

A Method of Small Object Detection and Tracking Based on Particle Filters

Yu Huang*, Joan Llach*, Chao Zhang**

**Thomson Corporate Research, Princeton, NJ08540, US*

***State Key Lab of Machine Perception, Peking University, Beijing 100871, China*

E-mails: yu.huang07@gmail.com, joan.llach@thomson.net, chzhang@cis.pku.edu.cn

Abstract

In this paper an efficient method of small object localization is proposed that integrates detection and tracking. The system is initialized using a strong detector and then it locates the object over time using a weak detector and a temporal tracker. Both of strong and weak detectors are based on foreground-background segmentation. The strong detector is created from shape analysis of foreground blobs and used to trigger the object tracker. The weak detector is built with outputs from the foreground detection likelihood and integrated into the tracker's observation likelihood. In the particle filter-based object tracker, motion estimation is embedded to generate a better proposal distribution and a mixture model is tailored to handle the ambiguity of template matching due to cluttered background. As a case study, the proposed scheme is applied to ball detection and tracking in soccer game videos. Promising results are presented to illustrate the proposed method effectively handles heavy clutter, occlusion and motion blur.

1. Introduction

Object detection and tracking in digital videos provide important information about the object locations and temporal correspondence over the time. Lying at two extremes, traditional tracking utilizes every assumption of temporal continuity, while usual detection aims at discrimination of the target from the background. When the object is small in appearance, its continuity turns to be weak in tracking since cluttered background and occlusion result in severe ambiguity; meanwhile reliable detection is often unaffordable due to deficient features extracted from the object's small region in the image. Integration of detection and tracking offers the capability to overcome this difficulty, achieving strong discriminative power while maintaining the use of weak spatial-temporal continuity.

It is seen that incorporating classifiers into the tracker has become a popular approach [4, 9, 2]. The main idea is to formulate the task as a segmentation/classification problem in the sense of distinguishing the object from the background.

Statistically-based object detectors, such as face detection using boosting [13] and human detection using support vector machine [5], provide relatively good performance. However, they are not appropriate for small objects. Instead, the segmentation-based method [14], using color, depth or motion cues, is a good choice in these situations.

Particle filters provide a convenient Bayesian filtering framework of integrating the detector into the tracker. There are two basic schemes: applying a mixture proposal distribution by combining the dynamic model with the output of the detector [11]; or sending the output of the detector into the measurement likelihood [3].

In this paper the problem of detecting and tracking a small object is addressed. The key idea is to clarify the different roles of detectors in the integration: The detector triggering the tracker should be strong to verify the object's presence at a very low false positive detection rate, whereas the detector integrated into the tracker should be weak with a very low false negative detection rate and allow statistical fusion with temporal inference; Both detectors are based on foreground/background segmentation; The strong detector is created from shape analysis of foreground blobs and the weak detector is built with outputs from the foreground detection likelihood; In the proposed particle filter-based tracker, motion estimation is embedded to generate a better proposal distribution and a mixture model is tailored to handle the ambiguity of template matching due to cluttered background.

2. Segmentation-based Detection

Various segmentation-based detection algorithms [14] have been proposed based on background subtraction, frame differencing, motion estimation and color segmentation. As a case study, ball detection in soccer game video is analyzed in this section and color segmentation is applied to build a blob-level strong detector and a pixel-level weak detector.

2.1. Field segmentation

Soccer is normally played on a grass field; therefore a useful first step is to detect the pixels that form the playfield. In this paper, a color histogram learning technique is employed to detect the playfield pixels [6].

A playfield pixel classifier is derived through the standard likelihood ratio approach as

$$G(x, y) = \begin{cases} 1, & \text{if } like_ratio(x, y) \geq \theta \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where the likelihood ratio for pixel (x, y) is

$$like_ratio(x, y) = \frac{P(rgb|playfield)}{P(rgb|nonplayfield)},$$

with $P(rgb|playfield)$, $P(rgb|nonplayfield)$ converted from playfield and non-playfield histograms, and $\theta \geq 0$ is a threshold optimized from the receiver operating characteristic (ROC) curve [6].

Ideally, the non-playfield pixels inside the extracted playfield areas should be the foreground pixels that can be grouped into different foreground blobs by connected component analysis (CCA).

