

 Open access • Proceedings Article • DOI:10.1109/ICALIP.2008.4589956

Tracking the small object through clutter with adaptive particle filter

— [Source link](#) 

Yu Huang, J. Llach

Institutions: Thomson Corporation

Published on: 07 Jul 2008 - International Conference on Audio, Language and Image Processing

Topics: Video tracking, Motion estimation, Video game, Adaptive filter and Object detection

Related papers:

- [Adaptive dynamic model particle filter for visual object tracking](#)
- [Visual object tracking using particle filter](#)
- [Robust online tracking using orientation and color incorporated adaptive models in particle filter](#)
- [Detection-based particle filtering for real-time multiple-head tracking applications](#)
- [Learning and regularizing motion models for enhancing particle filter-based target tracking](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/tracking-the-small-object-through-clutter-with-adaptive-5eq7zazz0s>

Tracking the Small Object through Clutter with Adaptive Particle Filter[@]

Yu Huang, Joan Llach

Thomson Corporate Research, Princeton, New Jersey, USA

yu.huang07@gmail.com, joan.llach@thomson.net

Abstract

Cluttered background and occlusion cause large ambiguity in the tracking of video objects. When the object is small (like a soccer ball in broadcast game video signals), the ambiguity gets even more severe. In this paper, we propose an adaptive particle filter with effective proposal distribution to handle these situations. In the proposed tracking approach, motion estimation is embedded into the state transition to tackle abrupt motion changes and generate good proposal distributions. We also propose a mixture model to account for multiple hypotheses in the template correlation surface when estimating the appearance likelihood. In addition, motion continuity and trajectory smoothness are combined with template correlation in the observation likelihood to further filter out visual distracters. As an example of small object tracking, promising results of the ball tracking (as small as 30 pixels) in soccer game videos are presented to illustrate that the proposed scheme handles the cluttered background and occlusion effectively.

1. Introduction

Visual tracking is a crucial element of many computer vision systems. A lot of applications such as visual surveillance, smart rooms, video compression and vision-based interfaces often require a visual tracker to be robust in complex environments and efficient in computation [1, 4-7, 10, 12, 14].

Recently, particle filters (PF) have gained more attention in visual tracking [1, 4, 7, 10, 12, 14]. The efficiency and accuracy of the particle filter depends on two key factors: how a swarm of particles are generated by a proposal distribution and how these particles are weighted to approximate the real posterior state distribution. The pioneering work of *Condensation* [7] uses the state transition prior as the proposal distribution. This type of particle filter is prone to be distracted by background clutter because the state transition does not take into account the most recent observation.

As an alternative, the unscented particle filter was used in [12] to generate importance densities. However, this approach needs to convert likelihood evaluations into state space measurements. Furthermore, it is still likely to fail in the presence of abrupt motion changes. An adaptive particle filter

was proposed in [14], where adaptive state transition utilizing the new observation data on the affine flow constraints is realized and the diversity of particles is adapted based on motion estimation errors.

However, how to estimate the particle weight from the current observation data to differentiate the object from the cluttered background is more critical. Histogram and contour have been proved to be robust features in object tracking through clutter [4, 7, 11, 12], but they may not characterize well the appearance of small objects. The intensity-based appearance model is used in [14] and a mixture model is employed to handle its variation. Nonetheless, cluttered background is not coped with explicitly in [14].

Likewise, in [1] motion estimation is embedded into the state transition, too. Its likelihood function accounts for uncertainty in template matching based on correlation surface (with a fixed size) [9]. Despite that, no motion or trajectory information is involved into the likelihood calculation even though motion continuity and trajectory smoothness are helpful to filter out the visual distracters in complicated cluttered background. Moreover, this approach doesn't take into account multiple candidates [11] in the correlation surface.

In comparison, in [10] not only patch correlation (normalized cross correlation), shape and color information, but also motion measurements are introduced into the likelihood function. Even so, no template is adapted in this method. In contrast, "naïve update" [8] was performed. For this reason, it cannot handle severe occlusion or "drifting" artifacts explicitly.

