A New Method to Segment Playfield and Its Applications in Match Analysis in Sports Video

Shuqiang Jiang[1,2]; Qixiang Ye[1,2]; Wen Gao[1,2]; Tiejun Huang[2]
[1](Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)
[2](Graduate School of Chinese Academy of Sciences, Beijing 100039, China)
Email:{sqjiang, qxye, wgao, tjhuang}@jdl.ac.cn

ABSTRACT
With the growing popularity of digitized sports video, automatic analysis of them need be processed to facilitate semantic summarization and retrieval. Playfield plays the fundamental role in automatically analyzing many sports programs. Many semantic clues could be inferred from the results of playfield segmentation. In this paper, a novel playfield segmentation method based on Gaussian mixture models (GMMs) is proposed. Firstly, training pixels are automatically sampled from frames. Then, by supposing that field pixels are the dominant components in most of the video frames, we build the GMMs of the field pixels and use these models to detect playfield pixels. Finally region-growing operation is employed to segment the playfield regions from the background. Experimental results show that the proposed method is robust to various sports videos even for very poor grass field conditions. Based on the results of playfield segmentation, match situation analysis is investigated, which is also desired for sports professionals and longtime fanners. The results are encouraging.

Categories and Subject Descriptors
I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis – color, object recognition. I.4.6 [**Image Processing and Computer Vision**]: Segmentation – pixel classification, region growing, partitioning.

General Terms
Algorithms, Experimentation.

Keywords
GMMs, Region growing,Sports video, Match analysis

1. INTRODUCTION
Sports video always appeals to large audiences. In recent years, the amount of digitized video content has been increasing rapidly, and users need to access these content through various network solutions by various digital equipments. Therefore, automatically extracting useful information from sports video to facilitate retrieval and organization is an important problem. In fact, this has emerged as a hot research area during the past few years.

Gong's work [1] use object color and texture features to parse TV soccer programs. In [2], the authors use playfield zone classification, camera motion analysis and player's position to infer highlights of soccer video by FSM (finite state machine). While Hanjalic[3] use audio, motion and shot features to extract soccer highlights. In [4], Xie et al. firstly detect dominant ratio using HSV (Hue-Saturation-Value) color model, then calculate the motion intensity on macro-block, and finally use HMM (hidden Markov model) to segment play and breaks. Duan and Tian et al. [5] propose a mid-level framework that can be used for event detection, highlight extraction, summarization and personalization of sports video, the information they employed include low-level features, mid-level representation and high-level event. In [6] cinematic feature as shot type, slow plays and object-based features are used to analyze and summarize sports video.

(a) (b) (c) (d) (e) (f)

Figure 1: Different appearances of playfield

It could be observed that playfield plays the fundamental role in analyzing many kinds of sports video such as soccer, tennis, and basketball. It conveys the basic information of the match course. Players and ball are always surrounded by playfield region; mouth goal (soccer), basket (basketball) and sideline are always at the boarder of the playfield region; and audiences always locate outside of this important region. Results of playfield segmentation could be used to classify different types of shots or frames [2], to identify players [10] [2] and ball [11], and to further extract highlight and events [4]. In these sports, the color of playfield is generally uniform and playfield often occupies dominant color region. Some researchers have used this knowledge to extract playfield; and most of them use histogram-based method [4][6][7][9][10]. In [4], the authors accumulate hue histograms of HSV color space on the initial 50 frames to yield definition of dominant color and use this definition to identify grass pixel of soccer programs. In [9][6], Ekin el al. proposes an algorithm to automatically learn the dominant color statistics of playfield by using two color spaces: a control space and a primary space and information from these two color spaces are combined. In [7], the authors establish the green color model as a convex set in HSI space, then the ratio of the number of pixels are computed as the criterion of the grass pixel in soccer program. Other playfield detection methods include Green Color Table (GCT) [8] and region growing methods [11].

