A Probabilistic Framework for Multitarget Tracking with Mutual Occlusions

Menglong Yang*†, Yiguang Liu*, Longyin Wen†, Zhisheng You*, Stan Z. Li†
* Key Laboratory of Fundamental Synthetic Vision Graphics and Image for National Defense,
School of Aeronautics and Astronautics & Computer Science, Sichuan University, China
† Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
steinbeck@163.com, lygpapers@aliyun.cn, lywen@cbsr.ia.ac.cn,
zsyou@mail.sc.cninfo.net, szli@cbsr.ia.ac.cn

Abstract

Mutual occlusions among targets can cause track loss or target position deviation, because the observation likelihood of an occluded target may vanish even when we have the estimated location of the target. This paper presents a novel probability framework for multitarget tracking with mutual occlusions. The primary contribution of this work is the introduction of a vectorial **occlusion variable** *as part of the solution. The occlusion variable describes occlusion states of the targets. This forms the basis of the proposed probability framework, with the following further contributions: 1)* **Likelihood***: A new observation likelihood model is presented, in which the likelihood of an occluded target is computed by referring to both of the occluded and occluding targets. 2)* **Priori***: Markov random field (MRF) is used to model the occlusion priori such that less likely "circular" or "cascading" types of occlusions have lower priori probabilities. Both the occlusion priori and the motion priori take into consideration the state of occlusion. 3)* **Optimization***: A realtime RJMCMC-based algorithm with a new move type called "occlusion state update" is presented. Experimental results show that the proposed framework can handle occlusions well, even including long-duration full occlusions, which may cause tracking failures in the traditional methods.*

1. Introduction

Multitarget tracking is one of the most fundamental and important problems in computer vision. It has been researched for a long time, but it is still a challenging task in some complex scenarios. For example, when there are occlusions among targets, especially full occlusions, many traditional methods may fail because the occluded targets could not be well observed, which is the main problem addressed in this work. We propose a novel framework for multitarget tracking, which can handle mutual occlusions, no matter which are slight occlusions, heavy occlusions or even long-duration full occlusions.

1.1. Contributions of This Work

In summary, this paper presents a novel occlusion handling multitarget tracking framework with the following contributions.

The first contribution is the novel probabilistic framework for multitarget tracking, because of the introduction of "occlusion variable" (section 2). By using the occlusion variable, we can expediently describe the occlusion relationships among targets and observe all the targets in a more reasonable way even if some of them are occluded. We present a new appearance observation likelihood model (section 3), in which the appearance similarity of an occluded target is calculated by comparing the observation feature with both occluded and occluding targets, rather than only the feature of the occluded target.

The second contribution is the design of the priori models (section 4). The newly priori model called "occlusion priori model" is modeled by an MRF model, which eliminates the possibility of occlusion between far-away targets and restrains the uncommon occlusion relationships among the targets. In addition, the motion priori model is modified in section 4.2. The occlusion variable is incorporated by placing the occluded targets in lower priori probabilities.

The final contribution is the real-time RJMCMC algorithm of the proposed tracking framework (section 5), which includes a totally new move type called "$occlusion\ state\ update$" to sample occlusion relationships among targets.

1.2. Related Work

Our approach is based on the well-known MCMC particle filter [9], which could successfully track up to 20

targets. Traditionally, a probabilistic graphical framework called "recursive Bayesian inference" is adopted in (MCMC) particle filters, as shown in fig. 3(a). In the recursive Bayesian inference, the primary task for multitarget tracking is to determine the posteriori distribution $P(X_t \mid Z^t)$ over the current joint configuration of the targets $X_t \triangleq \{\mathbf{x}_{it}\}$ at the current time step t, given all observations $Z^t \triangleq \{Z_1, Z_2, \ldots, Z_t\}$ up to that time step.

$$P(X_t \mid Z^t) = cP(Z_t \mid X_t) \times \int P(X_t \mid X_{t-1}) P(X_{t-1} \mid Z^{t-1}) dX_{t-1} \quad (1)$$

where $c = 1/P(Z_t \mid Z^{t-1})$ is a normalization constant, $P(Z_t \mid X_t)$ is the likelihood of the observation, and the motion priori $P(X_t \mid X_{t-1})$ predicts the state X_t given the last state X_{t-1}. In fact, eq. (1) could be regarded as two steps: 1) Prediction: $P(X_t \mid Z^{t-1}) = \int P(X_t \mid X_{t-1}) P(X_{t-1} \mid Z^{t-1}) dX_{t-1}$; 2) Verification: $P(X_t \mid Z^t) = cP(Z_t \mid X_t) P(X_t \mid Z^{t-1})$. When an target is heavily occluded, especially fully occluded, the verification step may be not reliable, because the observation likelihood $P(Z_t \mid X_t)$ for an occluded target may drastically decrease even if the estimated location is accurate, as shown in fig. 1.

