AUTOMATIC MULTI-PLAYER DETECTION AND TRACKING IN BROADCAST SPORTS VIDEO USING SUPPORT VECTOR MACHINE AND PARTICLE FILTER

Guangyu Zhu[1], *Changsheng Xu*[2], *Qingming Huang*[3], *Wen Gao*[1,3]

[1]Harbin Institute of Technology, Harbin, P.R. China
[2]Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore
[3]Graduate School of Chinese Academy of Sciences, Beijing, P.R. China
{gyzhu, qmhuang, wgao}@jdl.ac.cn, {xucs}@i2r.a-star.edu.sg

ABSTRACT

In this paper, a novel multiple objects detection and tracking approach based on support vector machine and particle filter is proposed to track players in broadcast sports video. Compared with previous work, the contributions of this paper are focused on three aspects. First, an improved particle filter called SVR particle filter is proposed as the player tracker by integrating support vector regression (SVR) into sequential Monte Carlo framework. SVR particle filter enhances the performance of classical particle filter with small sample set and improves the efficiency of tracking system. Second, support vector classification combined with playfield segmentation is employed to automatically detect the players in sports video as the initialization of tracker. Third, a unified framework for automatic object detection and tracking is proposed based on support vector machine and particle filter. The experimental results are encouraging and demonstrate that our approach is effective.

1. INTRODUCTION

The explosive growth of the amount of broadcast video necessitates indexing and storage of them to facilitate efficient querying and browsing. In order to enable effective indexing, it is important to understand the semantics of the video. In this paper, we focus on broadcast sports video which is a kind of popular television programs and needs to understand the semantics of the content to handle various queries.

Semantic understanding of sports video can be considered as understanding the process and strategy of the game. The position and trajectory of players in the game can provide useful information to facilitate such semantic analysis. In order to achieve such performance from the video, detection and tracking of objects (players, ball, and lines) is required. The detection and tracking reveals the movement of the players on the field, which can be used to detect certain events or understand the overall trend and strategy of the game. Although detection and tracking of objects have been a poplar topic in computer vision and image processing, most of existing approaches assume special conditions such as fixed or multiple cameras, single moving object, and relative static background. In broadcast sports video, these strict conditions are of relaxation. First, the cameras used to capture sports game are not fixed and they are always moving to follow the players. Second, the broadcast video is an edited video where the broadcast sequence feed is selected from frequent switches among multiple cameras according to broadcast director's instruction. Third, there are numerous players moving in various directions in broadcast video. Finally, the background in broadcast sports video changes sharply. Thus detection and tracking of players in broadcast sports video remains difficult. There have been many efforts devoted to player tracking in sports video such as [1, 2, 3, 4, 5]. Some approaches track players using single camera [1, 2, 3] while others [4, 5] attempted to track players from the video fed by multiple cameras.

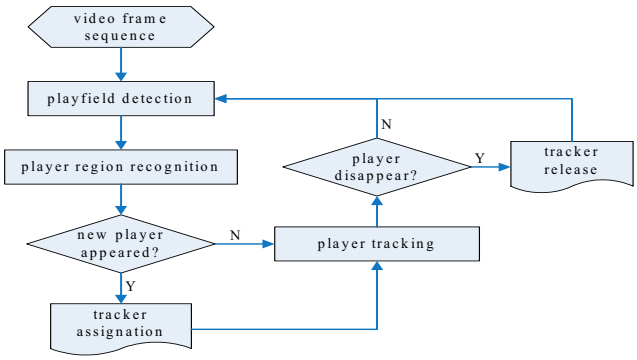
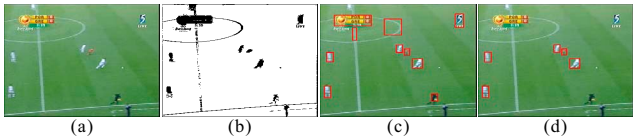


Fig. 1 Framework of multi-player detection and tracking

In this paper, we propose a novel multi-player detection and tracking approach based on support vector machine and particle filter. Fig. 1 shows the framework of our approach. For each video frame, playfield detection is first performed. The regions inside the extracted field are considered as the player candidates. Then recognition module is employed to eliminate the non-player candidates. For each of selected player regions, if it is identified as a new appeared player, a tracker is assigned. Filtering based trackers keep tracking players in the frames. After each tracking iteration, the player disappearance module evaluates whether the current



(a) (b) (c) (d)

Fig. 2 Player detection results. (a) original frame, (b) after playfield detection, (c) after post-processing, (d) after support vector classification

tracked players leave the scene. If it does, the corresponding tracker is released.