Although playfield usually occupies dominant color regions in many sports videos, it may have various appearances due to different types of sports (Fig.1 (a)&(b)), different stadium (Fig.1 (b)&(c)), and different weather and lighting conditions within the same stadium (Fig.1 (c)&(d)). Besides, some playfields do not show consistence as a whole, there may have some patterns on them, as illustrated in Fig.1 (e)&(f). Therefore, the method chosen to estimate playfield regions needs to be sufficiently general in order to model various color distributions generated by different kinds of sport videos. In this paper, we will give a novel approach to robustly segment various kinds of playfield in sports video. The proposed method is based on Gaussian mixture models (GMMs) accompanied with a region-growing algorithm. Color histogram methods can be viewed as simple, non-parametric forms of density estimation in color space, and they are direct reflections of color distributions. The powerful attribute of GMMs is its ability to form smooth approximations to arbitrarily shaped densities. Density estimation using GMMs is performed in a semi-parametric way so that the number of mixture models scales with the complexity of the data rather than with the size of the data set. The GMMs method is sufficiently general to model highly complex, non-linear distributions. After the GMMs detection process, discrete playfield pixels are detected; noises may exist in them and they do not form regions. The function of region growing operation is to connect playfield pixels into regions, eliminate noises, and smooth the boundaries, thus makes the final segmentation results be competent for sports video content analysis. In a word, the proposed playfield segmentation method has the following two advantages: (1) It is robust independent of sports types and appearances even for very poor grass conditions; (2) the segmentation results can not only obtain the dominant color regions, but also well reserve objects in the playfield thus facilitating further analysis of sports match.

In the literature, the final goal of sports video analysis may have two main purposes: (1) to extract high-level events or objects (such as some sports stars) that users might interested in; (2) to generate highlight summaries in various aspects based on users' intentions. Besides, results of events and objects detection are helpful for highlight summaries generation. Playfield segmentation plays an important role to achieve the aims described above. While for professional sports person and most of the longtime sports fanners, match situation analysis is very useful for them. Who is superior between two teams in the course of the match or what strategy that a team makes use of in a specific time is desired. As the second contribution of the paper, match situation analysis is investigated on soccer video as the example. Given a series of video shots, playfield segmentation is first performed; then players and playfield in each frame of the shot are analyzed; and finally which team is superior in the shots is obtained. Match situation analysis proposed in this paper is not only an application to show the effectiveness of the playfield segmentation, but also explores a new and important aspect that automatic sports video analysis may be of help to the users.

The rest of the paper is organized as follows. Playfield segmentation is presented in detail in Section 2. Match analysis is discussed in Section 3. Section 4 concludes the paper.

2. PLAYFIELD SEGMENTATION

Figure 2 is the algorithm flowchart of the segmentation method. For each sports video clip, the sampling prepares training data for GMMs. About 100 frames are drawn as training data. As in most cases, playfield occupies largest regions in sport videos. The pixels are evenly sampled from each frame. Color distributions of pixels are modeled in hue-luminance space because hue reflects color of the court and luminance corresponds to the lightness of the color that reflects lighting conditions of the stadium. Through experiments, the HL color model is comparable with HSL and outperforms other color representation methods.

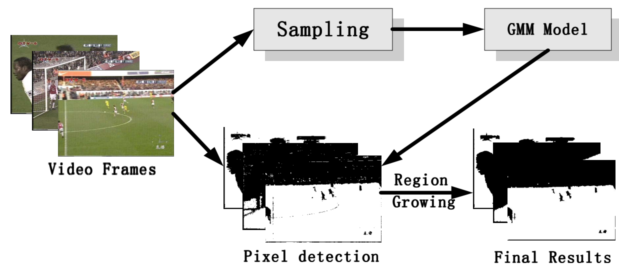


Figure 2: Flowchart of segmentation Algorithm

2.1 Gaussian Mixture Models

The conditional density of a pixel ξ belongs to the playfield region Φ is modeled with a convex combination of M Gaussian densities:

$$p(\xi \mid \Phi) = \sum_{i=1}^{M} w_i b_i(\xi)$$

where ω_i are the mixture weights, and $\sum_{i=1}^{M} w_i = 1$; $b_i(\xi)$ are mixture components, i=$1,2,\ldots,M$. Each component density is a Guassian with mean μ_i and covariance matrix Σ_i:

$$b_i(\xi) = \frac{1}{2\pi |\Sigma_i|^{1/2}} \exp\{-\frac{1}{2}(\xi - \mu_i)' \Sigma_i^{-1} (\xi - \mu_i)\}$$