1.2.1 Particle Filter and MCMC PF

There are many approaches based on sequential importance resampling (SIR), i.e. the well-known particle filters (PF), having been proposed to solve the tracking problem [7, 11, 17]. In these approaches, the posteriori $P(X_t \mid Z^t)$ is approximated by a set of N weighted particles, i.e. $P(X_t \mid Z^t) \approx \{X_t^{(r)}, w_t^{(r)}\}$, where $w_t^{(r)}$ is the weight of the rth particle. The drawback of SIR is that it is hard to track more than 2 or 3 interacting targets in real applications, because the number of the required particles grows exponentially as a function of the state-space dimension[9].

Markov Chain Monte Carlo Particle Filters (MCMC for short in this work) have been shown to successfully track multiple targets in real time, and the required number of particles for MCMC tracking is only a linear function of the number of tracked targets when they do not interact [2, 9]. Different from SIR, the posteriori in MCMC methods is approximated by a series of unweighted samples, i.e., $P(X_t \mid Z^t) \approx \{X_t^{(r)}\}$.

1.2.2 Occlusion Handling

Recently, occlusion handling in multitarget tracking approaches can be mainly divided into two classes, according to whether the occlusion relationships or depth ordering of the targets is inferred.

The first class of solutions are called "*implicit*", in which the occlusion relationships among targets are not inferred explicitly. Some approaches try to construct robust appearance features and the classifiers[11, 17]. In fact, these methods are sometimes available because of two factors. One is that the partial features of occluded targets are well used, which may lose its ability in the case of full occlusions. The other is the online learning of target features, which may fail when heavy occlusion happens, especially long-duration heavy occlusion, because it may learn a false object for an occluded target.

The second class of solutions are called "*explicit*", in which the occlusion relationships among targets or depth ordering of targets is inferred explicitly. Some approaches rely on the spatial structure of the scene [19]. Thus these methods are difficult to generalize. In [3], video sequences are automatically decomposed into constituent layers sorted by depths by combining spatial information with temporal occlusions. But it is hard to handle a long-duration occlusion. Senior et al. [16] use appearance models to localize the targets and use disputed pixels to resolve their depth ordering during occlusions. However, it is feasible only in the scenarios without full occlusions, because it is difficult to localize a fully occluded target.

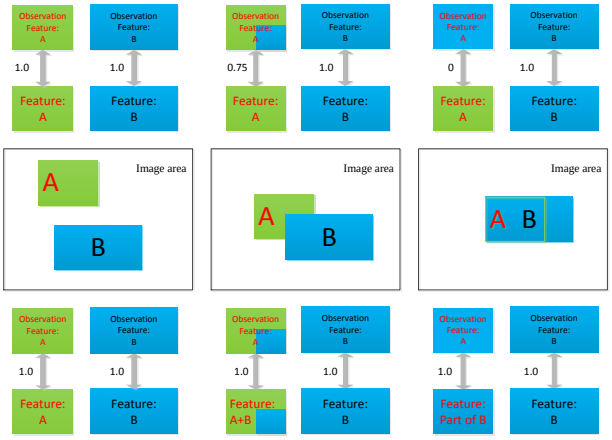


Figure 1. An instance of occlusions between two targets is shown in the middle row. The target A is unoccluded initially and fully occluded by B finally. The top row shows the measurement of observation likelihood in traditional methods. The observation likelihood of A may become not reliable owe to occlusions, even the estimated location of A is accurate. The bottom row shows the idea of measuring observation likelihood in this work. If a target A is estimated to be partially occluded by B, its observation likelihood should be measured by comparing the observation feature with the fusion feature of A and B, rather than only the feature of A. If the target A is estimated to be fully occluded by B, its observation likelihood should be measured by comparing the observation feature with the feature of the area which B occludes A. Such a measurement can ensure that the observation likelihood of an occluded target does not drastically decrease if its estimated location is accurate.

Note there are another class of "explicitly" methods, i.e. data association based methods, which are of offline methods, e.g., [1, 6, 18]. They can handle occlusions well, but the targets' trajectories can be got only when all frames of the whole video (at least a fragment) have been analyzed. In contrast, this work is of the online method, which gets the targets positions immediately when a new frame comes. Besides, the speeds of offline algorithms are often slow[1] and cannot satisfy the real-time requirements of some applications, thus they are ignored in this paper.

