BOOSTED INTERACTIVELY DISTRIBUTED PARTICLE FILTER FOR AUTOMATIC MULTI-OBJECT TRACKING

Yi Wu¹, Xiaofeng Tong², Yimin Zhang², Hanqing Lu¹

¹National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science
{yw, luhq}@nlpr.ia.ac.cn
²Intel China Research Center, Beijing, P.R. China
{xiaofeng.tong, yimin.zhang}@intel.com

ABSTRACT

In this paper, we propose a Boosted Interactively Distributed Particle Filter (BIDPF) to address the problem of automatic multi-object tracking in the application of player tracking in broadcast soccer video. The interactively distributed particle filter technique (IDPF) is adopted to handle the mutual occlusions among targets. The proposal distribution using a mixture model that incorporates information from the dynamic model and the boosting detection is introduced into the IDPF framework. The boosting proposal distribution quickly detects targets, while the IDPF process keeps the identity of targets during mutual occlusions. Moreover, the foreground observation is extracted by using the color model of the playfield to speed up the boosting detection and reduce false alarms. The foreground is also used to develop a data-driven potential model to improve the IDPF performance. We test the proposed approach on several video sequences and the results demonstrate that our system is able to track a variable number of objects in a dynamic scene and correctly maintain their identities regardless of camera motion and frequent mutual occlusions.

Index Terms—Boosting, Proposal distribution, Particle filter, Distributed multi-object tracking, Data-driven potential model

1. INTRODUCTION

Multiple object visual tracking is an important research area of computer vision. A multi-object tracking system should be able to track a variable number of objects in a dynamic scene and maintain the identity of targets regardless of occlusions and visual perturbations. The occlusion and presence of a varying number of objects create complex interactions with overlap and ambiguities.

In [1], Pérez et al. used the particle filter [2] to track multiple targets. Okuma et al. [3] further extended it by incorporating a boosting detector [4] into the particle filter, which is called boosted particle filter (BPF). However, as their system did not have explicit mechanisms to model mutual occlusions among targets, it can not maintain the identity of each target after occlusions.

Various approaches have been proposed to model the interaction among targets in multi-object tracking. In [5] a MCMC-based particle filter was proposed to track multiple interacting objects based on the assumption that there is no overlap among objects. Yu et al. [6] allowed for collaboration among different trackers by modeling the joint prior of objects using a Markov Random Network to prevent the trackers from falsely merging. However, they did not deal with the data association between targets and observations so that their approach may switch the labels of targets after occlusions as illustrated in [7].

As pointed out in [7], multi-object tracking of similar objects in appearance (such as soccer players in the same team) fails when the objects present mutual occlusions. To circumvent this problem, Qu et al. introduced an interactively distributed particle filter (IDPF) using an interaction model in the particle filtering implementation to adjust the weight of each particle.

In this work, we propose an automatic multi-object tracking method and apply it for player tracking in broadcast soccer video. Incorporating the boosting detector into the particle filter is helpful for automatic initialization of a variable number of targets, while the IDPF is a powerful approach to model the interaction among targets. The two approaches are complementary thus fusion of them will significantly improve the multi-object tracking performance. We use a boosting detector to guide the IDPF. The proposal distribution consists of a probabilistic mixture model that incorporates information from the dynamic model and the boosting detection. This enables us to quickly detect and track players in a dynamically changing scene. Moreover, the color model of the playfield is learned to detect the foreground to improve the boosting detection. The foreground observation is also used to develop a data-driven potential model to prevent the tracker from falsely merging. We
call the resulting algorithm the Boosted Interactively Distributed Particle Filter (BIDPF).

![Image](image_url)

Figure 1. (a) Field extraction. (b) Foreground detection. (c) Player detection.