Ordinary EM (Expectation Maximization) algorithm is used to estimate three parameters of the GMMs: Mean vectors, covariance matrices and mixture weights from all component densities. GMMs with 4 components are obtained after the training process. The above procedure could be conducted not only at the beginning of the program, but also in the course of the play. As an example, Figure 3(b) is the GMMs detection results of 3(a). Region growing is used after the GMMs detection.

2.2 Region Growing

After playfield pixel detection process, 2-D binary signals are obtained from the input frames where value 1 denotes for playfield pixel and 0 for non-playfield pixel. To get more refined detection result, region-growing procedure is used which is a general technique for image segmentation. Based on the traditional region growing methods, we propose a new region-growing algorithm to perform the segmentation [12]:

1) Search the unlabeled pixels in a binary image in order from the top left corner to the bottom down corner;

2) If a pixel ξ is not labeled, a new region is created. Then we iteratively collect unlabeled pixels that have the same value

and are connective to ξ. All these pixels are labeled with same region label, this label is same to the value of pixels;

3) If there are still existing unlabeled pixels in the image, go to 2);

4) If pixel number of region R bellows a given threshold, this region will be deleted and merged to the neighboring region. Threshold for regions labeled with 1 is different from regions with 0. This is because some regions labeled with 0 surrounded by playfield regions are meaningful information such as players. While in most cases the same size of regions labeled with 1 surrounded by non-playfield regions are nonsense and noise regions.

After the above process, regions labeled with 1 are identified as playfield. Figure 3(c) is the final segmentation result of 3(a). From this example, it could be observed that region growing could get better results compared with only using GMMs detection. More convincing analysis is given in next section through experiments.

(a) (b) (c)

Figure 3: A segmentation result

2.3 Segmentation Results

To show the validity of our method, more than 17 hours (18 clips) of various sports program from various sources are used. The clips are MPEG format including soccer, tennis, badminton, and basketball. To quantitatively evaluate the performance of the method, frames from 6 clips are manually annotated as the test set that consists of three soccer clips, one badminton, one tennis and one basketball clip.

The segment precision *SP* and segment confusion *SC* criterions are employed to evaluate the segmentation performance that is defined as follows:

SP= detected playfield pixels / annotated playfield pixels

SC= detected non-playfield pixels/ annotated non-playfield pixels

The overall experimental results are presented in Table2 and 3.

	Histogram	GMMs	Final Results
Soccer-1	71.8149%	85.4374%	86.1945%
Soccer-2	94.8604%	98.8452%	98.8945%
Soccer-3	95.2487%	98.8412%	99.2011%
Badminton	91.5885%	96.8472%	97.4873%
Tennis	87.8405%	95.6811%	96.7629%
Basketball	92.8569%	96.0376%	97.2847%

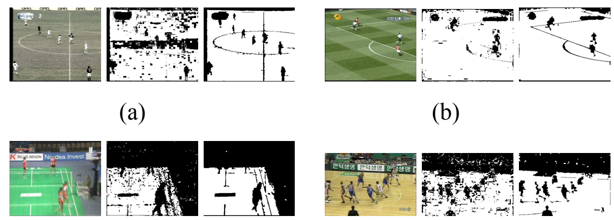
Table 2. *SP* results of various methods

	Histogram	GMMs	Final Results
Soccer-1	96.441%	95.6361%	97.3759%
Soccer-2	91.4%	88.8630%	90.1846%
Soccer-3	93.1318%	92.7659%	93.7582%
Badminton	88.9148%	98.6928%	99.0720%
Tennis	94.0844%	96.2096%	98.3313%
Basketball	82.9019%	91.4943%	95.5389%

Table 3. *SC* results of various methods

It could be observed that the segmentation performance of GMMs method is generally better than that of histogram method. The region growing process enhances the final segmentation performance. Entirely all the six *SP* and five *SC* results are higher that only using GMMs detection method alone.