2. A New Multitarget Tracking Framework

2.1. Occlusion Variable

The proposed approach belongs to the class of "explicitly" handling occlusions. In this work, a vectorial variable $O_t \triangleq \{\mathbf{o}_{it}\}$ expressing the occlusion relationships among the multiple targets is introduced, which is named as "occlusion variable". In this paper, a target A is considered to be occluded by another target B, only if we only see the part of B rather than A in the occluded area [2], as shown in fig. 2. Combining X_t with O_t, the new configuration of multiple targets is denoted as $\{X_t, O_t\}$. Herein the occlusion variable is defined as follows:

$$\mathbf{o}_{it} = \begin{cases} i, & \text{if target } i \text{ is not occluded,} \\ j, j \neq i, & \text{if target } i \text{ is occluded by } j. \end{cases} \quad (2)$$

2.2. The Probabilistic Framework

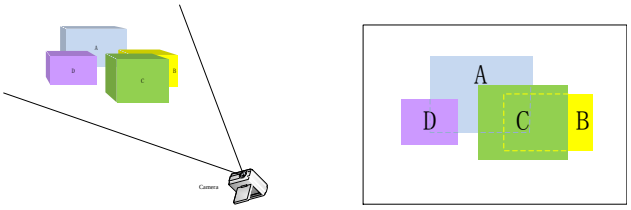
As a result of adding the occlusion variable, we adopt a novel probabilistic framework as shown in fig. 3(b). In this framework, we suppose that the priori probability of the occlusion states is independent of previous occlusion states O_{t-1}, i.e. it is only determined by the current state X_t[3]. According to the probabilistic framework, the posteriori $P(X_t, O_t \mid Z^t)$ is expressed as:

$$P(X_t, O_t \mid Z^t) = cP(Z_t \mid X_t, O_t)P(X_t, O_t \mid Z^{t-1}) \quad (3)$$

[1]We tested two offline trackers, i.e. Discrete-Continuous Optimization algorithm [1] and GMCP-Tracker [18]. The speeds of both algorithms are less than 10 fps, which are much slower than our algorithm.

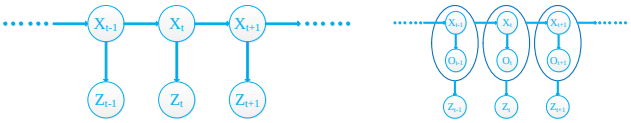
[2]The mutual occlusion relationship of the targets can be totally described by a boolean matrix. In this work, we adopt a vector to label each target is occluded by which one or un-occluded. This simplification greatly reduce the computational cost and it is enough for our most applications. In fact, a single occlusion variable cannot describe the occlusion of one target by 2 or more other targets. However, at each frame the proposed algorithm get a lot of samples with different posteriori probabilities, which may assign different (2 or more) occluders to an occluded target.

[3]Temporal independence of occlusion variable is also a practical simplification in order to decrease the calculation cost. In fact, this assumption has few infection to the tracking results. Generally, the inappropriate occlusion relationship gets a low posteriori likelihood, because the observation likelihood may be very small. In addition, when the video frame rate is low or the speed of the targets are large, there is little correlation between the occlusion states in different frames.



(a) (b)

Figure 2. (a) An instance of occlusions, where target A is occluded by B, C and D, and B is occluded by C. (b) is the projection image in the camera. Note there is only a part of C being seen in the lower right area of A, therefore in the opinion of this work, the lower right of A could be considered to be occluded not by B, but by C. In this case, the occlusion variable of A may be $\mathbf{o}_A = A$ (not occluded), $\mathbf{o}_A = C$ (occluded by C) or $\mathbf{o}_A = D$ (occluded by D), with different posteriori probabilities. The occlusion variable of B may be $\mathbf{o}_B = B$ or $\mathbf{o}_B = C$, while the occlusion variables of C and D could only be $\mathbf{o}_C = C$ and $\mathbf{o}_D = D$, respectively.



(a) (b)

Figure 3. (a) Recursive Bayesian Inference. (b) The proposed probabilistic graphical framework. Here O_t is the "occlusion variable" at time step t.

where

$$P(X_t, O_t \mid Z^{t-1}) = P(O_t \mid X_t) \int\int P(X_t \mid X_{t-1}, O_{t-1}) \times P(X_{t-1}, O_{t-1} \mid Z^{t-1}) dX_{t-1} dO_{t-1} \quad (4)$$

We can see that the observation likelihood model $P(Z_t \mid X_t, O_t)$ incorporates the occlusion variable, which will be presented in section 3. Compared with eq. (1), there is a new term in eq. (4), i.e. the occlusion priori model $P(O_t \mid X_t)$. Similar as the observation likelihood model, the motion priori model $P(X_t \mid X_{t-1}, O_{t-1})$ incorporates the occlusion variable. These two priori models will be presented in detail in section 4.

