Demo: Spatio-temporal template matching for ball detection

Amit Kumar K.C., Pascaline Parisot and Christophe De Vleeschouwer
ICTEAM Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
amit.kc@uclouvain.be

***Abstract*—This paper considers the detection of ball in a basketball game covered by multiple loosely synchronized cameras. First, plausible ball candidates are detected on the nodes of a 3D grid defined around the basket. This is done by correlating independently in each view the spatial template of the ball with a precomputed foreground mask. Efficient implementation of this step relies on integral images. Afterwards, false positives are filtered out based on a temporal analysis of the ball trajectory. This analysis builds on the Random Sample Consensus (RANSAC) method, with a ballistic trajectory model. The integrated approach is demonstrated on a real-life dataset, and appears to be both effective and efficient.**

***Index Terms*—Multiview, template-matching, RANSAC.**

I. Introduction and overview

Detection of objects is a major component of video analysis systems. In case of sport events, the detection of the players and the ball is essential to interpret the actions in the game. Lots of efforts have been devoted to soccer [1] or tennis games [2]. This paper considers a basketball game. Specifically, the detection of the ball near the basket is useful for interpreting two main events of the game, namely *throw-attempt* and *rebound*.

The proposed algorithm uses only foreground mask. A foreground mask is a binary mask that distinguishes the moving objects from the stationary ones. Assuming that the color of the ball is known a priori, the foreground mask is adapted to highlight foreground pixels that share the color of the ball. The detection module investigates exhaustively a set of candidate points inside a 3D volume. Then, it chooses the point for which the sum of correlation between the 3D ball silhouette templates, projected in the multiple views, and the corresponding foreground mask is maximum and exceeds a given threshold. This method is inherently exhaustive, and thus computationally heavy. However, it can be accelerated by using integral images [3]. The detection rate is further boosted by analyzing temporal consistency of the ball candidates. Most often, the false positives do not have a coherent trajectory whereas the ball follows a ballistic model. Thus, the false positives can be pruned out by using RANSAC method [4]. Results show that the proposed technique yields true detection rate of about 80% for a low false detection rate (around 4%), despite adverse real-life acquisition conditions (poor contrast, loose synchronization between cameras etc.).

The rest of the paper is organized as follows: Section II presents the ball detection technique using spatial template matching. Trajectory analysis of detections is discussed in Section III. Finally, results are presented in Section IV.

II. Ball detection with spatial template matching

Given a calibrated camera $i \in [1, K]$, where K is the number of cameras, and its projection function $f^{(i)} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$, a 3D candidate position $\mathbf{x} \in \mathbb{R}^3$ (that corresponds to the centre of the ball) can be projected on the image plane to obtain a projected point $\mathbf{x}_p \in \mathbb{R}^2$. Then, the projected silhouette of the ball, located in $\mathbf{x}$, would correspond to a disk with center $\mathbf{x}_p$ and radius r, where r can be estimated based on $\mathbf{x}$ and the known ball size. Hence, one can simply verify the presence of a ball in $\mathbf{x}$ by correlating the foreground mask with the spatial template $T^{(i)}$ which is a white disk, with center $\mathbf{x}_p$ and radius r, surrounded by a black annulus of radius $2r$.

To compute the correlation of the template $T^{(i)}$ with the foreground mask, the disk is approximated by a square (such that the square circumscribes the disk), which allows the use of integral images. Formally, let the area of the inner square be $A_1^{(i)}$ and that of the outer square be $A_2^{(i)}$. Moreover, let $S_1^{(i)}$ and $S_2^{(i)}$ be the number of white pixels in the foreground mask with respect to the inner and the outer squares, respectively. $S_1^{(i)}$ and $S_2^{(i)}$ can be computed efficiently by using standard *four-point formula of integral images* [3]. Subsequently, the correlation metric for i^{th} camera, $C^{(i)}$, and the corresponding normalizing factor, $N^{(i)}$, are defined by:

$$C^{(i)} = S_1^{(i)} + \alpha \left[\left(A_2^{(i)} - A_1^{(i)} \right) - \left(S_2^{(i)} - S_1^{(i)} \right) \right] \tag{1}$$

$$N^{(i)} = A_1^{(i)} + \alpha \left(A_2^{(i)} - A_1^{(i)} \right) \tag{2}$$

where $0 \leq \alpha \leq 1$ allows us to give different weights to the inner and the outer square.

Then, the correlation at $\mathbf{x}$, denoted as $C(\mathbf{x})$, can be obtained as:

$$C(\mathbf{x}) = \frac{\sum_i w^{(i)} \left(\frac{C^{(i)}}{N^{(i)}} \right)}{\sum_i w^{(i)}} \tag{3}$$

where the summation is done over the cameras for which $\mathbf{x}_p$ lies inside the field of view. The weight $w^{(i)}$ reflects the reliability of $C^{(i)}$ and thus depends on the level of noise affecting the foreground mask in view i. Here, anything that does not correspond to the ball silhouette is considered as noise. When $\mathbf{x}$ is close to the camera, $C^{(i)}$ is more likely to be affected by players' silhouette because the rays, connecting

978-1-4577-1707-9/11/$26.00 ©2011 IEEE

the camera to $\mathbf{x}$, end up inside the game field area. In contrast, when $\mathbf{x}$ is far from camera, the rays end up in the public area where the foreground mask is reliable (no interference due to players). So, a lower weight is assigned to the camera for which $\mathbf{x}$ is closer (and hence $N^{(i)}$ is larger) as:

$$w^{(i)} = \left(N^{(i)}\right)^{\beta} \tag{4}$$

where $\beta < 0$. Our experiments have revealed that $\beta = -0.3$ provides the best performance in our acquisition settings.