2.2. Strong detector

This kind of detector functions by finding the isolated object (ball) among the foreground blobs in the playfield. The technique uses shape descriptors, such as perimeter P , area A , major/minor axes C_L/C_S , roundness $F = P^2/(4\pi A)$ and eccentricity $E = C_L/C_S$.

Besides, the ball is nearly white. A simple method to identify white pixels in each foreground blob is [7]

$$X(x, y) = \begin{cases} 1, & \text{if } rb(x, y) < a \text{ AND } I(x, y) > b \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

with

$$rb(x, y) = [r(x, y) - 1/3]^2 + [b(x, y) - 1/3]^2,$$

$r(x, y)$, $b(x, y)$ being normalized red and blue components for pixel (x, y) , $I(x, y)$ denoting the intensity value in the range $[0, 255]$ and $X(x, y)$ denoting the white pixel mask (For all the experiments in this paper, $a=0.6$, $b=120$). Consequently, the proportion of detected white pixels for each blob is $p_w = C\{(x, y) | X(x, y) = 1\} / A$, where $C\{\cdot\}$ is the cardinal of $\{\cdot\}$.

It can be shown that there is a predefined range for the ball's blob area A according to the camera configuration. Roundness F and eccentricity E for a blob candidate should be close to 1.0, different from disconnected segments of field lines or noise. So, the output of the ball detector is

$$B = \begin{cases} 1, & p_w > r_w \text{ AND } A \in [ar_{\min}, ar_{\max}] \text{ AND} \\ & F < f_{\max} \text{ AND } E < e_{\max}, \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

(Note that the parameters are set as $ar_{\min} = 3$, $ar_{\max} = 80$, $f_{\max} = 2$, $e_{\max} = 3$, $r_w = 0.5$ in this paper.)

2.3. Weak detector

This type of detector is defined to output a likelihood score for each search window W in the image as

$$L_{\text{det}}(W) = \sum_{(x, y) \in W} [1 - G(x, y)] N(d_l, \sigma_l) N(d_w, \sigma_w), \quad (4)$$

where $d_l = like_ratio(x, y) / \theta$, $d_w = rb(x, y) / a$, and $N(\cdot, \cdot)$ denotes the normal function.

3. Adaptive Particle Filter for Tracking

3.1. State space and dynamics

The object's state vector is defined as $X = (x, y)$, where (x, y) is the window center of the object. The window size is assumed to be constant during tracking.

In this paper, motion estimation is embedded into the dynamic model as in [10]. If the estimated motion for the object is denoted as V_t , the dynamics can be formulated as

$$X_{t+1} = X_t + V_t + \mu_t, \quad (5)$$

with μ_t denoting the state prediction error. To vary the diversity of particles, variance of μ_t is adapted [15] during a range as $R_t \in [R_{\min}, R_{\max}]$, proportional to the motion estimation error, i.e. $|\mu I_x + \nu I_y + I_t|$, where I_x, I_y, I_t is partial derivatives of the intensity function I with respect to x, y and t .

3.2. Observation likelihood

It is assumed the intensity measurement Z_t^{int} , the motion measurement Z_t^{mot} and the detector measurement Z_t^{det} are independent. Then it results in the following likelihood:

$$P(Z_t | X_t) = P(Z_t^{\text{int}} | X_t) P(Z_t^{\text{det}} | X_t) P(Z_t^{\text{mot}} | X_t)^{O_t-1}, \quad (6)$$

where $Z_t = \{Z_t^{\text{int}}, Z_t^{\text{mot}}, Z_t^{\text{det}}\}$, $O_t = 0$ if the object is occluded, and 1 otherwise. When the object is occluded or motion estimation fails, the motion continuity term is not feasible in the likelihood equation (8). The likelihood term from the detector measurement is given in (4), i.e. $P(Z_t^{\text{det}} | X_t) = L_{\text{det}}(W)$. Other two likelihood components are discussed below.