3. Observation likelihood Model

In this paper, two kind of observation likelihoods are taken into account, i.e. Global Mask Similarity and Appearance Similarity. The observation likelihood model is defined as

$$P(Z_t \mid X_t, O_t) = \mathcal{M}(X_t)\mathcal{S}(X_t, O_t) \quad (5)$$

where $\mathcal{M}$ is the Global Mask Similarity defined in [2], and $\mathcal{S}$ is the Appearance Similarity incorporating occlusion relationships, which is defined as

$$\mathcal{S}(X_t, O_t) = P(\eta) \prod_i \Theta(\mathbf{x}_{it}, \mathbf{o}_{it}) \quad (6)$$

where Θ is the appearance similarity of a single target. $P(\eta)$ is a penalty factor for invisible parts of the targets. The variable η is the proportion of total invisible area and it depends on X_t and O_t. Here invisible parts mean the target parts that are occluded by some other targets or out of the camera region. The penalty factor is a monotonic decreasing function, and it aims at expressing the thought of "seeing is believing". In this paper, it is simply defined as $P(\eta) = \frac{1}{1+\eta}$.

Due to introduction of the occlusion variable, the appearance model of single target is different from other methods, which is defined as

$$\Theta(\mathbf{x}_{it}, \mathbf{o}_{it}) = \begin{cases} \pi(z_{it}, z_i), & \text{if } \mathbf{o}_{it} = i \\ \pi(z_{it}, \mathcal{F}_{ij}), & \text{if } \mathbf{o}_{it} = j \text{ and } j \neq i \end{cases} \quad (7)$$

where z_i is the storage image feature vector of the target i, such as the well known RGB or HSV histogram, LBP, HOG feature etc. z_{it} denotes the image observation of target i, at the time step t. π is a similarity function measures the similarity between two features. $\mathcal{F}_{ij}$ is the *fusion feature* of the targets i and j. For example, as fig. 1 shows, if $\mathbf{o}_A = B$ (target A is occluded by B), the appearance similarity of A should be calculated by $\Theta(\mathbf{x}_A, \mathbf{o}_A) = \pi(z_A, \mathcal{F}_{AB})$.

In this paper, HSV histogram is used as the target feature, and Bhattacharyya coefficient is used to measure the similarity. The fusion feature of targets i and j is simply calculated as $\mathcal{F}_{ij} = (1 - r_{ij})z_i + r_{ij}z_j$, where r_{ij} is the "degree of the occlusion" defined as the proportion of the area occluded by j to the total area of the target i.

4. Priori Models

4.1. An MRF Occlusion Priori Model

At each frame, to model the occlusion relationships among the multiple targets, we dynamically construct an MRF that addresses the possible occlusions between nearby targets. Specifically, each target could be regarded as a node and the links between the nodes determine whether occlusion may occurs between each pair of targets, as shown in fig. 4.

In this work, a pairwise MRF is adopted, where the cliques are restricted to the pairs of nodes that are directly connected in the graph. The joint probability of occlusion state O_t is defined as the following model, where the $\psi(\mathbf{o}_{it}, \mathbf{o}_{jt})$ are pairwise interaction potentials:

$$P(O_t \mid X_t) \propto \prod_i P(\mathbf{o}_{it} \mid X_t) \prod_{ij \in \mathcal{N}} \psi(\mathbf{o}_{it}, \mathbf{o}_{jt}) \quad (8)$$

Here the $P(\mathbf{o}_{it} \mid X_t)$ is the priori probability of target i either being not occluded ($\mathbf{o}_{it} = i$) or being occluded by another target ($\mathbf{o}_{it} \neq i$), without considering the occlusion relationships of the rest targets. In this paper, a simple model

Figure 4. To model occlusion relationships, an MRF is constructed at each frame, with edges for close target couples, such as the target couples $(1, 2)$, $(2, 3)$ and $(5, 6)$. The far-away target couples are not linked, e.g. $(1, 3)$, $(3, 4)$ etc., indicating that it is impossible to exist occlusions between them. The ellipses indicate that predicted possible positions of the targets.

is adopted for this individual probability, shown as following.

$$P(\mathbf{o}_{it} = j \mid X_t) = \begin{cases} 0, & j \notin \hat{\mathcal{N}}_i, \\ \frac{\Phi(\mathbf{x}_{it}, \mathbf{x}_{jt})}{\sum_{k \in \hat{\mathcal{N}}_i} \Phi(\mathbf{x}_{it}, \mathbf{x}_{kt})}, & j \in \hat{\mathcal{N}}_i, \end{cases} \quad (9)$$

and

$$\Phi(\mathbf{x}_{it}, \mathbf{x}_{jt}) = \begin{cases} 1, & \text{if } i = j \text{ or the targets } i \text{ and } j \text{ overlap,} \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

where $\hat{\mathcal{N}}_i = \mathcal{N}_i \bigcup \{i\}$, N_i is the target set connecting with the target i in MRF.

Using this form, when a target is overlapped with some other targets, the priori probability of this target being not occluded or being occluded by any other is identical. Intuitively, for the overlapped targets, they have the same opportunity to become the "most front" (closest to the camera) one.