2. SUPPORT VECTOR CLASSIFICATION BASED PLAYER DETECTION

In player detection task, we assume that the playfield pixels are the dominant components in most of the video frames and players are always located within the area of playfield. Since player region in video frame is small and very easy to be blurred due to the camera motion, this may result in the failure of player segmentation. Here we adopt a converse coarse-to-fine method that segments playfield first to obtain the player candidate regions. Then player model based on support vector classification (SVC) is employed to finely recognize the final player regions.

2.1. Playfield detection

The playfield color model is constructed using Gaussian mixture models (GMMs) [6]. GMMs have powerful ability to form smooth approximations and arbitrarily shaped densities.

Color distributions of playfield pixels are modeled in the 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. The conditional density of a pixel ϕ belonging to playfield region ψ is modeled with a convex combination of M Gaussian density components as illustrated by Eq. (1).

$$P(\phi \mid \psi) = \sum_{i=1}^{M} w_i g_i(\phi) \tag{1}$$

where w_i is the mixture weight, and $\sum_{i=1}^{M} w_i = 1$; $g_i(\bullet)$ is the mixture component, $i = 1,\ldots,M$. Each component density is a Gaussian with mean vector μ_i and covariance matrix Σ_i:

$$g_i(\phi) = \frac{1}{(2\pi)^{d/2} \mid \Sigma_i \mid^{1/2}} \exp\left\{-\frac{1}{2}(\phi - \mu_i)^T \Sigma_i^{-1} (\phi - \mu_i)\right\} \tag{2}$$

where d is the dimension of sample data. Fig. 2(b) shows a result of playfield detection. GMMs with 4 components are utilized in our experiments.

After playfield pixel detection, we employ the region growing algorithm [7] to perform the post-processing so as to connect playfield pixels into regions, eliminate noises, and smooth the boundaries. Fig. 2(c) shows the corresponding experimental result. We can see that the player regions are extracted successfully. But some false regions, such as

Fig. 3 Training samples of SVC for player detection

super-imposed caption, logo and line mark in the field, are also segmented.

2.2 Player region recognition

Usually in tracking applications, we are quite often interested in tracking a particular class of objects such as people or vehicles. In this case, we can train a classifier in advance to distinguish between the interested and uninterested object.

Support vector machine (SVM) [8] is used in our approach to train the classifier. SVM has been successfully applied to a wide range of pattern recognition and classification problems. The advantages of SVM over other methods consist of: 1) providing better prediction on unseen test data, 2) providing a unique optimal solution for training problem, and 3) containing fewer parameters compared with other methods. Another important motivation of using SVM is that support vector regression (SVR) is applied in our particle filter tracking framework. Consequently, a unified detection and tracking framework is constructed based on support vector machine and particle filter. Some training samples for support vector classification (SVC) are illustrated in Fig. 3. The collected samples are manually generated from the original frames and false segments of playfield detection.

The feature fed into SVC utilizes the color model proposed in [9] which is obtained by histogramming technique in the Hue-Saturation-Value (HSV) color space. We first define $b(\mathbf{u}) \in \{1,\ldots,N\}$ as the bin index of histogram associated with the color vector $c(\mathbf{u})$ at pixel location $\mathbf{u}$ in each sample image. Given sample s_i in training set, $i = 1,\ldots,S$ where S is the total number of samples, a kernel density estimate $\mathbf{q}(s_i) = \{q(n;s_i)\}_{n=1}^{N}$ of the color distribution for s_i is given by Eq. (3).

$$q(n;s_i) = C \sum_{\mathbf{u} \in s_i} \eta[b(\mathbf{u}) - n] \tag{3}$$

where η is the Kronecker delta function, C is the normalization constant ensuring $\sum_{n=1}^{N} q(n;s_i) = 1$. This model associates a probability to each of the N color bins.

Fig. 2(d) shows the final result after support vector classification. It is obvious that the false segments after playfield detection have been eliminated. The final segmented players are identified as new appearance if there are no associated trackers. Then the new trackers are assigned and initialized for the succedent tracking iterations.

3. PARTICLE FILTER BASED PLAYER TRACKING WITH SUPPORT VECTOR REGRESSION

Particle filter utilizes sequential Monte Carlo method for online inference within Bayesian framework and has been successfully applied for visual tracking [9, 10] due to its ability to carry multiple hypotheses and relaxation of linearity/Gaussian assumption. For classical particle filter, large numbers of samples are needed in practice to properly approximate density over time steps. However, the size of sample set is directly related to the computational cost and should be kept as small as possible. We adopt a simple but effective sample re-weighting scheme based on support vector regression (SVR) [8] to tackle this problem which results in the proposal of SVR particle filter [11].