2. TARGET DETECTION

Target detection is achieved by running a boosted cascade of Haar features [4]. To improve the performance the foreground is extracted to filter out the background areas. The color model of the field is learned by accumulating HSV color histograms. Then the field is extracted through dominant color segmentation, morphological filtering and connect-component analysis (Fig. 1(a)). We obtain the foreground is extracted to filter out the background areas. The color model of the field is learned by accumulating foreground. Target detection is achieved by running a boosted cascade of Haar features [4]. To improve the performance the color model (Fig. 1(b)). The boosting detector is then formulated by modeling the interaction among the objects’ occlusion problems effectively. In this approach, the interactively degraded to multiple independent Bayesian trackers. One key issue in multi-object tracking is how to model the interaction among targets. This model could not interpret the state not the observation, and need a balance process to get the new state. This model is based on the state not the observation, and need a balance process to get the new state. This model is based on the traditional Bayesian tracking approaches.

3.1 Bayesian Formulation

Denote the state of target \( i \) in time \( t \) and the corresponding observation as \( x^i_t \) and \( y^i_t \) respectively. The state set up to time \( t \) is \( x^i_{0:t} \), where \( x^i_0 \) is the initial state, and the observation set up to time \( t \) is \( y^i_{0:t} \). The interactive observations of \( y^i_t \) at time \( t \) is denoted by \( y^c_t \) where the elements of \( J_c \) are the indexes of objects whose observations interact with \( y^i_t \).

The purpose of tracking is to estimate the posterior based on all the observations \( p(x^i_{0:t} | y^i_{0:t}, y^c_{1:t}) \). When \( y^i_t \) has no interaction with other objects all the time, \( p(x^i_{0:t} | y^c_{1:t}) = p(x^i_{0:t}) \), this formulation is consistent with the regular particle filter approach.

Using the conditional independence properties, we can formulate the density propagation for each interactive tracker as follows [7]:

\[
p(x^c_{0:t}, y^i_{1:t}) = k_t p(y^i_t | x^c_t) p(x^c_t | x^c_{t-1}, y^c_{t-1}, y^i_{t-1}) p(y^i_t | x^c_t, y^i_t)
\]

where \( p(y^i_t | x^c_t) \) is the local likelihood, \( p(x^c_t | x^c_{t-1}) \) is the state transition density, which are similar with the traditional Bayesian tracking approaches. \( p(y^i_t | x^c_t, y^i_t) \) is called “interactive likelihood” [7], which characterizes the interaction among targets. When there is no interaction among targets, this formulation will be degraded to multiple independent Bayesian trackers.

3.2 Particle filter implementation

In the particle filter implementation, the posterior probability \( p(x^c_{0:t}, y^i_{1:t}) \) is characterized by a set of weighted samples \( x^c_{0:t}, w^i_{1:t} \), where \( \sum_{n=1}^{N_s} w^i_{1:n} = 1 \).

According to the importance sampling theory, the weights are updated as [7]:

\[
w^i_{1:n} \propto \frac{p(y^i_t | x^i_t) p(x^c_t | x^c_{t-1}, x^c_{t-2}) p(y^i_t | x^c_t, y^i_t)}{q(x^c_t | x^c_{t-1}, y^c_{t-1}, y^i_t)} w_{t-1}^i
\]

Figure 2. These three objects have interaction with each other and they form a group \( G \).

3.2.1 The Interactive likelihood \( p(y^i_t | x^c_t, y^i_t) \)

A data-driven potential model is proposed based on the foreground observation to estimate the interactive likelihood \( p(y^i_t | x^c_t, y^i_t) \) to adjust the weights of particles.

Denote \( G \) as the image group formed by the interactive objects, \( F_G \) as the image foreground and \( S_i \) as the union of all player regions in \( G \) (see Fig. 2 for illustration). \( p_{11} \) is the probability that a pixel in a player belongs to the foreground, and \( p_{01} \) is the probability that a pixel in a player does not belong to the foreground. \( p_{00} \) is
the probability that a pixel outside players is from the foreground, and \( p_{b0} \) is the probability that a pixel outside players is not from the foreground. \( p_{11}+p_{01}=1 \), \( p_{10}+p_{00}=1 \), and \( p_{01}<p_{00} \). When computing \( \phi^{t,n}_{i} \) we fix the state \( x_{j} \). The \( N \) particles of \( x_{j} \) incorporation with \( x_{j} \) form \( N \) coverage for the \( G \), namely \( N \) interpretations for the foreground observation.