Figure 4 is four segmentation results on various sports clips. On each example, the first one is the original frame, the second is the segmentation result with histogram method, and the last one is our method. The result not only enhances the performance of only using GMMs alone, but also gets more clear boundaries of players and other objects in the playfield regions. On clips with very poor grass conditions and with obvious court patterns, the method also exhibits good adaptability as illustrated in Figure 4(a) and 4(b).



(a) (b)

(c) (d)

Figure 4: Segmentation results of some frames

3. MATCH ANALYSIS

The proposed match analysis framework is given in Figure 5. The motivation is to find which team is superior in long-view soccer shots, thus could automatically annotate the shots to facilitate semantic video retrieval or to help to analyze the whole match.

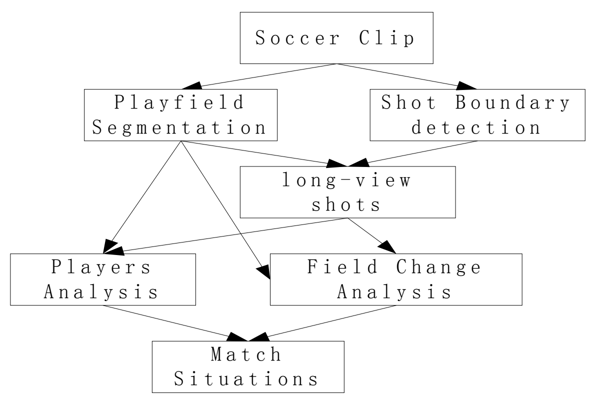


Figure 5: Framework of Match Analysis

Shot boundary detection has been performed in advance, and super-imposed captions have been excluded. Based on the results of playfield segmentation, we could identify long-view shots, which is also considered in [2][6], the feature that we used is dominant color ratios. For each long shot, player analysis and playfield change analysis are performed.

Suppose a video shot *VS* has *n* number of frames*:* $\{f_1, f_2...f_n\}$, for each frame f_i, field segmentation described above is performed, generating playfield regions: R_{fi} and a serious regions inside of the R_{fi}. As the playfield segmentation method could well preserve player regions and eliminate noises, player regions and which team that the players belongs to could be identified based on the color feature of players uniforms. Thus in the procedure of player analysis, number of players existed in f_i could be estimated and represented as TA_{fi} and TB_{fi}, where A and B represent two teams. It should be noted that full occlusions rarely happen in long-view frames. We use global motion estimation to implement field change analysis as playfield regions occupy most of the places in long-view frames. Global motion is calculated by estimating several model parameters [13]. The parameters include motions in three aspects: zooming, rotation and translation. Here we use horizontal translation parameter H_{fi} to model the field change in f_i. Without generality, suppose if H_{fi} is positive, the camera motion is to the mouth goal of team B. Thus match situation of VS is modeled as:

$$MS_{VS} = \sum_{i=1}^{n} (a(TA_{fi} - TB_{fi}) + bH_{fi}) / n$$

Where a and b are two parameters, to make results of player analysis and field change analysis comparable and to make player analysis more important than field change. If MS_{VS} is above a threshold t, then team A is identified as superior to B; else if MS_{VS} is below $-t$, team B is superior; otherwise they are inextricably involved.

Video clip of soccer-2 (a recent match between China and Iran) introduced in section 2.3 is used to experiment on match analysis. In player analysis process, detection precision criterion is described as:

$$dpc = \left(\sum_{i=1}^{m} \left(\frac{\left|TA_{fi} - \left|TA_{fi}^{T} - TA_{fi}\right|\right|}{TA_{fi}} + \frac{\left|TB_{fi} - \left|TB_{fi}^{T} - TB_{fi}\right|\right|}{TB_{fi}}\right)\right) / 2m$$

We select eighteen long-view frames as the test set for player analysis; a 92.2% dpc precision is obtained, which shows that results of player analysis are effective for match analysis. Seven shots are used to evaluate the proposed match analysis method. The situations of these shots are evaluated and have a consensus by several subjects. Reports are given in Table 4, which shows that good analysis results are achieved. Only one shot does not accord with the evaluation by subjects.