1.1 Ball tracking in soccer game video

In this paper, our experiments focus on the soccer ball in game videos, an instance for small object tracking. Actually soccer video analysis is receiving increasing attention from researchers [2, 3, 13] as a convergence of computer vision and multimedia technologies, mainly motivated by applications such as event analysis, automatic indexing and object-based encoding. In particular, significant research work aims at obtaining the ball's position since it plays an important role in detecting key events and improving object-based compression performance.

However, detecting and tracking the ball in sports video from broadcast signals is a really challenging problem. The ball may look very similar in appearance to other regions of the

[@] Yu Huang is now working at Futurewei Technologies Inc., NJ, USA

image; for example, portions of the players' jerseys could trigger false alarms. It is also frequently merged with field lines or occluded by the players and, additionally, the ball moves fast most of the time and is quite small (less than 30 pixels in size) when the camera is capturing a wide view of the playfield.

1.2 Overview of our approach

In this paper, we propose an adaptive particle filter with effective proposal distribution to deal with severe cluttered background in small object tracking. Our proposal is inspired by [14] with combination of work in both [1] and [10]. To start with, motion estimation is embedded into state transition to tackle abrupt motion changes and generate a good proposal distribution. Then, we propose a mixture model to account for multiple hypotheses in the template correlation surface for estimating the appearance likelihood. Different from [1] and [10], we utilize the explicit motion measures through both the dynamic model and the likelihood function.

The paper is organized as follows. Section 2 discusses the proposed tracking algorithm, an adaptive particle filter. In Section 3, results using broadcast soccer videos are shown. Finally, conclusions are drawn in Section 4.

2. The Proposed Tracking Approach

Particle filter is a state space method for implementing a recursive Bayesian filter by Monte Carlo simulations. The key idea is to approximate the posterior probability distribution by a weighted particle set. Each particle represents one hypothetical state of the object, with a corresponding discrete sampling probability (weight). The mean state of an object is estimated at each time step by weighted average of all the particles. Usually resampling is used to alleviate particles' degeneracy.

The efficiency and accuracy of a particle filter for tracking relies on the definition of a good proposal distribution and an effective observation model for particle weights. Below we will give details on these issues in our proposed algorithm.

We define the object's state vector as $X = (x, y)$, where (x, y) is the window center of the object. The state space model for object tracking is formulated as

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

$$Z_t = g(X_t, \xi_t), \quad (2)$$

where X_t represents the object state vector, Z_t is the observation vector, f and g are the dynamic model and the observation model, respectively, and μ_t and ξ_t represent the process and observation noise, respectively.

2.1. Dynamic Model

The dynamic model characterizes the object state change between frames. Similar to [1, 10, 14], we directly obtain the apparent motion of the object by a hierarchical estimation framework [5, 6]. Since the most recent observation is used for state transition, a better proposal distribution is generated.

2.1.1. Adaptive Motion Model

Let us denote the estimated motion for the object as V_t . Accordingly, the dynamic model in (1) can be reformulated as

$$X_{t+1} = X_t + V_t + \mu_t, \quad (3)$$

with μ_t denoting the state prediction error.

2.1.2. Adaptive Process Noise Variance

To vary the diversity of particles, $\mu_t \in [\mu_{\min}, \mu_{\max}]$ is proportional to the motion estimation error, i.e. a defined residual measure $|uI_x + vI_y + I_t|$, where I_x, I_y, I_t are partial derivatives of the intensity function I with respect to x, y and t .

If motion estimation fails (by thresholding the average of absolute difference), motion is set zero like a "random walk" and μ_t is set as the maximum value.

