TRACKING OF MULTIPLE INTERACTING OBJECTS USING A NOVEL PREDICTION MODEL

Zhijie Wang, Hong Zhang, Nilanjan Ray

Department of Computing Science, University of Alberta
221 Athabasca hall, University of Alberta, Edmonton, Alberta, T6G 2E8

ABSTRACT

Tracking multiple interacting objects is an interesting and difficult task in computer vision. Two common problems in this field are a single object with multiple tracks and a single track with multiple objects. Most of the existing algorithms address the first problem but not the second one. In this paper, to solve the second problem we propose a new algorithm with a novel prediction model, which exploits the idea of penalizing outliers in statistics. The experiments show that our proposed algorithm is more robust than the existing algorithms in tackling both the aforementioned problems.

Index Terms— Tracking, interacting objects, particle filter

1. INTRODUCTION

This work deals with the multiple-object tracking problem. Our objective is to estimate all the objects’ trajectories over time and maintain a unique and correct identity for each of them throughout. This tracking problem becomes difficult when objects with similar appearance interact with each other, i.e. move close to or cross each other.

Classical methods for multiple-object tracking include multiple hypothesis tracker [1] and joint probabilistic data association filter [2, 3]. These methods require an external procedure to detect objects in every frame and they have difficulty in coping with nonlinear models and non-Gaussian noises. Therefore, a class of new Bayesian sequential estimation approaches have become popular recently. They make decisions based on the integrated information over time; as a result they tend to be more accurate. Additionally, they avoid the problem of explicit data association. These approaches require two models, a prediction model predicting the current state based on the previous state and a measurement model updating the predicted state according to the current observation. These Bayesian sequential estimation approaches must deal with the problems that arise when objects interact with each other. For example, Czyz’s algorithm [4] adds a minimum distance constraint between two tracks to the measurement model. In Mac Cormick’s algorithm [5], the authors use the probabilistic exclusion principle in the measurement model so that measurement from a certain region can only be associated with one object. MRF-based algorithm [6] introduces a prediction model using a Markov random field to model the interaction among objects. All these algorithms are able to solve the problem of a single object with multiple tracks; i.e. multiple tracks tend to be attracted to the same object with the best likelihood score when objects interact. These algorithms however have difficulties to deal with another problem when objects interact: a single track with multiple objects; i.e. a single track gets attracted by multiple objects resulting in meaningless tracking information. The goal of our research is to solve this problem by proposing a new prediction model that exploits the idea of penalizing outliers in statistics [9]. This prediction model encourages each track to focus on one object rather than spread to multiple ones incorrectly. As a result, our proposed algorithm is able to deal with the problem of a single track with multiple objects.

Other related works [7, 8] also exist that try to solve the problems resulting from multiple objects interacting with each other. For example, BraMBLE [7] incorporates 3D geometrical information to solve the occlusion problem. 3D scene modelling however is difficult and problematic. Therefore those multiple-object tracking algorithms are not considered in this paper.

The rest of this paper is organized as follows. In Section 2 the complete Bayesian sequential estimation solution is first presented followed by our new prediction model. The solution is then implemented with a particle filter in Section 3. Section 4 includes the experimental results showing the performance of our proposed algorithm, and finally the conclusion is drawn in Section 5.

2. BAYESIAN MULTIPLE-OBJECT TRACKING

In this section, we first describe how Bayesian sequential estimation works in general to track multiple objects, and then we propose our prediction model that can solve the two problems associated with object interaction: a single object with multiple tracks and a single track with multiple objects.

Bayesian sequential estimation solves the multiple-object tracking problem by recursively computing the posterior distribution \( p(X_t|Z^t) \) according to

\[
p(X_t|Z^t) = \frac{k p(Z_t|X_t) p(X_t|Z^{t-1})}{k p(z_t|X_t) \int_{X_{t-1}} p(X_{t-1}|X_{t-1}, Z^{t-1}) p(X_{t-1}|Z^{t-1})}
\]

\[(1)\]

[6]. \( X_t \) is the joint state of all \( n \) objects \( \{X_{it}\}_{i=1}^n \) at time \( t \), and \( Z^t \) includes all the observations \( \{z_1, \ldots, z_t\} \) up to time \( t \). \( p(z_t|X_t) \) is the measurement model defined as \( p(z_t|X_t) = \Pi_{i=1}^n p(z_{it}|X_{it}) \), assuming that the appearances of objects are conditionally independent. \( p(X_t|X_{t-1}, Z^{t-1}) \) is the prediction model we focus on in this paper. There have been different works on defining the prediction model to solve the two aforementioned problems. Therefore, we will first discuss the effects of some existing prediction models, and then propose our prediction model for the problem remaining in the existing models.