In the actual scenarios, the "*circular*" relationships are rare, e.g., target a and b occlude each other, or target a occludes b, b occludes c and c occludes a. Furthermore, when the occlusion relationship of the targets is "*cascading*" as the A, B and C in fig. 2, the bottom targets A and B are often both considered being occluded by the most front one, i.e. C. Namely, in the opinion of this paper, the priori probability for an occluded target occluding another one is low. This can be ensured by using an MRF interaction potential. It is defined as

$$\psi(\mathbf{o}_{it}, \mathbf{o}_{jt}) = \begin{cases} \psi_0, & \mathbf{o}_{it} = j \text{ and } \mathbf{o}_{jt} \neq j, \\ 1, & \text{otherwise.} \end{cases} \quad (11)$$

where $\psi_0 < 1$. In the experiment, we empirically select $\psi_0 = 0.1$.

4.2. Motion Priori Model

In the motion priori model, a common form except for incorporating the occlusion variables is adopted, which is

shown as follows:

$$P(X_t \mid X_{t-1}, O_{t-1}) = \prod_i P(\mathbf{x}_{it} \mid \mathbf{x}_{i(t-1)}, \mathbf{o}_{i(t-1)}) \quad (12)$$

Gaussian distribution is often adopted as the motion priori model of a single target. In this paper, Discretized Wiener velocity model [15] is used, i.e.,

$$P(\mathbf{x}_{it} \mid \mathbf{x}_{i(t-1)}, \mathbf{o}_{i(t-1)}) \sim N(\mathbf{x}_{it} \mid A\mathbf{x}_{i(t-1)}, q\Sigma) \quad (13)$$

where A is a constant transmission matrix of the motion priori model, q is the diffusion coefficient and Σ is a constant symmetry matrix. The detailed form of A and Σ is referred to [15]. Different diffusion coefficients result different noise covariance matrices. We set a larger diffusion coefficient if the target was occluded at last time step, which means that the transmission are less reliable when the target was occluded, i.e.,

$$q = \begin{cases} q_1, & \text{if } \mathbf{o}_{i(t-1)} = i, \\ q_2, & \text{otherwise.} \end{cases} \quad (14)$$

where $q_2 > q_1$. In fact, the thought of "seeing is believing" is expressed again.

5. Multitarget Tracking Algorithm

The proposed algorithm is an extension of the method in [9]. The posteriori $P(X_{t-1}, O_{t-1} \mid Z_{t-1})$ at time step $t-1$ is approximated as a set of N samples $P(X_{t-1}, O_{t-1} \mid Z_{t-1}) \approx \{X_{t-1}^{(r)}, O_{t-1}^{(r)}\}_{r=1}^{N}$. The MCMC approximation of (3) can be obtained as following equation.

$$P(X_t, O_t \mid Z^t) \approx cP(Z_t \mid X_t, O_t) \prod_i P(\mathbf{o}_{it} \mid X_t) \times \prod_{ij \in \mathcal{N}} \psi(\mathbf{o}_{it}, \mathbf{o}_{jt}) \sum_{r=1}^{N} P\left(X_t \mid X_{i(t-1)}^{(r)}, O_{i(t-1)}^{(r)}\right) \quad (15)$$

In this paper, we adopt Metropolis-Hastings algorithm to approximate the posteriori distribution.

5.1. Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm is the most popular MCMC method [5, 13]. An iteration step in the MH algorithm involves sampling a candidate value x' given the current value x according to a proposal distribution $q(x' \mid x)$. The Markov Chain then moves towards x' with the acceptance probability $\mathcal{A}$. Otherwise it remains at x. More details could be seen in [5, 9, 13].

The MH algorithm adopted in this work is a "Reversible Jump" MCMC (RJMCMC) version for handling a variable number of targets. In RJMCMC, the MH algorithm starts the chain in an arbitrary configuration and then selects a move type m from a finite defined set of moves that may change the dimensionality of the state. A move that changes the dimensionality of the state is referred to as a "jump". Each jump has a corresponding reversed jump defined, e.g., a target adding move should have a corresponding target deleting move. In other words, if it can jump from state A to state B, it must be able to reversed jump from state B to state A. After selecting the move type and proposing the candidate configuration, the acceptance ratio is defined as follows

$$\mathcal{A} = \min\left\{1, \frac{p(x')}{p(x^i)} \frac{p_{m'}}{p_m} \frac{q_{m'}(x^i \mid x')}{q_m(x' \mid x^i)}\right\} \quad (16)$$

where m and m' is the reversed jump move to each other, and p_m and $p_{m'}$ are the corresponding proposal densities of the move types m and m', respectively.