Having computed the correlation for all points of the grid, the one having the largest value (and exceeding some threshold, τ_{corr}) is selected as the most plausible ball position.

III. Temporal Filtering with RANSAC

To remove false positive detection, temporal analysis is performed. Generally, the sources of false positives can be accredited to players' heads, hands, etc. Fortunately, these false detections do not follow the typical ballistic trajectory of the ball in the air. The RANSAC approach [4] is chosen to remove outliers among the detected ball candidate positions and is applied on the following ballistic trajectory model:

$$\mathbf{b}(t) = \frac{1}{2}\mathbf{g}t^2 + \mathbf{v}t + \mathbf{p} \tag{5}$$

where $\mathbf{g}$ is the gravitational acceleration ($\mathbf{g} = [0\;0\;-g]^T$; $g = 9.81m/s^2$), t is the time and $\mathbf{v}$ and $\mathbf{p}$ ($\in \mathbb{R}^3$) are the six unknowns/parameters to estimate.

This process is iterative. It is applied on a window of one second with half-second overlap. On each period, the parameters are initially estimated from a set of three candidates taken at random. To be less sensitive to noisy detections, any two candidates randomly selected should not be from consecutive timestamps. Then, the data that fit this estimated model (with a certain threshold τ_{fit}) are considered as hypothetical inliers; the model is considered and is re-estimated from these inliers if their number is greater than a threshold $\tau_{inliers}$. Finally, the mean squared error between the estimated model and the inliers is computed. This procedure is repeated for a determined number of times and the best estimated model in term of error is kept.

IV. Results

Results are provided with respect to APIDIS dataset [5]. The basketball game has been captured by K (=7) loosely synchronized cameras. Detailed results are provide in [6]. The parameters used in the algorithm are:

- The size of 3D grid is 2 m $\times$4 m$\times$2 m around the basket, and the grid step is 10 cm.
- α = 1/3 and β = -0.3
- Correlation threshold, $\tau_{corr} = 0.53$.
- Maximum error of inliers, τ_{fit} = 15 cm.
- Minimum size of consensus set, $\tau_{inliers}$ = 8.

A significant improvement in the detection rate can be observed in Figure 1 due to temporal filtering. Figure 2 visually depicts the performance of the detector. As depicted

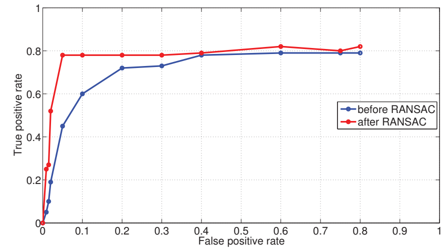
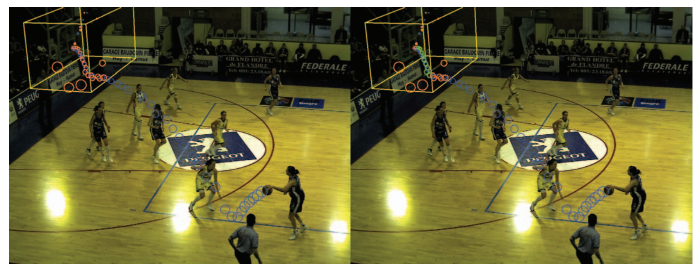


Fig. 1: Performance comparison of the detector before and after temporal analysis

in Figure 2b, false positives (which generally do not follow ballistic trajectory) are discarded by temporal filtering.



(a) (b)

Fig. 2: (a) Ball candidates after spatial template matching (in orange) and ground truth (in blue), (b) Ball candidates classified as ballistic (in green) and non-ballistic (in orange) by temporal analysis.

V. Conclusion

Using a correlation metric to detect the ball around a predefined area of interest appears to be relevant. Even though the search is exhaustive, the computation of correlations can be accelerated by using integral images. Temporal filtering drastically reduces the false alarm rate. The proposed algorithm achieves a true detection rate of around 80% at a low false detection rate (around 4%).

Acknowledgment

The work has been funded by the Walloon Region WIST3 SPORTIC project, and by the Belgian NSF.

References

[1] Y. Huang, J. Llach, and S. Bhagavathy, "*Players and ball detection in soccer videos based on color segmentation and shape analysis*", in Proc. of the Intl. Conf. on Multimedia Content Analysis and Mining. 2007, pp. 416-425, Springer-Verlag.

[2] F. Yan, W. Christmas, and J. Kittler, "*A tennis ball tracking algorithm for automatic annotation of tennis match*", BMVC, pp. 619-628, 2005.

[3] P. Viola and M. Jones, *Robust real time object detection*, Second Intl. Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing and Sampling, 2001.

[4] M. A. Fischler and R. C. Bolles, *"Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography"*, Comm. of the ACM 24: 381-395, 1981.

[5] http://www.apidis.org/Dataset/

[6] http://thetis.tele.ucl.ac.be/Apidis/amit/results/results/Results.html