Playfield Detection Using Adaptive GMM and Its Application

Yang Liu[1], Shuqiang Jiang[2], Qixiang Ye[2], Wen Gao[1,2], and Qingming Huang[2]

[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[2] Graduate School of Chinese Academy of Sciences, Beijing 100080, China
liuy@vilab.hit.edu.cn, {sqjiang, qxye, wgao, qmhuang}@jdl.ac.cn

ABSTRACT

Playfield detection is a key step in sports video content analysis, since many semantic clues could be inferred from it. In this paper we propose an adaptive GMM based algorithm for playfield detection. Its advantages are twofold. First, it can update model parameters by incremental expectation maximization (IEM) algorithm, which enable the model to adapt to the playfield variation with time; Second, online training is performed, which save buffer for training samples. Then, the playfield detection results are applied in recognizing key zone of current playfield in soccer video, in which a fast algorithm based on playfield contour and least square is proposed. Experimental results show that the proposed algorithms are encouraging.

1. INTRODUCTION

Every day plenty of sports videos are captured, and how to fast access their content is increasingly demanded. Therefore many researchers focus their work on it.

In [1], Gong et al adopts Laplacian operator to extract edges and find lines in the court by subtracting those edges with non-white pixels. Through analyzing relationship among lines, circle and semicircles with domain knowledge, the zone of playfield is recognized. In the same year, Yow [2] proposes a method to synthesize a panorama near the goalmouth area to observe the ball's trajectory and attacker's track. Ekin [3] uses dominant color for shot type classification, since in general playfield occupies dominant color regions in many sports videos.

It could be observed that playfield plays essential role in analyzing many kinds of sports, such as soccer tennis and so on. Further more the result can be used for identifying players [4,5], and ball [6]. In the literatures, researchers usually use histogram to model the playfield [7], [8], [9]. In [8], Xie also uses the dominant color in HSI space to analyze soccer video. In [9], the authors exploit the green color as a convex set in HSI space. In [7], the dominant color detection algorithm is first described in detail. The authors exploit two color spaces, one is used as

control space, and the other is primary space. The dominant color is represented by the mean value of the main region in the histogram. The algorithm calculates the cylindrical metric to determine whether a pixel belongs to the playfield.

Compared with these methods, our algorithm exploits Gaussian mixture model (GMM). As described in our prior work [10], playfield color will vary in the same stadium under different conditions, for example even the same region in a playfield will present different color in the conditions of different whether and different lighting. Even if all the conditions are identical, the color of playfield may be different, for example the color in global view is not consistent with that in close up view. To solve these problems, we propose a novel algorithm that can adapt the color model (GMM) to new conditions, in which the color model's parameters are reestimated online by incremental expectation maximization algorithm [11].

Our algorithm has the following contributions for playfield detection. 1) To more accurately model the playfield, GMM is adopted, which is more sufficiently general to model complex distribution density than histogram. 2) To adapt the model to playfield evolution, incremental EM (IEM) algorithm is employed to update the GMM. 3) Since one-pass training is performed, the algorithm can save buffer for training set.

The rest of this paper is structured as follows. In section 2, we propose the adaptive GMM for field detection in detail. A new coarse court position detection algorithm is discussed in section 3, which is based on the adaptive detection result. Experiments results and analysis are presented in section 4. The last section concludes the paper.

2. ALGORITHM for PLAYFIELD DETECTION

In this section, we propose an adaptive GMM based framework for playfield detection. The framework is shown in figure 1.

2.1. Dominant color region detection

We exploit the main property of sports video, in which the playfield color dominates most shots in sports video.
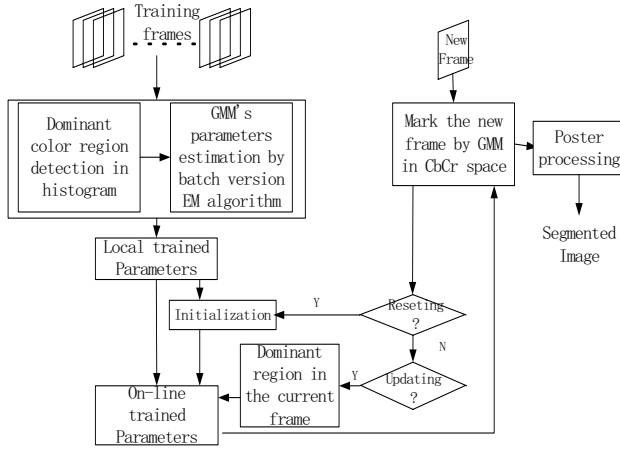
Figure 1. The proposed framework for playfield detection.