In the present case, the posteriori distribution $P(X_t, O_t \mid Z_t)$ could be approximated, in which the acceptance ratio should be calculated as follows.

$$\mathcal{A} = \frac{P(Z_t|X_t', O_t')}{P(Z_t|X_t, O_t)} \frac{\prod_i P(\mathbf{o}_{it}'|X_t')}{\prod_i P(\mathbf{o}_{it}|X_t)} \frac{\prod_{ij \in \mathcal{N}} \psi(\mathbf{o}_{it}', \mathbf{o}_{jt}')}{\prod_{ij \in \mathcal{N}} \psi(\mathbf{o}_{it}, \mathbf{o}_{jt})} \times \frac{\sum_r P\left(X_t' | X_{t-1}^{(r)}, O_{t-1}^{(r)}\right)}{\sum_r P\left(X_t | X_{t-1}^{(r)}, O_{t-1}^{(r)}\right)} \frac{p_{m'}}{p_m} \frac{Q(X_t, O_t | X_t', O_t')}{Q(X_t', O_t' | X_t, O_t)} \quad (17)$$

5.2. Data Driven Proposals

At each frame, a target detector could detect out some targets by the target detectors to provide a set of targets with noises can be used to update the current target set X_t. Different strategies may be used for the specific applications. In this paper, we adopt the background subtraction described in [12] as the target detector.

There are a total of six move types in the proposed algorithm, i.e. $\mathcal{M} = \{\ Add, Delete, Stay, Leave, Perturb, Occlusion\ state\ update\ \}$. The first five models are similar as [9] except the computation of the acceptance ratio (17). The new move type "$Occlusion\ state\ update$" aims at drawing the occlusion relationships among the targets.

At each step, only one target's occlusion state could be changed. The proposal is done as follows: randomly select a target $\mathbf{x}_o$ from the target set $\{\mathbf{x}_{it}\}_{i \in \hat{\mathcal{N}}} \bigcap \{\mathbf{x}_{jt}\}_{\Phi(\mathbf{x}_{it}, \mathbf{x}_{jt})=1} \bigcap X_t$, that denotes all the targets, each of which has been already added to X_t, and there are some targets "near" it in MRF, as well overlapping with it. The definition of Φ is referred to (10). Then randomly sample its occlusion state $\mathbf{o}_o'$ from $\{\mathbf{x}_{it}\}_{i \in \hat{\mathcal{N}}_o} \bigcap \{\mathbf{x}_{jt}\}_{\Phi(\mathbf{x}_o, \mathbf{x}_{jt})=1} \bigcap X_t$, where $\{\mathbf{x}_{it}\}_{i \in \hat{\mathcal{N}}_o} \bigcap \{\mathbf{x}_{jt}\}_{\Phi(\mathbf{x}_o, \mathbf{x}_{jt})=1}$ denotes all the targets "near" $\mathbf{x}_o$ in MRF, overlap with $\mathbf{x}_o$, and have already been added to current sampler X_t.

6. Experimental Results

We first evaluated the occlusion module of our approach by tracking a long-duration full occluded target and comparing the proposed framework with traditional framework.

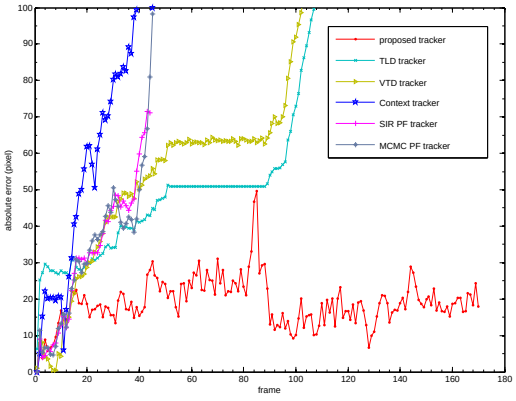


Figure 6. Tracking errors of the occluded target.

Then a real-time multitarget tracking system was demonstrated in some real outdoor scenarios under the platform of Mac OS X 10.7, 2.4GHz Intel Core 2 Duo CPU, 4GB RAM without GPU or any other parallel computing device.

6.1. Evaluation of Occlusion Module

We evaluated the availability of the occlusion module in an indoor environment (fig. 5), under Matlab R2011b. The tracking results are evaluated in comparison with several state-of-art trackers with default parameters in their articles, including VTD [10], Context [4] and TLD [8], besides the SIR PF [14] and MCMC [2] [4] under the traditional framework. In this experiment, the black card was occluded by the white one for a long duration (exceeds 100 frames). The proposed method can not only track the two targets, but also indicate the occlusion relationship and occlusion degree between them.

The tracking results of occluded target are compared in fig. 6 and Table 1. We stored the tracked center locations of target 2, and calculated absolute error in each frame and the mean error in a certain interval frames before the target is failed tracked. From the figure, we can see the tracking errors were close before the target was occluded (before 14th frame). During the target was occluded, the absolute errors of the proposed method almost kept within a certain range. However, the performances of other methods tobogganed from the frame when the target started to be occluded. The results exhibited the effect of the occlusion module, which can be seen from the table.