Shot-1	Shot-2	Shot-3	Shot-4	Shot-5	Shot-6	Shot-7
C-S	*C-S*	*I-S*	*I-S*	*I-S*	*Equv.*	*I-S*
C-S	*C-S*	*C-S*	*I-S*	*I-S*	*Equv.*	*I-S*

Table 4: Results of match analysis. The second line is the situation of the shot evaluated by subjects; the third line is the estimated situation by our method. *C-S* represent that China is superior; *I-S* represent that Iran is superior and *Equv.* represents that the two teams are equivalent.

4. CONCLUSION

This paper proposes a new method to segment playfield of sports video based on GMMs. The method is robust to multifarious sports videos that playfield occupies dominant regions. The segmentation results could facilitate further analysis of sports video such as player or ball detection, shot-type classification et al. Besides, match analysis of soccer video based on the playfield segmentation is proposed and promising results are obtained. Further researches include constructing more match analysis methods based on basic semantic clues of different sports videos.

5. ACKNOWLEDGMENTS

This work was supported by NEC-JDL sports video analysis and retrieval project; partly supported by China-American Digital Academic Library (CADAL) project.

6. REFERENCES

[1] Y. Gong, T.S. Lim, and H.C. Chua, "Automatic Parsing of TV Soccer Programs", IEEE International Conference on Multimedia Computing and Systems, May, 1995

[2] J. Assfalg, Marco Bertini,Carlo Colombo, Alberto Del Bimbo, Walter Nunziati, "Semantic annotation of soccer videos: automatic highlights identification", Computer Vision and Image Understanding, Special isssue on video retrieval and summarization, Volume 92,Issue 2/3,November/ December 2003

[3] Alan Hanjalic, "Generic approach to highlights extraction from a sports video," IEEE ICIP03

[4] L. Xie, P. Xu, S.-F. Chang, A. Divakaran and H. Sun, "Structure Analysis of Soccer Video with Domain Knowledge and Hidden Markov Models," Pattern Recognition Letters, to appear, 2004

[5] Ling-Yu Duan, Min Xu, Tat-Seng Chua, Qi Tian, Chuang-Sheng Xu, "A mid-level representation framework for semantic sports video analysis", ACM Multimedia 2003

[6] A. Ekin, A. Murat Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," IEEE Trans. on Image Process, Volume: 12, Issue: 7, July 2003

[7] M Luo, Y.F. Ma, H.J. Zhang, "Pyramid wise Structuring for Soccer Highlight Extraction", IEEE PCM 2003

[8] H. Sun, J.H. Lim, Q. Tian, M. Kankanhalli, " Semantic Labeling of Soccer Video", IEEE PCM 2003

[9] A. Ekin, and A. Murat Tekalp, "Robust dominant color region detection and color-based applications for sports video", International Conference on Image Processing 2003

[10] Utsumi, O., Miura, K., Ide, I., Sakai, S., Tanaka, H. "An object detection method for describing soccer games from video", ICME '02, Proceedings International Conference on Multimedia and Expo, Volume: 1 , 26-29 Aug. 2002

[11] Xinguo Yu, Qi Tian, Kong Wah Wan,"A novel ball detection framework for real soccer video", ICME '03, Proceedings International Conference on Multimedia and Expo, Volume: 2 , 6-9 July 2003

[12] Qixiang Ye, Wen Gao, Wei Zeng. "Color Image Segmentation Using Density-Based Clustering", International Conference on Acoustic, Speech and Signal Processing. ICASSP2003.

[13] Srinivasan M., S. Venkatesh, and R. Hosie, "Qualitative estimation of camera motion parameters from video sequences", Pattern Recognition, vol.30, 1997.