The posterior density of particle filter $p(\mathbf{x}_t \mid \mathbf{z}_{1:t})$ at time step t is given by a weighted sample set $\{(\mathbf{x}_t^i, w_t^i)\}_{i=1}^N$ in a discrete form:

$$p(\mathbf{x}_t \mid \mathbf{z}_{1:t}) \approx \sum_{i=1}^{N} w_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i) \tag{4}$$

where $\delta(\bullet)$ is the Dirac delta function, $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \ldots, \mathbf{z}_t\}$ is the history of observation up to time step t. Considering Eq. (4), samples need to be weighted more accurately if we reduce the size of sample set. In classical particle filter, it is separately that the weights are evaluated by likelihood function $p(\mathbf{z}_t \mid \mathbf{x}_t)$. However, the robustness of $p(\mathbf{z}_t \mid \mathbf{x}_t)$ in practice is influenced significantly by the noise of circumstance such as background clutter, deformation of non-rigid object in visual tracking. Therefore, a few weights will be corrupted which mislead the corresponding false samples close to the true state. The small set with false samples will result in a noisy posterior distribution leading to significant sample impoverishment consequently.

SVR particle filter introduces support vector regression into sequential Monte Carlo which re-weights samples accurately by smoothing the distribution via regression function to eliminate the noise in posterior. With the context of particle filter, the posterior density is usually nonlinear/non-Gaussian. We cannot specify its parameterized representation in advance. Support vector regression is a nonparametric technique. Therefore it is more suitable to tackle our problem than other existing parametric regression methods. Moreover, SVR is robust to noisy distribution and can work effectively even with small sample set.

Considering the weighted sample set $\{(\mathbf{x}_t^i, w_t^i)\}_{i=1}^N$ and Eq. (4), $\mathbf{x}_t^i$ can be treated as a sample point drawn from the current posterior distribution and w_t^i as the corresponding density value. Intuitively, we can construct the regression function f_t over $\{(\mathbf{x}_t^i, w_t^i)\}_{i=1}^N$ using SVR technique [8]:

$$f_t(\mathbf{x}_t) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}_t, \mathbf{x}_t^i) + b_t \tag{5}$$

where $K(\bullet)$ is the kernel function adopted by SVR. Afterwards, each sample is re-weighted using Eq. (5) to generate new weighted set $\{(\mathbf{x}_t^i, \tilde{w}_t^i)\}_{i=1}^N$ where each weight is refined. Essentially, SVR particle filter utilizes the regression function obtained by Eq. (5) to transform samples from a discrete space into a continuous space. Kernel function in regression is used for smoothing so that each sample is re-weighted using the information of its neighboring samples rather than separately in classical particle filter. Since re-weighting is implemented at every time step, the efficiency and diversity of sample set is maintained and sample impoverishment is avoided. More detail about SVR particle filter for player tracking can be found in [11].

After each tracking iteration, player models generated by support vector classification from player detection module are employed to check each tracked region. If the region is classified as non-player for L consecutive frames, we treat the player to be disappeared or lost from tracker. Then the corresponding tracker is released in order to perform next detection and tracking iteration.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance and scalability of our approach, the test clips in our experiment are extracted from 4 videos recorded from live broadcast television including the games of soccer, hockey, American football, and basketball. The videos are compressed in MPEG-2 standard with the frame rate of 25fps and the resolution of 352×288. 13 clips for consecutive far view scene in the matches are derived from the whole videos which the total number of frames is 3599.

The detailed experimental statistics are listed in Table 1 where SOPs means Sum Of Players appeared in all the frames, e.g. three players in the first frame and four players in the second frame, then SOPs is seven. As shown in Table 1, the average accuracy of detection and tracking for 13 clips is 90.5%. This demonstrates that our multi-player detection and tracking approach is able to achieve good performance.

Due to the limitation of space, Fig. 4 just illustrates the result of one clip. We can see that our approach is able to detect all the players in the scene including the situation of player coming in (at frame 30 for goalkeeper and frame 120 for white player) and out of frame (at frame 90 for white player). The players were tracked successfully as long as they were detected. This demonstrates SVR particle filter is effective for player tracking. However, in some situation, the player cannot be detected and tracked. This is because 1) the player region is so small that it is falsely treated as noise in the background, 2) the player region in the frame is very close to the caption, logo, or line mark in the field that may lead to its merger with adjacent region when performing post-processing after playfield detection.