The causes of the two problems can be easily understood by analyzing two interacting objects. Assume we are tracking two objects experiencing random walk with a joint state variable \( X_t \). Two current observations are \( v_1 \) and \( v_2 \). For simplicity and without loss of
generality, we consider the object states and observations in this case to be scalar numbers. Then the simplified problem is as follows. The previous state is $X_{t-1} = [a, b]$ and each individual object has a static motion model with a zero mean Gaussian noise. We are interested in estimating the current state $X_t$ given the current two observations at $a$ and $b$. An observation is supposed to be drawn from a Gaussian distribution centered at the real state; then the joint measurement model is shown in Fig. 1 where each track is assumed to be able to associate with any observation. The simplest prediction model is the independent model

$$p(X_t | X_{t-1}, Z^{t-1}) \propto \Pi_{i} p(X_{t|t-1})$$  

where $p(X_{t|t-1})$ is the individual object motion model. Fig. 2a shows the predicted distribution when this model is used. In this case, no constraint is applied to the predicted states $[a, a]$ and $[b, b]$ where two tracks are locating at the same object. Therefore in its posterior distribution (Fig. 2b), the two peaks at $[a, a]$ and $[b, b]$ indicate that the problem that two tracks are attracted by the same object; i.e. a single object with multiple tracks. Additionally, no constraint is applied to the predicted state $[b, a]$ which results in the peak at $[b, a]$ in the posterior distribution. This peak along with the peak at $[a, b]$ show that each track is tracking both objects; i.e. a single track with multiple objects. Therefore, the alternative existing algorithms in [5, 4, 6] proposed prediction models to solve the problems, and most of them are based on similar ideas. Take MRF-based algorithm as an example. It adds a pairwise interaction function $\phi(X_{it}, X_{jt})$ to the independent prediction model, so that the prediction model becomes

$$p(X_t | X_{t-1}, Z^{t-1}) \propto \Pi_{i} p(X_{t|t-1}) \cdot \Pi_{ij \in \mathcal{T}} \phi(X_{it}, X_{jt}).$$  

(3)

Since the interaction function $\phi(X_{it}, X_{jt})$ is aware that multiple tracks are interacting with each other, this prediction model is able to penalize the states $[a, a]$ and $[b, b]$ where the two tracks get too close to each other as shown in its predicted distribution (Fig. 3a). Therefore in its posterior distribution (Fig. 3b), the two peaks at $[a, a]$ and $[b, b]$ are suppressed, and this implies the problem of a single object with multiple tracks can be dealt with by this prediction model. However the peak at $[b, a]$ still can not be suppressed because two tracks are not interacting.

In order to solve the remaining problem of a single track with multiple objects, we introduce one more term to the prediction model in MRF-based algorithm. This term exploits the idea of penalizing outliers in statistics [9]. Therefore, the peak at $[b, a]$ in Fig. 3a can actually be considered as an outlier, which may be due to flaws in the prediction model (Eq. 3) that generates an assumed family of distributions. Penalizing this outlier can be expected to solve the problem of a single track with multiple objects. However we do not know the actual centre $E(X_t | Z^{t-1})$ of the real predicted distribution $p(X_t | Z^{t-1})$; instead it is approximated by the center $E(X'_t | Z^{t-1})$ of the distribution in Fig. 2a; i.e. the predicted distribution $p(X'_t | Z^{t-1})$ using the independent prediction model. Our prediction model is finally defined as

$$p(X_t | X_{t-1}, Z^{t-1}) \propto \Pi_{i} p(X_{t|t-1}) \cdot \Pi_{ij \in \mathcal{T}} \phi(X_{it}, X_{jt}) \cdot \psi(X_t, E(X'_t | Z^{t-1})).$$

$$E(X'_t | Z^{t-1}) = \int_{X'_t} X'_t \cdot p(X'_t | Z^{t-1})$$

is a conditional expectation of $X'_t$ with respect to the conditional probability distribution:

$$p(X'_t | Z^{t-1}) = \int_{X_{t-1}} \Pi_{i} p(X'_{it} | X_{(t-1)}) p(X_{t-1} | Z^{t-1}).$$