The intensity measurement is computed based on the correlation surface [1] which better measure the uncertainty in cluttered background. The SSD-based correlation surface for each particle is defined around the SSD peak in a small neighborhood $Neib$ as

$$r(X_t) = \sum_{\chi \in W} [T(\chi) - I(\chi + X_t)]^2, \quad X_t \in Neib. \quad (7)$$

where W is the object window, T is the object template and $I(\cdot)$ is the image at current time.

In this correlation surface, it is assumed J candidates. As a result, $J+1$ hypothesis can be defined as [12]:

$$H_0 = \{c_j = C : j = 1, \dots, J\},$$

$$H_j = \{c_j = T, c_i = C : i = 1, \dots, J, i \neq j\}, j = 1, \dots, J,$$

where $c_j = T$ means the j th candidate is associated with the true match, $c_j = C$ otherwise. Hypothesis H_0 means that

none of the candidates is associated with the true match. The clutter is assumed to be uniformly distributed as $U(\cdot)$, and hence the true match-oriented measurement is Gaussian distributed as $N(\cdot, \cdot)$. Consequently, the intensity likelihood term is formulated as

$$P(Z_t^{\text{int}} | X_t) = q_0 U(\cdot) + C_N \sum_{j=1}^J q_j N(r_t, \sigma_t), \quad (8)$$

where C_N is the normalization factor, and q_j is the prior probability for hypothesis H_j , $j=0, \dots, J$ ($q_0=0.5$, $q_j=(1-q_0)/J$ for all the examples in this paper).

The motion likelihood term is calculated based on the difference between the particle's speed (position change) and the average object speed in recent past, i.e.

$$d_{\text{mod}}^2 = (|\Delta x_t| - \overline{\Delta x})^2 + (|\Delta y_t| - \overline{\Delta y})^2, \quad t > 2$$

where $(\Delta x_t, \Delta y_t)$ is the particle's speed with respect to (x_{t-1}, y_{t-1}) , and $(\overline{\Delta x}, \overline{\Delta y})$ is the average object speed in recent past, i.e.

$$\overline{\Delta x} = \sum_{s=t-k}^{t-1} |x_s - x_{s-1}| / k, \quad \overline{\Delta y} = \sum_{s=t-k}^{t-1} |y_s - y_{s-1}| / k.$$

Hence the motion likelihood is calculated as

$$P(Z_t^{\text{mot}} | X_t) = N(d_{\text{mot}}, \sigma_{\text{mot}}). \quad (9)$$

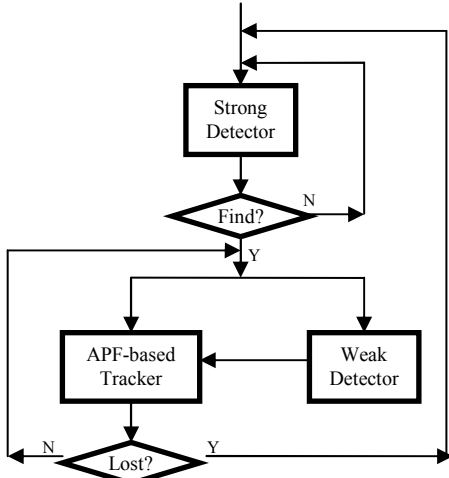


Figure 1. Integration of object detection and tracking

3.3. Occlusion detection and template update

In this paper, occlusion is declared by applying the response distribution given by [1]. If the object is declared “occluded,” we stop estimating its motion V_t and discard the motion likelihood term.

When the tracker is initialized, the occlusion state of the object is set to 1, i.e. $O_1 = 1$. Afterwards, the tracker has to estimate O_t at each time $t > 1$ by occlusion detection. Actually when the object is quite small, it is hard to distinguish between true occlusion and heavily ambiguous clutter (the latter case is called *virtual occlusion*).

A template updating technique in [8] is used in a conservative way to handle the drifting problem in tracking.

4. Integration of Detection and Tracking

The flowchart of the whole framework is illustrated in Figure 1, called “Detect-then-Detect-and-Track” (DtDaT). When the small object declares “occluded”, a counter reckons the number of continuous frames in which it is

“occluded”. Once the number is over a threshold (it is 5 in this paper), the object is “lost” in tracking. The tracking algorithm using APF is shown in Figure 2.