6.2. A Real-time Tracking System

We demonstrated the results of the proposed method for tracking multiple targets in three outdoor scenarios, including a campus scenario of the public dataset PETS2001, a real surveillance scenario of a street and a more complicated scenario of a square, as shown as fig.7. There were some occlusions in all the three videos. For example, in the second video, the target 270 successively interacted with 266, 264, 265 and 269. As fig.7(b) shows, there were tracking failures for the interacting targets using traditional methods. During a target was occluded, its appearance feature was interfused much appearance information of occluding target when updating the appearance features of targets. However, the proposed tracker knew the occlusion relationship between the interacting targets, thus it can avoid leaning false information when updating appearance features (in this work, the appearance feature did not update when a target was occluded). Therefore the proposed algorithm successfully tracked all the targets during the whole process of their interactions.

[4] In this experiment, since the targets are manually labeled at first frame rather than detected by object detector, in both the proposed tracker and RJMCMC tracker [2], we set $\mathcal{M}(X_t) = 1$ in (5).

Table 2. Processing Capability of the System

Number of Samples	Mean Frame Rate of the System (fps)	Mean Frame Rate of the Tracker (fps)
100	76.07	243.32
200	73.52	211.18
500	64.59	151.62
1000	53.28	102.54
1500	45.74	79.06
2000	39.82	62.54
3000	31.69	44.76

Then the processing capability of the proposed system was evaluated on the first scenario (320×240), using a 100 to 3000 particles chain. Table 2 shows the mean frame rates for different number of samples. The last column shows the processing capability of the tracker without target detection, while the middle column is the processing capability of the whole system including the target detection [12]. The system can be run in real time even if the number of samples achieves 3000.

6.3. Failures

Of course, the proposed method can not successfully work for all scenarios. In fact, only the simple HSV histogram and the Bhattacharyya coefficient were respectively adopted as the appearance feature and the measurement of appearance similarity, a scenario with drastic illumination and pose variations is hard to handle. In addition, the occlusion studied in this paper is the mutual occlusion among targets. The targets occluded by the background may cause tracking failures. Fig.8 shows some examples of the tracking failures.

7. Conclusions

This paper presented a novel probabilistic framework for multitarget tracking, which could handle mutual occlusions among targets. By introducing a vectorial variable called "occlusion variable", the occlusion state of each target can be expressed. In the new framework with this "occlusion

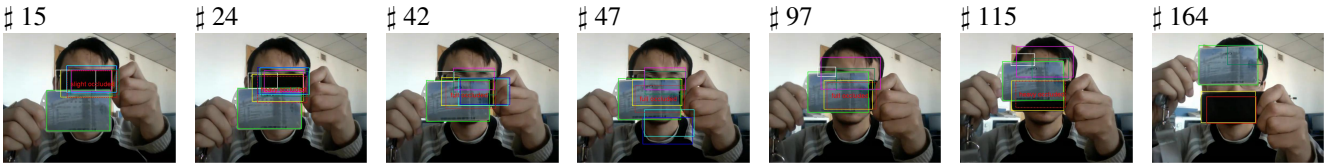
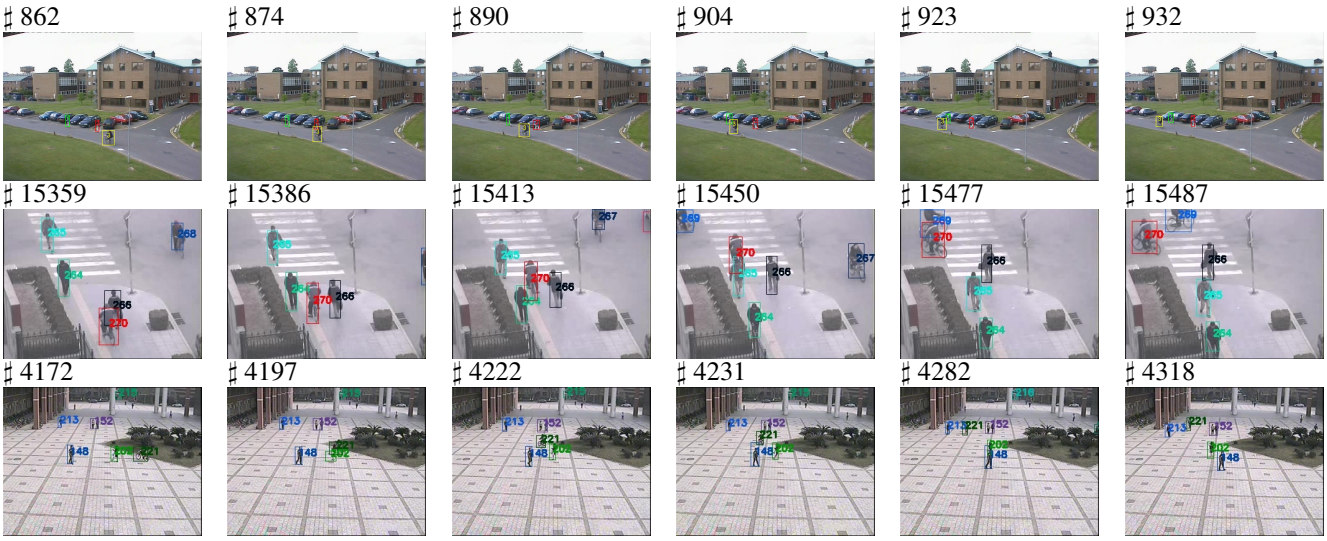