With $\psi(X_t, E(X'_t | Z^{t-1}))$, we penalize those predicted states $X_t$ far away from $E(X'_t | Z^{t-1})$ as outliers. Fig. 4 shows the predicted and posterior distributions when our prediction model is used. Comparing our predicted distribution with the one in Fig. 3a, we can see the peak at $[b, a]$ is penalized as an outlier, because it is far away from the center $E(X'_t | Z^{t-1})$ at $[a, b]$. Consequently the posterior distribution only peaks at $[a, b]$ as it should. To implement a penalty function, we define $\psi(X_t, E(X'_t | Z^{t-1}))$ as the Gibbs distribution:

$$\psi(X_t, E(X'_t | Z^{t-1})) \propto \exp(-\alpha \cdot \text{dist}(X_t, E(X'_t | Z^{t-1})))$$

$$\propto \Pi_{i} \exp(-\frac{\alpha \cdot \text{dist}(X_{it}, E(X'_{it} | Z^{t-1}))}{\sum_{j} \text{dist}(X_{it}, E(X'_{it} | Z^{t-1}))}).$$

Fig. 1. 2-object measurement model distribution $p(z_t | X_t)$.

Fig. 2. (a) The predicted distribution $p(X_t | Z^{t-1})$ in Eq. 1. (b) The posterior distribution $p(X_t | Z^t)$ obtained by multiplying the predicted distribution (left) with the measurement distribution in Fig. 1. The independent prediction model in Eq. 2 is used in this case.

Fig. 3. (a) The predicted distribution $p(X_t | Z^{t-1})$ in Eq. 1. (b) The posterior distribution $p(X_t | Z^t)$ obtained by multiplying the predicted distribution (left) with the measurement distribution in Fig. 1. The prediction model in MRF-based algorithm is used in this case.
Here, the distance between \( X_t \) and the expectation \( E(X_t | Z^{t-1}) \) is approximated as the sum of the distances between all the individual object states \( X_{it} \) and their corresponding individual state expectations \( E(X_{it} | Z^{t-1}) \). \( \psi(X_t, E(X_t | Z^{t-1})) \) encourages the predicted state \( X_t \) with all the individual object states \( X_{it} \) to be close to their corresponding expectations \( E(X_{it} | Z^{t-1}) \). This is consistent with the intuition that if \( X_{it} \) is far away from \( E(X_{it} | Z^{t-1}) \) and close to \( E(X_{it} | Z^{t-1}) \), then \( X_{it} \) is more probably tracking a wrong object \( j \) rather than the object \( i \) it should track, and this case should be penalized. In this way, each track is encouraged to concentrate on one object and is discouraged to get close to the objects being tracked by other tracks.

\[
p(X_t | Z^{t-1}) 
\]

(a) The predicted distribution \( p(X_t | Z^{t-1}) \) in Eq. 1. (b) The posterior distribution \( p(X_t | Z^t) \) obtained by multiplying the predicted distribution (left) with the measurement distribution in Fig. 1. Our prediction model is used in this case.

### 3. PARTICLE FILTER IMPLEMENTATION

The Bayesian sequential estimation with our new prediction model will now be implemented with a particle filter. Briefly, it is assumed that the posterior distribution at the previous time is approximated by a set of weighted particles \( p(X_{t-1} | Z^{t-1}) \approx \{X_{t-1}^{(r)}, w_{t-1}^{(r)}\}_{r=1}^{N} \). We then represent the posterior distribution at the current time as:

\[
p(X_t | Z^t) \approx k \cdot p(z_t | X_t) \cdot \sum_r w_{t-1}^{(r)} p(X_t | X_{t-1}^{(r)}, Z^{t-1}). \tag{4}
\]

We substitute our prediction model into Eq. 4 and rearrange terms that do not depend on the previous state \( X_{t-1} \) out of the mixture distribution. Then the posterior in the equation above becomes:

\[
p(X_t | Z^t) \approx k \cdot p(z_t | X_t) \cdot \Pi_{i \in T} \phi(X_{it}, X_{jt}) \\
\cdot \psi(X_t, E(X_t | Z^{t-1})) \cdot \sum_r w_{t-1}^{(r)} \Pi_i p(X_{it}|X_{i,t-1}^{(r)}).
\]