2.2. Observation Model

The observation model measures the weights of particles based on a predefined likelihood function. Here it is defined as

$$P(Z_t | X_t) = P(Z_t^{\text{int}} | X_t) P(Z_t^{\text{mot}} | X_t)^{O_t-1} P(Z_t^{\text{trj}} | X_t)^{1-O_t-1}, \quad (4)$$

where $Z_t = \{Z_t^{\text{int}}, Z_t^{\text{mot}}, Z_t^{\text{trj}}\}$ and the intensity measurement

Z_t^{int} is assumed to be independent from either the motion measurement Z_t^{mot} or the trajectory measurement Z_t^{trj} , $O_t = 0$

if the object is occluded, and 1 otherwise. When the object is 'visible' trajectory constraints are not enforced, which avoids violating the temporal Markov chain assumption; on the other hand, when the object is occluded or motion estimation fails, the trajectory smoothness takes the place of motion continuity in the observation likelihood. Details of each likelihood component are given below.

The intensity measurement is computed with the similarity between the target model (template) and the candidate particle. A simple metric is the sum-of-squared-differences (SSD) for each particle as

$$Z_t = \arg \min_{X_t \in \text{Neib}} \sum_{\chi \in W} [T(\chi) - I(\chi + X_t)]^2 \quad (5)$$

where W is the object window, Neib is a small neighborhood around X_t , T is the object template and $I(\cdot)$ is the image in the current time.

This metric cannot be used directly for intensity likelihood in the case of highly cluttered background. Instead, the correlation surface [9] can better measure the uncertainty and generate a reasonable estimate.

The SSD-based correlation surface for each particle in its support area Neib is defined as

$$r(X_t) = \sum_{\chi \in W} [T(\chi) - I(\chi + X_t)]^2, X_t \in \text{Neib}. \quad (6)$$

Compared with the fixed size of the correlation surface in [1, 9], the surface size in our proposal varies (from 3x3 to 11x11 pixels in the examples) proportionally to the motion estimation error as well given in formula (3), similar to the process noise variance, which provides a flexible measure of the ambiguity in template matching.

Inspired by [7, 11], we assume having detected J candidates from the correlation surface inside Neib . As a result, $J+1$ hypothesis can be defined as:

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

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

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 over $Neib$ and hence the true match-oriented measurement is Gaussian distributed. Consequently, the combined intensity likelihood is formulated as

$$P(Z_t^{int} | X_t) = q_0 U(\cdot) + C_N \sum_{j=1}^J q_j N(r_t, \sigma_t), \quad (7)$$

$$\text{s.t. } q_0 + \sum_{j=1}^J q_j = 1,$$

where C_N is the normalization factor, and q_j is the prior probability ($q_0=0.5$ in the examples show in this paper) for hypothesis H_j , $j=0, \dots, J$. This mixture model gives a reasonable approximation to the intensity likelihood for each particle under the clutter.

Similar to [1, 9], "occlusion" detection (object's locating in highly cluttered regions is called "virtual" occlusion) is implemented as follows. To approximate the probability distribution on the true match location, a response distribution is defined as

$$D(X_t) = \exp(-\rho \cdot r(X_t)), \quad (8)$$

where ρ is the normalization factor, chosen to ensure $\sum_{(x,y) \in Neib} D_i(x,y) \approx 1.0$. We use the maximal eigenvalue of the covariance matrix R_i associated to measurements Z_t^{int} as confidence metric: if it is bigger than a given threshold, we detect the presence of "occlusion". In fact, when more peaks occur in the response distribution, the approximated Gaussian distribution gets flat, and accordingly the eigenvalues get bigger.

If the particle is regarded as "occluded," a minimal weight is set to it, which may result in its rejection in resampling.

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

$$d_{mod}^2 = (\Delta x_t | -\overline{\Delta x})^2 + (\Delta y_t | -\overline{\Delta y})^2, \quad t > 1$$

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

$$\overline{\Delta x} = \sum_{s=t-k}^{t-1} |x_s - x_{s-1}| / k, \quad \overline{\Delta y} = \sum_{s=t-k}^{t-1} |y_s - y_{s-1}| / k, \quad (k=10)$$

Hence the motion likelihood is calculated as

$$P(Z_t^{mot} | X_t) = \frac{1}{\sqrt{2\pi}\sigma_{mot}} \exp\left(-\frac{d_{mot}^2}{2\sigma_{mot}^2}\right). \quad (9)$$

This component accounts for constraints from motion continuity of the object. In comparison to [10], our motion likelihood takes into account the recent motion history of the object.