As we have observed that when the court region is large enough, the histogram of the frame has a distinct cluster region in CbCr space. The reason to select CbCr space is that it can better describe the court color. In section 4, we will provide a comparison between CbCr and Hue-Intensity space, and the results show that performance in CbCr space is better than that in HSI.

As we have observed that only some small regions (bins) of a histogram have non-zero values, and in general there are some peaks in the two-dimensional histogram. Although most of the main peaks correspond to grass color, exceptions could be found. Thus, we have to determine the main region from histogram, which corresponds to the playfield region in the frame. The procedure is described as follows:

1. Determine the main peak P_1;

2. Find connected region (4-connected region) around the P_1, only the bins with values larger than $T * Value(P_1)$ are considered. Compute the sum of the connected bins, noted as Sum_1, then subtract the connected region, where T is a ratio. In this paper we set it 0.05.

3. Similar to 1 and 2, find the main peak P_2 in the remaining histogram and compute the sum of the connected bins around it, denoted as Sum_2.

4. Return the connected region in the histogram corresponding to the larger of Sum_1 and Sum_2.

2.2. Adapting GMM by incremental EM

In order to model the playfield color, Gaussian mixture model (GMM) is adopted. There are two reasons for us to adopt the model. First, in some stadiums the court colors are not uniform. Second, GMM is a good model for describing probability density with multiple peaks.

● Gaussian mixture model

A GMM G is the mixtures of multiple Gaussian functions $G_1, G_2, \cdots, G_k$, which is described as (1)

$$G = \sum_{i=1}^{k} \pi_i G_i ,$$

where $\quad G_i(X;\theta_i) = \dfrac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp^{-\frac{1}{2}(X-\mu_i)^T (\Sigma_i)^{-1}(X-\mu_i)}$,

and $\quad \sum_{i=1}^{k} \pi_i = 1$

(1).

Each component G_i is a Gaussian function, parameterized by θ_i, which consists of the mean vector μ_i, and the covariance matrix Σ_i. The dimension of sample data X is d. Thus, the set $\{\pi_i, \theta_i\}$ of all unknown parameters belongs to some parameter space. Generally, those parameters are estimated by expectation maximization (EM) algorithm. In the algorithm, we first estimate the model's parameters as the initial settings on the initial accumulated frames.

We need to point out that one match lasts long time. Training data coming from some initial shots can not cover the court well. So, the system should update the GMM continuously to adapt new situations. Because video data is too much, two reasons restrict us from retraining the model from scratch. One is that we have no enough memory space for storing samples. The other is that frequently retraining GMM will cost computational resource. These drive us to use an incremental EM algorithm to continuously update the model.

● Incremental EM algorithm

Radford [11] has shown that the incremental EM algorithm can speed up the convergence of maximum likelihood estimation. For online adaptation, faster speed is one of the critical factors. In the view of [11], E step and M step can maximize the same function, and both the batch and incremental versions of EM can be derived from the function (2).

$$F(\widetilde{P}, \theta) = E_{\widetilde{P}}[\log P(y, z \mid \theta)] + H(\widetilde{P}) \qquad (2),$$

where $\quad H(\widetilde{P}) = -E_{\widetilde{P}}[\log \widetilde{P}(y)]$ is the entropy of the distribution $\widetilde{P}$, $E_{\widetilde{P}}[\bullet]$ is the expectation of $\bullet$ computed under the distribution $\widetilde{P}$.

The difference between the two versions is in the E step. For batch version, an iteration of it can be written in terms of the function F:

E step: Set $\widetilde{P}^{(t)}$ to that $\widetilde{P}$ which maximizes $F(\widetilde{P}, \theta^{(t-1)})$.

M step: Set $\theta^{(t)}$ to that θ which maximizes $F(\widetilde{P}^{(t)}, \theta)$.

In the batch version, the E step version is performed on all data items.

