	12.1
	e_3	0.0	3.2	84.1	12.7
	e_4	3.8	7.1	9.2	79.9
Overall accuracy: 77.7%					
MEIs	e_1	96.3	2.6	0.4	0.7
	e_2	4.6	84.6	13.2	17.6
	e_3	4.0	6.2	72.7	17.1
	e_4	1.0	21.1	20.6	57.3

e_1: Break, e_2: Run, e_3: Shoot and e_4: Slow-motion replay

Table 2: Semantic class association results in teunis domain

		Associated class			
Method	Actual class	e_1 (%)	e_2 (%)	e_3 (%)	e_4 (%)
Overall accuracy: 78.9%					
Proposed approach	e_1	91.6	2.3	4.4	1.7
	e_2	0.4	88.1	3.6	7.9
	e_3	0.8	6.2	73.6	19.4
	e_4	9.4	12.6	15.9	62.1
Overall accuracy: 63.8%					
MEIs	e_1	49.6	10.5	13.4	26.5
	e_2	2.9	87.3	4.7	5.1
	e_3	1.6	12.5	69.7	16.2
	e_4	46.3	3.2	1.6	48.9

e_1: Break, e_2: Rally, e_3: Serve and e_4: Slow-motion replay

Table 1 and 2 list the semantic analysis results for Football video and Tennis video, respectively. It can be observed that from the viewpoints of average semantic

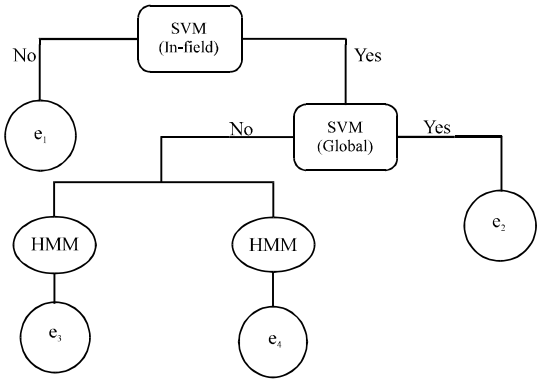

Fig. 9: The specific block diagram of semantic model for football and tennis video based on the basic block diagram

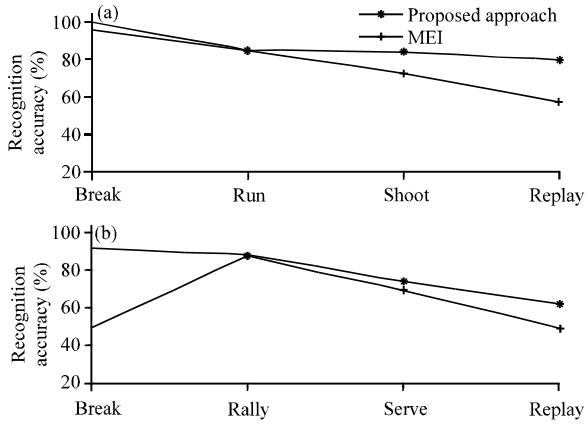

Fig. 10(a-b): The comparison of shot classification between the proposed approach and MEIs in domains of (a) Football and (b) Tennis according to Table 1 and 2

recognition accuracy, the proposed approach is 9.5 and 15.1% better than MEIs. The main reasons for this improvement are two-fold: First, compared with single model, the fusion of multiple models can extract more rich semantic information from low-features which can also reduce the dependence degree for any of statistical models. Second, the introduction of image registration and mathematical morphology reduces the interferences due to noise and shot movement. It improves the accuracy of motion region and motion feature extraction to some extent.

It is observed from Fig. 10 that, these two approaches are high in the accuracy of recognizing "Break", "Run" and "Rally" shots. This is because that the ratio of semantic color is the minimum or maximum in these shots which can be classified easily using SVMs. However, for the recognition of "Shoot", "Serve" and "Slow-motion Replay" shots, their precisions are greatly decreased mainly because of the mis-classifications between them. By careful observations, it can be found that the shots of "Slow-motion Replay" are mainly slow-motion replay the shots of "Shoot", "Serve" and so on. Therefore, it will lead to confused semantics. However, the proposed approach can still achieve an average accuracy of about 82.0% and 67.9% which is 17.0% and 8.6% higher than MEIs methods, respectively.