The trajectory likelihood is estimated from the particle's closeness to a trajectory that is obtained from past positions of

the object. Let's denote the trajectory function in a polynomial form

$$y = \sum_{i=0}^m a_i x^i, \quad (10)$$

where a_i are the polynomial coefficients and m is the order of the polynomial function (for the examples in this paper, $m=2$).

Only past "visible" object positions are used for trajectory fitting. A forgotten factor $F = \lambda_f^{t-o}$ is defined, where λ_f is the forgotten ratio, ($0 < \lambda_f < 1$, $\lambda_f=0.9$ in this paper), and t_o is the number of frames the object has been occluded. This factor accounts for that, the more frames the object has been occluded, the less reliable the fitted trajectory is ($t_o=2$ illustrated in Fig. 1).

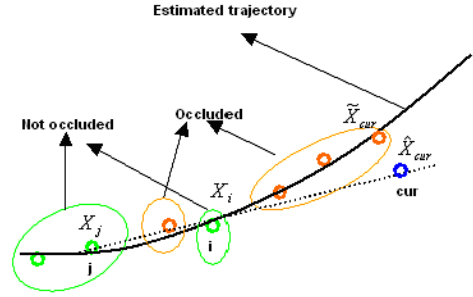


Fig. 1 Illustration of object trajectory fitting

As a result, the trajectory smoothness likelihood is given by

$$P(Z_t^{trj} | X_t) = \frac{1}{\sqrt{2\pi}\sigma_{trj}} \exp\left(-\frac{(d_{trj}/F)^2}{2\sigma_{trj}^2}\right), \quad (11)$$

where the closeness metric is $d_{trj} = |y - \sum_{i=0}^m a_i x^i|$.

2.3. Position Refining and Template Update

If the object is "visible" at the previous frame $cur-1$ ($O_{cur-1}=1$), we get the state estimate by weighted average of all particles; otherwise, taking the particle with the maximum weight.

After state estimate, we run "occlusion" detection again (explained in Section 2.2.3) to set O_{cur} for the current frame cur . If $O_{cur}=0$, we refine it by projecting a predicted one onto the estimated trajectory to keep its motion smoothness. Illustrated in Fig. 1: given the two last reliable positions X_j and X_i at frames j and i respectively ($i > j$), the predicted one \hat{X}_{cur} is calculated as

$$\hat{X}_{cur} = X_i + (X_j - X_i) * (cur-i)/(i-j).$$

Its projection \tilde{X}_{cur} is defined as the point on the trajectory closest to \hat{X}_{cur} . Hence, we refine the object position as

$$X_{cur} = (1 - \lambda_f^{t-o}) \hat{X}_{cur} + \tilde{X}_{cur} * \lambda_f^{t-o}. \quad (12)$$

If $O_{cur}=1$, we employ the template update approach [8] in a conservative way to cope with the appearance variation, i.e. the "drifting" artifact in tracking.

2.4. The outline of the proposed method

Fig. 2 shows the framework of our proposed approach for small object tracking.

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:

- **Prediction:** If $O_{t-1} = 1$, estimate motion V_t and prediction error μ_t ; Otherwise $V_t = 0$, $\mu_t = \mu_{\max}$. For $i = 1 \dots N$, simulate $X_t^{(i)} \sim N(X_{t-1}^{(i)} + V_t, \mu_t)$;

- **Updating:** For $i = 1 \dots N$, $\pi_t^{(i)} = P(Z_t | X_t^{(i)})$ by (4), consisting of the intensity, motion and trajectory likelihood terms by (7), (9) and (11).

- **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. Otherwise, project the estimated position onto the trajectory by (12).