Assuming the pixels are independent, we get the following likelihood:

\[
p(y_{t}^{n}|x_{t}^{m}, y_{t}^{i}) = \prod_{g \in G} p(g|x_{t}^{m})
= \prod_{1}^{N} p_{11}^{N_{11}} p_{10}^{N_{10}} p_{01}^{N_{01}} p_{00}^{N_{00}}
= (1-p_{01})^{F_{c}}(1-p_{10})^{F_{c}} \left( \frac{p_{10}}{1-p_{01}} \right)^{N_{01}} \left( \frac{p_{01}}{1-p_{10}} \right)^{N_{11}}
= C_{g} e^{-\lambda_{10} N_{10} + \lambda_{01} N_{01}}
\]

Where \( C_{g} \) is a normalization constant, \( \lambda_{10} \) and \( \lambda_{01} \) are two coefficients depending on \( p_{10} \) and \( p_{00} \) respectively, and \( \cap \) is the intersection of two regions. The likelihood only depends on \( N_{10} \) and \( N_{01} \). As illustrated in Fig. 2(c), the white pixels outside the rectangles form \( N_{10} \) and the black pixels inside the rectangles form \( N_{01} \). This likelihood was also used in crowd segmentation [9] and is useful to interpret the foreground observation.

By exploiting the proposed potential model, the interactive likelihood \( p(y_{t}^{n}|x_{t}^{m}, y_{t}^{i}) \) reduces the probability that objects’ estimates do not interpret the observations well. Therefore, it can separate the observation in occlusion and thus solve the merging problem.

3.2.2 The state transition density \( p(x_{t+1}^{m}|x_{t}^{m-1}, x_{t}^{i-2}) \)

To solve the labeling problem, the state transition probability \( p(x_{t}^{m}|x_{t}^{m-1}, x_{t}^{i-2}) \) is estimated by

\[
p(x_{t}^{m}|x_{t}^{m-1}, x_{t}^{i-2}) = \frac{p(x_{t}^{m-1}|x_{t}^{m-1}, x_{t}^{i-2}) p(x_{t}^{m-2}|x_{t}^{m-1}, x_{t}^{i-2})}{p(x_{t}^{m-2}|x_{t}^{m-1})}
\]

where \( p(x_{t}^{m-1}|x_{t}^{m-1}, x_{t}^{i-2}) \) is the traditional state transition density and we estimate it using a constant acceleration model. The inertia weight \( \phi^{t}_{t}(x_{t}^{m-1}, x_{t}^{i-1}, x_{t}^{i-2}) \) is defined as [7]

\[
\phi^{t}_{t}(x_{t}^{m-1}, x_{t}^{i-1}, x_{t}^{i-2}) = p(x_{t}^{m-1}|x_{t}^{m-1}) \phi^{t}_{t}(x_{t}^{m-1}, x_{t}^{i-1}, x_{t}^{i-2})
\]

\[
\alpha \exp \left\{ -\frac{\left( \theta_{t}^{m-1} \right)^{2}}{\sigma_{1}^{2}} \right\} \exp \left\{ -\frac{\left( \| \tilde{v}_{t}^{m-1} - \| \tilde{v}_{t}^{m-1} \|^{2} \right)^{2}}{\sigma_{1}^{2}} \right\}
\]

where \( \alpha \) are prior constants which characterize the allowable variances of the motion vector’s direction and speed, respectively. \( \theta_{t}^{m-1} = x_{t}^{m-1} - x_{t}^{i-1} \) is the motion vector of the particle \( x_{t}^{m-1} \). \( \tilde{v}_{t}^{m-1} - x_{t}^{i-1} - x_{t}^{i-2} \) represents the reference motion vector from \( x_{t}^{i-2} \) to \( x_{t}^{i-1} \). \( \theta_{t}^{m-1} = \angle(\tilde{v}_{t}^{m-1}, \tilde{v}_{t}^{m-1}) \) is the angle between \( \tilde{v}_{t}^{m-1} \) and \( \tilde{v}_{t}^{m-1} \). \( \| \| \) represents the \( l_{2} \) norm.