With the particle set $\{(X_{t-1}^{(i)}, \pi_{t-1}^{(i)}) | i = 1, \dots, N\}$ at time $t-1$, we proceed at time t as follows:

- **Predict:** If $O_{t-1} = 1$, estimate motion V_t and prediction error μ_t ; Otherwise $V_t = 0$, $R_t = R_{\text{max}}$. For $i=1, \dots, N$, simulate $X_t^{(i)} \sim N(X_{t-1}^{(i)} + V_t, R_t)$;
- **Update:** For $i=1 \dots N$, $\pi_t^{(i)} = P(Z_t | X_t^{(i)})$ by (6), consisting of the intensity, motion and detection likelihood terms.
- **Resample** (if necessary): with the particle weight set $\{\pi_t^{(i)} | i = 1, \dots, N\}$, run residual resampling (its virtue lie in insensitivity to the particle order compared with other techniques). Replace $\{(X_t^{(i)}, \pi_t^{(i)}) | i = 1, \dots, N\}$ by $\{(\tilde{X}_t^{(i)}, 1/N) | i = 1, \dots, N\}$.
- **Estimate:** If $O_{t-1} = 1$, output the average of all particles; Otherwise, select the particles (one or more) with the maximum weight and output the average of them. Detect occlusion for O_t setting. If $O_t = 1$, handle drifting by template update.

Figure 2. APF-based small object tracker with detection

5. Experiment Results

As a case study, the proposed algorithms are applied for ball localization in soccer game videos. The two testing videos in the following examples are “France” (260 frames at 360x288) and “Belgium” (375 frames at 360x240).

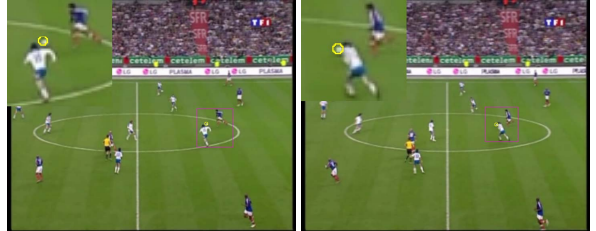


Figure 3. Tracking (Video “France”, frames 146, 154)

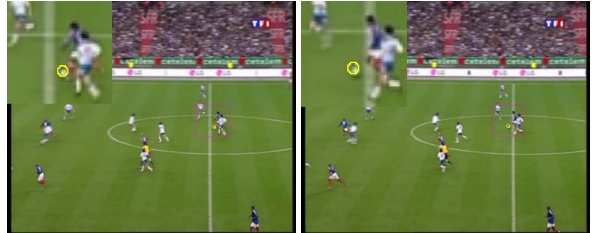


Figure 4. Tracking (Video “France”, frames 174, 177)

The object template size is decided by the strong detector, estimated as 5x5 and 7x7. The number of particles is 200. Results using the proposed algorithm are given in Figures 3-7. For all the figures, the yellow ellipse shows the ball position and size, and ellipse in black means a state estimate with the lowest confidence. For clarity, a zoomed portion

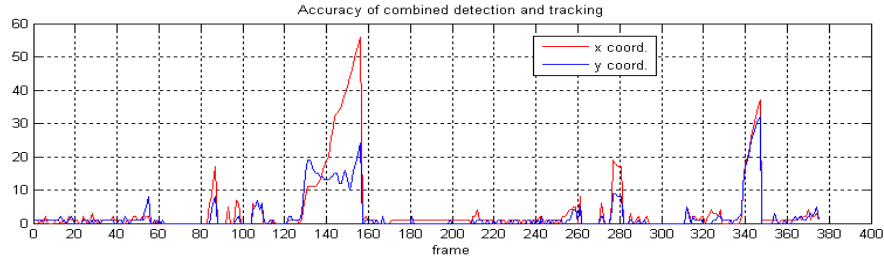


Figure 7. Detection and tracking accuracy in ball localization. Vertical axis corresponds to location error in pixels.

from the pink area is shown on the top left corner of each image.

For the occlusion case in video “France”, Figure 3 shows that the proposed method finds the ball when it reappears from occlusion as the most ambiguous case in tracking; while Figure 4 illustrates the proposed method follows the ball when it has left the field lines for the “clean” grass field.