Suppose that $y = (y_1, y_2, \cdots y_n)$ and $z = (z_1, z_2, \cdots z_n)$ are unobserved variables and observations. In general, the given data items are independent, thus the function F can be written in the following form:

$$F(\widetilde{P}, \theta) = \sum_i F_i(\widetilde{P}_i, \theta) \quad .$$

where $F_i(\widetilde{P}_i, \theta) = E_{\widetilde{P}_i}[\log P(y_i, z_i | \theta)] + H(\widetilde{P}_i)$

The incremental EM algorithm adopts the structure. Once one or a small group data items is visited in the E step, the M step will be performed. If the joint distribution function of unobserved variables and observations are in the exponential family, the E step can compute a sufficient statistics instead of computing the distribution $\widetilde{P}$ and the M step computes the maximization likelihood estimation of θ given the statistics. Using $s_i(y_i, z_i)$ to denote the sufficient statistics of i-th data item and its corresponding unobserved variable, the E step and the M step of incremental EM are described as follows (details in [11]):

E step: Select the i-th data item

Set $\widetilde{s}_j^t = \widetilde{s}_j^{t-1}$ for $j \neq i$

Set $\widetilde{s}_i^t = E_{\widetilde{P}_i}[s_i(y_i, z_i)]$, where $\widetilde{P}_i(y_i) = P(y_i | z_i, \theta^{t-1})$.

Set $\widetilde{s}^t = \widetilde{s}^{t-1} - \widetilde{s}_i^{t-1} + \widetilde{s}_i^t$.

M step: Set θ^t to that θ with maximum likelihood given s^t.

Note that, in our adaptive framework, the statistics $\widetilde{s}_i^{t-1}$ is zero, since one pass training is performed. In the proposed framework, when the E step explores each incremental data, the M step is performed. Thus only one pass training is executed.

● Criterion for parameters reestimation

We use the initial model and the current model simultaneously to label the current frame. The ratios of playfield pixels labeled by the two models' parameters are R_i and R_c respectively. When they are both larger than a threshold T_d and $|R_i - R_c| < \varepsilon$, the algorithm begins to detect the main region in the frame's histogram in CbCr space, and those pixels in this region are used as new incremental training data, then the current model is reestimated by incremental EM. When $|R_i - R_c| > T_r (T_r > \varepsilon)$, we reset the current model to the initial model.

3. KEY ZONE DETECTION in SOCCER

In general, highlights in soccer program happen near the penalty area. These areas can be represented by the following patterns, which are illustrated in figure 2.

As we observe, the lines in court are not always clear. Thus, it is difficult to detect the line through edge detection. Even if the pixels of the line are detected,

Hough transformation may not obtain the exact line because of many uncertain factors, including occlusion, noise, etc. To robustly detect the above mentioned patterns, outer layer points of playfield (red points in figure 2) are used. We use a line $y = K \cdot x + B$ to fit those points by least square method and convert the slope K of the line into words, *left, right, middle_left, middle_right,* according to its value. The conversion formula is described as

$$S_q = \begin{cases} left, & K \geq T_{left} \\ middle\_right, & 0 < K < T_{left} \\ middle\_left, & T_{right} < K < 0 \\ right, & K \leq T_{right} \end{cases} ,$$

where T_{left} and T_{right} are two thresholds for detection the left bottom line region and the right bottom line region. In the above formula, the four values correspond to left near bottom line, left near middle area, right near middle area and right near the bottom line. Figure 3 shows some samples extracted from our experiments.

Figure 2. Four patterns. From left to right: left-up, left-bottom, right-bottom and right-up positions in the court.
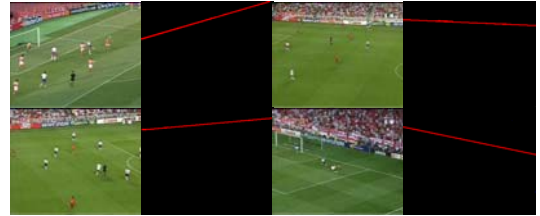

Figure 3. The fitting line pattern estimated by least square method.

4. EXPERIMENTAL RESULTS AND ANALYSIS

To verify the proposed algorithm, we test our system on real video recorded from CCTV-5 and BTV-6. The test set includes more than 10 hours soccer program. To quantify the detection results, we annotate 4 clips manually in pixel-wise fashion, including soccer and basketball.