3.2.3 The proposal distribution \( q(x_{t}^{m-1}|y_{t}^{i-1}, y_{t}^{i-2}) \)

A critical issue in keeping particle filtering effective is how to design the proposal distribution. We use a mixture of Gaussians model that combines both the dynamics prior and the boosting detections:

\[
q(x_{t}^{m-1}|y_{t}^{i-1}, y_{t}^{i-2}) = q_{g}(x_{t}^{m-1}|y_{t}^{i-1}, y_{t}^{i-2}) + (1 - \alpha_{q})p(x_{t}^{m-1})
\]

The parameter \( \alpha_{q} \) is dynamically updated by the overlap between the Gaussian distribution of boosting detection and the dynamics prior. The Nearest Neighbor algorithm is used to do the data association of assigning boosting detections to the existing tracks.

3.2.4 The local likelihood \( p(y_{t}^{i}|x_{t}^{i}) \)

The observation of the target is represented by a kernel density estimation of the color distribution extracted from the region \( R(x_{t}^{i}) \) centered at the location \( x_{t}^{i} \).
We apply the Bhattacharyya coefficient to evaluate the similarity between the current observation $K(x_t)$ and the reference model $K^* = \{K^*(n;x_0)\}_{n=1}^{N}$.

$$d[K^*, K(x_t)] = (1 - \rho[K^*, K(x_t)])^{1/2}$$

$$\rho[K^*, K(x_t)] = \sum_{n=1}^{N} \sqrt{k^*(n;x_0)k(n;x_t)}$$

In order to encode the spatial information of the observation, we adopt a multi-part color model [10], which splits an object vertically into two parts. Thus, the local likelihood of target $i$ is defined as:

$$p(y_i^b | x_i) \propto e^{-\frac{1}{2}\sum_{j=1}^{N} \rho^2 [K_i^*(n), K_i(n)]}$$

Figure 4. Two tracking results from the same sequence with 1047 frames in clip B.

4. EXPERIMENTAL RESULTS

The experiments are carried out on two video clips: one with the resolution of $720 \times 576$ and the other with the resolution of $960 \times 544$. For simplicity, we use $A$ and $B$ to denote the two video clips, respectively.

Fig.3 demonstrates the comparison between the tracking results of the system in [3] and our system. We can see that when new players enter the scene our system can quickly detect them and track them robustly even with heavy occlusions. Fig. 3(b) shows the close-up views of the elliptic region labeled in the first image of Fig. 3(a). The color of surrounding box for each player is unique if the player has no interaction with others. Otherwise, the color of the box is changed to black. Our system can detect targets entering the scene, while the IDPF process enables to keep the identity of each player during occlusions. Furthermore, the proposed data-driven potential model based on the foreground observations prevents the trackers from falsely merging and makes the tracking more robust.

5. CONCLUSION

In this paper, we developed an automatic multi-object tracking system that is able to track a variable number of objects in a dynamic scene and correctly maintain the identity of targets regardless of camera motions and mutual occlusions.

The proposed BIDPF is suitable for tracking a variable number of targets. The boosting proposal can quickly detect targets entering the scene, while the IDPF process enables to keep the identity of each player during occlusions. Furthermore, the proposed data-driven potential model based on the foreground observations prevents the trackers from falsely merging and makes the tracking more robust.

6. ACKNOWLEDGEMENTS

This work is partially supported by the National Hi-Tech Research and Development Program of China (863 Program, Grant No. 2006AA01Z315 and No. 2006AA01Z117), the National Natural Science Foundation of China (Grant No. 60470101) and the Natural Science Foundation of Beijing (Grant No. 4072025).

7. REFERENCES