When the player kicks the ball or strikes the ball strongly with the head, heavy motion blur happens. At this moment, the tracker cannot follow the ball since its appearance is extremely distorted. As explained, the proposed strategy resorts to restarting the strong detector; a couple of examples of such situations are shown in Figure 5 (video “France”) and Figure 6 (video “Belgium”).

Figure 7 shows the accuracy of combined detection and tracking for video “Belgium” (the ground truth was obtained manually). Several high peaks in the ball location error curve indicate tracking failures.

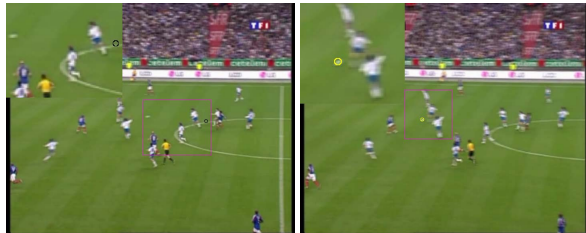


Figure 5. Motion blur (Video “France”, frames 198, 208)



Figure 6. Fast motion (Video “Belgium”, frames 84, 87)

6. Conclusions

Combination of detection and tracking provides the capability to handle small object localization. In this paper, a novel scheme of integrating them into an adaptive particle filter is proposed. Two different detector roles are clarified in this integration, i.e. a strong detector to trigger the tracker and a weak detector to enhance the tracker. In the tracker, some modifications in the particle filter are presented to cope with the cluttered background, occlusion and motion blur due to the object’s fast and abrupt motion. As a special case, the proposed framework is applied for ball localization in soccer game videos and experimental results demonstrate its efficiency and robustness.

Future work will focus on detection/tracking of a larger object around the detected/tracked small object since this larger object may provide more information related to the small object, meanwhile multiple-object-tracking is handled.

References

- [1] E. Arnaud, E. Memin, B. Cernuschi-Frias, “Conditional filters for image sequence based tracking application to point tracking”, *IEEE-T-IP*, 14(1):63-79, 2005.
- [2] S. Avidan, “Ensemble tracking”, *IEEE CVPR’05*, 2005.
- [3] T. Chateau, V. Gay-Belille, F. Chausse, J-T. Lapreste, “Real-time tracking with classifiers”, *ECCV’06*, 2006.
- [4] R. Collins, Y. Liu, “On-Line Selection of Discriminative Tracking Features”, *IEEE ICCV’03*, 2003.
- [5] N. Dalal, B. Triggs, “Histograms of oriented gradients for human detection”, *IEEE CVPR’05*, 2005.
- [6] Y. Huang, J. Llach, S. Bhagavathy, “Players and Ball Detection in Soccer Videos Based on Color Segmentation and Shape Analysis”, *Int. Workshop on Multimedia Content Analysis and Mining*, June, 2007.
- [7] D. Liang, Y. Liu, Q. Huang, and W. Gao. “A Scheme for Ball Detection and Tracking in Broadcast Soccer Video”. *PCM 2005*, pp. 864-875, 2005.
- [8] I. Matthews, S. Baker, T. Ishikawa, “The Template Update Problem”, *IEEE T-PAMI*, 26(6), 2004.
- [9] H. Nguyen, A. Smeulders, “Tracking aspects of the foreground against the background”, *ECCV’04*, 2004.
- [10] J. Odobez, D. Perez, S. O. Ba, “Embedding motion in model-based stochastic tracking”, *IEEE T-IP*, 15(11):3515-31, 2006.
- [11] K. Okuma, A. Taleghani, N. Freitas, J. Little, D. Lowe, “A boosted particle filter: multiple detection and tracking”, *ECCV’04*, 2004.
- [12] P. Perez, J. Vermaak, A. Blake, “Data fusion for visual tracking with particle filters”, *Proc. IEEE*, 92(3), 2004.
- [13] P. Viola, M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, *IEEE CVPR’01*.
- [14] L. Zhao, L. Davis, “Closely coupled object detection and segmentation”, *ICCV’05*, 2005.
- [15] S. Zhou, R. Chellappa, B. Maghaddam, “Appearance tracking using adaptive models in a particle filter”, *ACCV’04*, 2004.