4.1 Detection results by adaptive GMM

We test the performance of the proposed adaptive GMM algorithm for playfield detection. Since CbCr space has better differentiation ability (see table 1), we only compare the proposed method with unadaptable GMM, which is trained by batch version EM. Figure 4 shows some detection results by traditional GMM and adaptive GMM.

From the detection results, we can conclude that the adaptive GMM algorithm can fit the field color variation better. The overall detection results on more images are shown in table 1.
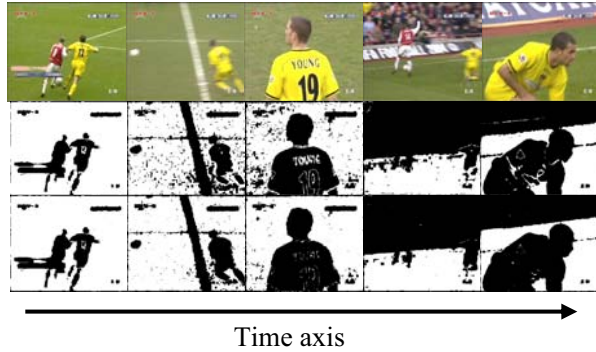

Time axis

Figure 4. First row: original images. Results detected by GMM and results detected by adaptive GMM are in the second and the third row.

Clip	GMM(hue-intensity)	GMM (CbCr)	Adaptive GMM (CbCr)
Soccer 1	90.96%	91.73%	93.67%
Soccer 2	85.38%	90.76%	93.55%
Basketball	84.84%	90.69%	93.48%

Table 1. Comparison of detection results of accuracy.

4.3 Key zone detection

In this section, we apply the adaptive field detection to zone identification in playfield. Since important events are generally described by medium or long view, we first filter out those shots of out of field and close up using the method in [3]. Table 2 shows the results tested on 5 hours video. As it illustrates, the algorithm has attained encouraging performance. The errors are mainly caused by the unfiltered out close-up shot.

Clip	Left penalty area (recall/precision)	Right penalty area (recall/precision)
Port. VS. Eng.	100% / 96.1%	100% / 89.7%
Switz. VS. Fr.	100% / 92.1%	95.2% / 93.5%
Spain VS. Russ.	94.4 % / 90.2%	96.5% / 91.6%

Table 2. Coarse position detection results by field contour and least square.

5. CONCLUSION AND FUTURE WORK

In this paper, we propose a new framework based on adaptive GMM for playfield detection, in which the GMM for playing is trained by incremental EM algorithm. By applying our algorithm, the initialized GMM can be adaptively online adjusted to fit new conditions. It avoids large buffer requirement for training data, since one-pass iteration is performed in incremental EM. To detect the key zone in playfield, the court contour and least square is adopted. Experimental results show that both the proposed algorithms have encouraging performance. In future work, we will add shadow region on playfield as incremental training data to the framework.

REFERENCE

[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, pp. 167 – 174.

[2] D. Yow, B. L. Yeo, M. Yeung, and B. Liu, "Analysis and presentation of soccer highlights from digital video," in ACCV95, 1995.

[3] A. Ekin and A. M. Tekalp, "Shot type classification by dominant color for sports video segmentation and summarization", ICASSP 2003.

[4] J. Assfalg, M. Bertini, C. Colombo, A. Del Bimbo, W. Nunziati, "Semantic Annotation of Soccer Videos: Automatic Highlights Identification", Computer Vision and Image Understanding, 92(2): 285-305.

[5] Utsumi, O., Miura, K., Ide, I., Skai, S., Tannaka, H., "an object detection method for describing soccer games from video", ICME 2002, Volume: 1, 26-29 Aug. 2002.

[6] X. Yu, Q. Tian, and K. W. Wan, "A novel ball detection framework for real soccer video," In Proc. ICME 2003, Vol. II, 265-268.

[7] A. Ekin and A. Murat. Tekalp, "Robust dominant color detection and color-based application for sports video", ICIP 2003.

[8] 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.

[9] M. Luo, Y. F. Ma, H. J. Zhang, "Pyramid wise structuring for soccer highlight extraction," IEEE PCM 2003.

[10] Shuqiang Jiang Qixiang Ye, Wen Gao. "A new method to segment playfield and its applications in match analysis in sports video," Accepted by ACM Multimedia 2004.

[11] Radford M. Neal and Geoffrey E. Hinton, "A new view of the EM algorithm that justifies incremental and other variants," Submitted to Biometrika, February, 1993