5. CONCLUSION

In this paper, we propose a novel multiple objects detection and tracking approach to track the players in broadcast

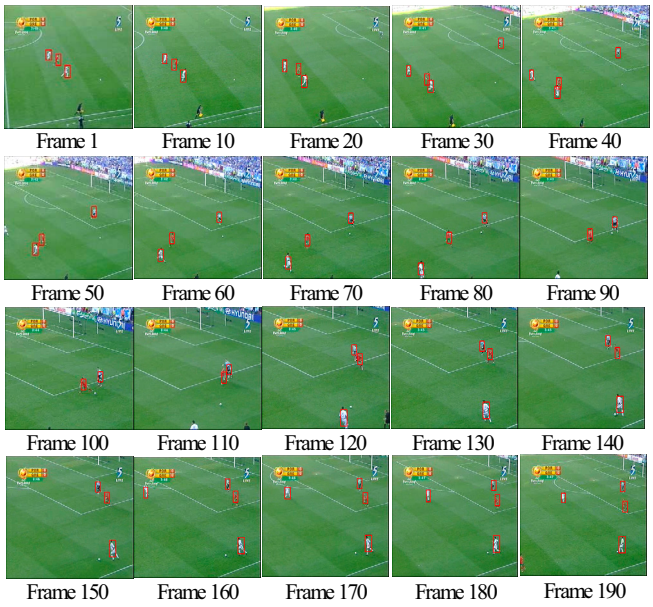


Fig. 4 Experimental results for multiple players detection and tracking

sports video. SVR particle filter by integrating support vector regression with classical particle filter is utilized as player tracker. Automatic detection using support vector classification combined with playfield segmentation is employed to extract the player region from the video frame as the initialization of tracker. As a consequence, a unified framework for automatic object detection and tracking based on support vector machine and particle filter is constructed. This framework can be widely used to solve other problems for multiple objects detection and tracking.

Due to many crowded scenes in the broadcast sports video, our future work will focus on dealing with occlusion issue for player detection and tracking.

6. REFERENCES

[1] C.J. Needham, R.D. Boyle, "Tracking multiple sports players through occlusion, congestion and scale," British Machine Vision Conference, vol. 1, pp. 93-102, 2001.

[2] O. Utsumi, K. Miura, I. Ide, S. Sakai, H. Tanaka, "An object detection method for describing soccer games from video," IEEE International Conference on Multimedia and Expo, vol. 1, pp. 45-48, 2002.

[3] J.I. Agbinya, D. Dees, "Multi-object tracking in video," Journal of Real-Time Image, vol. 5, pp. 295-304, 1999.

[4] P. Figueroa, N. Leite, R.M.L. Barros, I. Cohen, G. Medioni, "Tracking soccer players using the graph representation," IEEE International Conference on Pattern Recognition, vol. 4, pp. 787-789, 2004.

[5] S. Iwase, H. Saito, "Parallel tracking of all soccer players by integrating detected positions in multiple view images," IEEE International Conference on Pattern Recognition, vol. 4, pp. 751-754, 2004.

[6] S. Jiang, Q. Ye, W. Gao, T. Huang, "A new method to segment playfield and its applications in match analysis in sports video," ACM Multimedia, pp. 292-295, 2004.

[7] Q. Ye, W. Gao, W. Zeng, "Color image segmentation using density-based clustering," IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 345-348, 2003.

[8] V. Vapnik, The nature of statistical learning theory, Springer-Verlag, New York, 1995.

[9] P. Perez, C. Hue, J. Vermaak, M. Gangnet, "Color-based probabilistic tracking," European Conference on Computer Vision, pp. 661-675, 2002.

[10] M. Isard, A. Blake, "CONDENSATION—Conditional density propagation for visual tracking," International Journal of Computer Vision, vol. 29, no. 1, pp. 5-28, 1998.

[11] G. Zhu, D. Liang, Y. Liu, Q. Huang, W. Gao, "Improving particle filter with support vector regression for efficient visual tracking," IEEE International Conference on Image Processing, vol. 2, pp. 422-425, 2005.

Table 1 Multi-player detection and tracking results

video clip	# frame	# SOPs	# detected & tracked	accuracy
soccer-chunk-1	399	1232	1146	93.0%
soccer-chunk-2	271	795	713	89.7%
soccer-chunk-3	315	1241	1053	84.9%
soccer-chunk-4	230	763	694	91.0%
soccer-chunk-5	190	644	619	96.1%
hockey-chunk-1	419	1174	1098	93.5%
hockey-chunk-2	399	1237	1180	95.4%
hockey-chunk-3	327	1111	998	89.8%
hockey-chunk-4	176	688	673	97.8%
football-chunk-1	269	861	760	88.3%
football-chunk-2	301	813	702	86.3%
basketball-chunk-1	165	594	481	81.0%
basketball-chunk-2	138	287	238	82.9%
Total	3599	11440	10355	90.5%