Fig. 2 Particle filter-based small object tracking

3. Experimental Results

We have implemented the proposed tracking method in Visual C++ and it runs at about 15fps on a 3.2GHz PC platform. The video from broadcast soccer game signals in the following examples is 360x240, 30Hz and lasts 190 frames. The tracker is initialized manually. The template size for the soccer ball is 5x5. The number of particles is 200. In all the examples in this paper, the yellow ellipse shows the ball position and size (for clarity, a zoomed portion from the pink area of the frame is shown on the top left corner of each image).

To illustrate the performance of the proposed method, we've ran two classic tracking methods for comparison: one is an optic flow-based tracker with template update [8] (a KLT tracker) and the other is a traditional template matching tracker (the TM tracker) updated by an IIR filter (the update ratio is 0.15). Shown in Fig. 3, the template matching (TM) tracker first falls in a region on the player jersey and then drifts away (frame 43). The estimated normalized cross correlation is still high (0.86) on frame 43. In Fig. 4, we see that the optical flow-based (KLT) tracker drifts away on frame 36 due to partial occlusion by a player jersey area.

Tracking results of the proposed algorithm are given in Fig. 5-9. In this video, occlusion of the ball by the player occurs twice (frames 36-40, 147-153) and the field mark lines merge with the ball twice (frames 136-140, 175-176). Our proposal handles both cases successfully.



Fig. 3: Failure of the TM tracker (frames 35, 43)



Fig. 4: Failure of the KLT tracker (frames 35, 36)



Fig. 5 Tracking results (frames 34, 41, 146, 153)

For the first case, Fig. 5 shows that our method still finds the ball when it reappears from occlusion by players, the most complicated case in ambiguity. Furthermore, Fig. 6 shows the resampled particles (ellipses in blue) when the ball is physically occluded by players (ellipse in black means the final estimate with the lowest confidence). We can see that more diversified particles are retained to wait for the ball to reappear.

The second case is illustrated in Fig. 7, where our tracker follows the ball when it has left the field lines for the “clean” grass field. Since field lines look similar in color to the ball, template matching becomes more ambiguous when the ball approaches them. Likewise, Fig. 8 shows the resampled particles when the ball falls onto the field line. This case is regarded as “virtual” occlusion since its uncertainty is similar to a real occlusion. After resampling, particles close to the field lines are retained which will be propagated to detect the ball that goes into the “clean” grass area again.

The tracking error is given in Fig. 9 where the ground truth is generated manually (when the ball is occluded, its ‘real’ position has to be obtained by interpolation). Actually the several peaks in this error curve indicate the time periods when the ball is occluded by the player or merges with the field line.

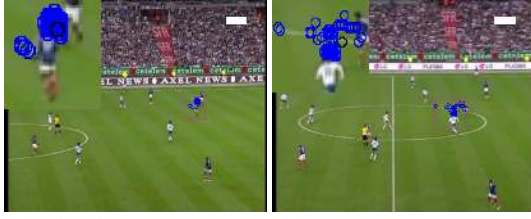


Fig. 6 Ball's occlusion by player (frame 39, 147)

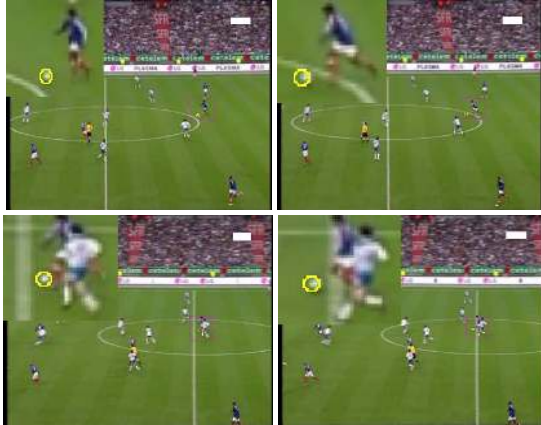


Fig. 7 Tracking results (frames 134, 141, 174, 177)