CONCLUSIONS

In this study, an effective algorithm for sports video semantic analysis is presented. It is based on the fusion and interaction of multiple models. By utilizing the semantic color ratio, the video shot is classified into global shot, in-field shot and out-of-field shot which facilitates the HMM-based classification. For shot corresponding to a specific scene, by introducing image registration, the artifacts of noise and camera movement are reduced and accurate local motion features are obtained. Then, Hidden Markov Models (HMMs) are exploited to associate every video shot with a particular semantic class. Experimental results show that the proposed approach can achieve a relatively high accuracy of semantic recognition.

ACKNOWLEDGMENTS

This study was supported by National Natural Science Foundation of China (Grant No. 61072122), Special Pro-phase Project on National Basic Research Program of China (Grant No. 2010CB334706) and Program for New Century Excellent Talents in University (Grant No. NCET-11-0134).

REFERENCES

Abdulfetah, A.A., X. Sun and H. Yang, 2010. Robust adaptive video watermarking scheme using visual models in DWT domain. Inform. Technol. J., 9: 1409-1414.

Bobick, A.F. and J.W. Davis, 2001. The recognition of human movement using temporal templates. Proc. IEEE Trans. Pattern Anal. Mach. Intel., 23: 257-267.

Cheng, C.C. and C.T. Hsu, 2006. Fusion of audio and motion information on HMM-based highlight extraction for baseball games. IEEE Trans. Multimedia, 8: 585-599.

Ekin, A., A.M. Tekalp and R. Mehrotra, 2003. Automatic soccer video analysis and summarization. IEEE Trans. Image Process., 12: 796-807.

Elgammal, A., R. Duraiswami, D. Harwood and L.S. Davis, 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proc. IEEE, 90: 1151-1163.

Gong, Y., L.T. Sin, C.H. Chuan, H. Zhang and M. Sakauchi, 1995. Automatic parsing of TV soccer programs. Proceedings of the International Conference on Multimedia Computing and Systems, May 15-18, 1995, Washington, DC USA., pp: 167-174.

Haering, N., P.L. Venetianer and A. Lipton, 2008. The evolution of video surveillance: An overview. Machine Vision Appl., 19: 279-290.

Ishak, K.A., S.A. Samad and A. Hussain, 2006. A face detection and recognition system for intelligent vehicles. Inform. Technol. J., 5: 507-515.

Jin, B., L. Feng, H.J. Piao and Z.Q. Diao, 2011. Research on information fusion model for patent retrieval. Inform. Technol. J., 10: 164-167.

Lavee, G., E. Rivlin and M. Rudzsky, 2009. Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video. IEEE Trans. Syst. Man. Cybern., 39: 489-504.

Lee, L.H., C.H. Wan, T.F. Yong and H.M. Kok, 2010. A review of nearest neighbor-support vector machines hybrid classification models. J. Applied Sci., 10: 1841-1858.

Lowe, D.G., 1999. Object recognition from local scale-invariant features. Proc. 7th IEEE Int. Conf. Comput. Vision, 2: 1150-1157.

Mei, W., T. Dawei and Z. Xuchao, 2011. Moving object detection by combining SIFT and differential multiplication. Optics Precision Eng., 19: 892-899.

Papadopoulos, G.T., A. Briassouli, V. Mezaris, I. Kompatsiaris and M.G. Strintzis, 2009. Statistical motion information extraction and representation for semantic video analysis. IEEE Trans. Circuits Syst. Video Technol., 19: 1513-1528.

Ping, Z., T. Li-Zhen and X. Dong-Feng, 2009. Speech recognition algorithm of parallel subband HMM based on wavelet analysis and neural network. Inform. Technol. J., 8: 796-800.

Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE, 77: 257-286.

Renuga, R. and S. Sadasivam, 2009. Data discovery in grid using content based searching technique. Inform. Technol. J., 8: 71-76.

Rui, Y., A. Gupta and A. Acero, 2000. Automatically extracting highlights for TV baseball programs. Proceedings of the 8th ACM International Conference on Multimedia, October 30-November 3, 2000, New York, USA., pp: 105-115.

Xu, G., Y.F. Ma, H.J. Zhang and S.Q. Yang, 2005. An HMM-based framework for video semantic analysis. IEEE Trans. Circuits Syst. Video Technol., 15: 1422-1433.

Yedjour, H., B. Meftah, D. Yedjour and A. Benyettou, 2011. Combining spiking neural network with hausdorff distance matching for object tracking. Asian J. Applied Sci., 4: 63-71.

Zhang, Y., D. Wei and J. Wang, 2010. Semantic analysis for soccer video based on fusion of multimodal features. Comput. Sci., 37: 273-276.