Now the filter can be simply implemented with the following two steps. First we sample from the proposal distribution

\[
X_t^{(s)} \sim q(X_t) = \sum_r w_{t-1}^{(r)} \Pi_i p(X_{it}|X_{i,t-1}^{(r)}).
\]

and then weight the samples according to the following likelihood:

\[
w_t^{(s)} = p(z_t | X_t^{(s)}) \cdot \Pi_{i \in T} \phi(X_{it}^{(s)}, X_{jt}^{(s)}) \cdot \psi(X_t^{(s)}, E(X_t^{(s)} | Z^{t-1})).
\]

Note \( X_t^{(s)} \), by definition, comes from the same distribution with \( q(X_t) \). Thus \( E(X_t^{(s)} | Z^{t-1}) \) can be simply calculated as the average of the samples drawn from the proposal distribution \( \sum_s X_t^{(s)} / N \).

### 4. EXPERIMENTAL RESULTS

In this section, our proposed algorithm is compared with three existing algorithms, Czyz’s algorithm [4], Maccormick’s algorithm [5] and MRF-based algorithm [6]. The results show that our algorithm has better performance than the three existing algorithms on tracking interacting objects.

In the experiments, we create a scenario where two objects move and intersect each other as shown in Fig. 5a. Each object is a square with the pattern as shown in Fig. 5b. The single-object measurement model is defined as

\[
p(z_{i,t} | X_{i,t}) = \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{\text{dist}(q_{i,t}, q^*)^2}{2\delta^2}\right).
\]

\( q_{i,t} \) is the intensity histogram computed from image \( z_{i,t} \), which is extracted from the object region specified by \( X_{i,t} \). \( \text{dist}(q_{i,t}, q^*) \) is the distance between the \( i \)th object’s histogram and the reference histogram \( q^* \). The left object in Fig. 5a moves to the right and the bottom object moves to the top. Both of them are at a constant speed of 3 pixels per frame plus a Gaussian noise. We generate 10 random sequences according to the above scenario so that two objects cross each other in 10 different cases. All the algorithms are then tested on these sequences. The performance metric is the sum of the distance error between two objects’ estimated positions and their groundtruth positions in each frame. For all the algorithms, two tracks are initialized on the two groundtruth positions in the first frame. Although this scenario of object interaction is simple, it is sufficient to illustrate the problems of interest, and compare the performances of the competing algorithms.

Fig. 5a shows the tracking results of Czyz’s algorithm for 10 runs. Before interaction around the 15th frame, identical to all the other algorithms, Czyz’s algorithm is able to track two objects without difficulty and the error is small. After interaction, however, the algorithm cannot keep correct tracks on the two objects, resulting in bigger errors after the 15th frame. The error comes from the problem of a single track with multiple objects. As shown in Fig. 5b, each track corresponding to the rectangles with the same appearance gets attracted by two objects wrongly. Therefore, the estimated position (represented by the dot) and the moving direction (represented by the arrow) of each object are clearly incorrect.

Fig. 7 shows the tracking results of Maccormick’s algorithm and MRF-based algorithm respectively. Although they use different
strategies from Czyz’s algorithm, their performances are similarly affected by object interaction. Therefore, they have the same drawbacks with Czyz’s algorithm.

Fig. 8a shows the results of our algorithm. Our proposed algorithm tracks two objects crossing each other successfully in all of the 10 sequences. Finally, Fig. 8b is the comparison of the mean error curves of all the four algorithms discussed above. It shows that our algorithm has a much smaller error than all the others. Therefore, our algorithm is more robust than the existing algorithms in dealing with the problem of tracking multiple interacting objects. Our algorithm is also tested on a real video; due to the limited space please refer to the result at http://www.cs.ualberta.ca/~zhijie/space/twoopocansresult.mp4.

5. CONCLUSION

This paper deals with the problem of tracking multiple interacting objects. By analyzing the problems that may arise when objects interact, we propose a new prediction model within a Bayesian sequential estimation framework. As shown in the experiments, our algorithm can solve both the problems of a single object with multiple tracks and a single track with multiple objects. Experiments on synthetic image sequences show that our algorithm has smaller error when tracking multiple interacting objects than three existing algorithms. Therefore, the final conclusion is our proposed algorithm can give better performance than the existing algorithms when tracking multiple objects with interactions.

6. REFERENCES