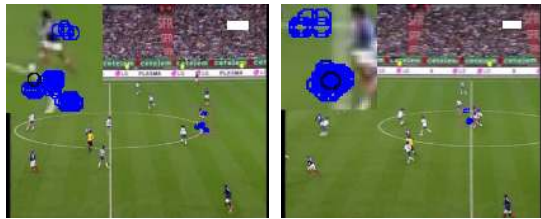


Fig. 8 Ball's merging with the line (frame 138, 176)

4. Conclusions

An adaptive particle filter for small object tracking which can effectively handle cluttered background and occlusion has been proposed. The adaptive motion model is applied to get better proposal distributions with varied diversity of particles. To further filter out visual distracters, motion continuity and trajectory smoothness are combined with the template correlation in the observation likelihood. Experimental results

when our approach applies for soccer ball tracking show that it can deal with challenging situations with success, e.g. ball merging with field lines or occlusions.

In future work, a small object detector [3] will contribute information into the particle filter-based tracker to handle motion blur and long-duration occlusions. Besides, other interacting objects at the neighborhood, such as player/referee, are needed to be tracked at the same time, expanding the system towards multiple object tracking.

5. 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] Y. Gong, L. T. Sin, C. H. Chuan, H. Zhang, and M. Sakauchi, "Automatic parsing of TV soccer programs", Proc. Multimedia Computing & Systems, pp167-174, 1995.
- [3] 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 (MCAM'07), June, 2007.
- [4] Y. Huang, J. Llach, "Variable Number of Informative Particles for Object Tracking". IEEE ICME'07, July, 2007.
- [5] Y. Huang, T. S. Huang, H. Niemann, "Segmentation-based Object Tracking Using Image Warping and Kalman Filtering", IEEE ICIP'02, Rochester city, US, Sept. 2002.
- [6] Y. Huang, T. S. Huang, H. Niemann, "Region-based Method for Model-free Object Tracking", IAPR ICPR'02, Quebec city, Canada, Aug., 2002.
- [7] M. Isard and A. Blake, "Condensation -Conditional density propagation for visual tracking," IJCV, 29(1), 1998.
- [8] I. Matthews, S. Baker, and T. Ishikawa, "The Template Update Problem", IEEE T-PAMI, 26(6), pp810- 815, 2004.
- [9] K. Nickels, S. Hutchinson, "Estimating uncertainty in SSD-based feature tracker", Image & Vision Computing, 20(1), pp47-58, 2002.
- [10] J. Odobez, D. Perez, S. Ba, "Embedding motion in model-based stochastic tracking", IEEE T-IP, 15(11), 2006.
- [11] P. Perez, J. Vermaak, and A. Blake, "Data fusion for visual tracking with particle filters", Proc. IEEE, 92(3), 2004.
- [12] Y. Rui, Y. Chen, "Better Proposal Distributions: Object Tracking Using Unscented Particle Filter", IEEE CVPR 2001.
- [13] X. Yu, C. Xu, Q. Tian, and H. W. Leong, "A ball tracking framework for broadcast soccer video", ICME'03.
- [14] S. Zhou, R. Chellappa, and B. Maghaddam, "Appearance tracking using adaptive models in a particle filter", ACCV, 2004.

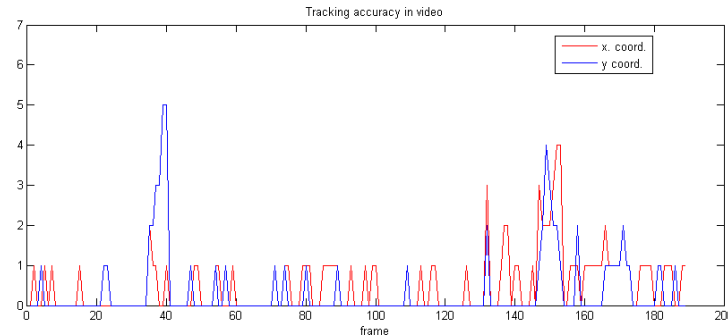


Fig. 9 Tracking accuracy in ball localization (vertical axis corresponds to location error in pixels).