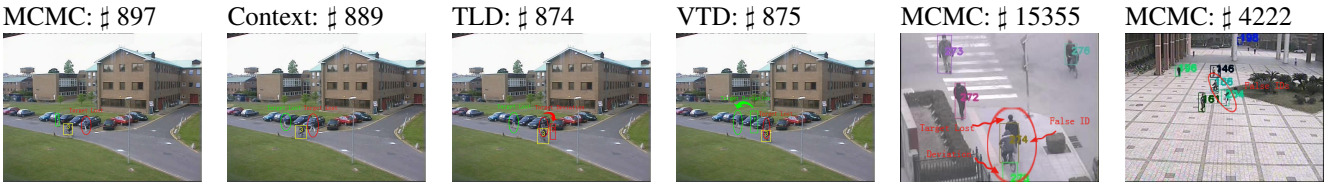
Figure 5. In the video (640×480), there are two targets need tracking, which are initial manually labeled at first frame. The green and yellow rectangles are the groundtruth of the two targets, respectively. The tracking results of occluding target are close to the groundtruth, thus we only show the tracking results of the occluded target. The results of the proposed tracker, MCMC [2], PF [14], TLD [8], VTD [10], Context [4] are depicted as red, cyan, blue, dark green, magenta, white rectangles respectively. The rectangles of dashes denote the target is considered to be occluded. The degree of the occlusion here is divided to three classes: slight occluded ($0.2 < r \leq 0.5$), heavy occluded ($0.5 < r \leq 0.8$) and full occluded ($r > 0.8$), where r is the ratio of the occluded target area.

Table 1. Mean error of the tracking results

Frames	$1 \sim 10$	$11 \sim 40$	$41 \sim 100$	$101 \sim 140$	$141 \sim 170$	$1 \sim 170$	total frames
Proposed	8.07	***17.74***	***23.02***	***15.96***	***19.7815***	***18.98***	***170***
MCMC	7.19	35.70	-	-	-	-	47
PF	6.83	36.96	-	-	-	-	76
Context	16.38	64.17	107.37	-	-	-	132
TLD	25.48	33.44	51.66	121.30	-	-	160
VTD	***3.60***	35.69	135.74	-	-	-	167



(a) The proposed method. Note there were some people being not tracked in the last row, because they were too small and they were regarded as noises by the target detector.



(b) Traditional methods

Figure 7. Tracking in the outdoor scenarios.

variable", the appearance observation model can be constructed to continuously compute the likelihoods of the occluded targets. The priori models were also designed to incorporate the occlusion variable. There is a new priori model called "occlusion priori model" being modeled using MRF. A real-time RJMCMC algorithm which can achieve the requirements of many applications was designed based on the proposed framework. Besides the real-time processing

♯ 1540 ♯ 318

Figure 8. Tracking Failures. The images are from the database of CAVIAR2004 and PETS2009, respectively.

capability, the main advantages of the proposed approach are summarized as follows.

The first advantage is the good performances in occlusion handling. Section 6.1 shows that the proposed tracking model is robust for full occlusion for a long duration, even though only the simple HSV histogram and Bhattacharyya coefficient were respectively adopted as the appearance feature and the measurement of appearance similarity.

The second advantage is brought by the "occlusion variable". At each video frame, we can know the degree of the occlusion of each occluded target from the tracking results, the appearance features of the targets can therefore be updated more flexibly.

In addition, the proposed multitarget tracking framework is general in that it caters for modification of any module in it. For example, if we adopt another appearance feature and similarity measurement (classifier), rather than the simple HSV histogram and Bhattacharyya coefficient, the observation likelihood model can be improved. It is hopeful that the more robust feature and classifier could bring better performance, which is our next work.

Acknowledgement

This work is partially supported by NSFC (No.61173182, 61375037 and 61179071), "863" Program (No.2011AA011804), Startup Foundation for Youth Scholars of Sichuan University (No.2013SCU11004), and the funding from Sichuan Province (No.2011JY0124, 2012HH0004, 2012HH0031 and 2012GZ0095).

References

[1] A. Andriyenko, K. Schindler, and S. Roth. Discrete-continuous optimization for multi-target tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1926–1933, 2012.

[2] F. Bardet, T. Chateau, and D. Ramadasan. Illumination aware MCMC particle filter for long-term outdoor multi-object simultaneous tracking and classification. In *IEEE International Conference on Computer Vision*, pages 1623–1630, 2009.

[3] G. J. Brostow and I. A. Essa. Motion based decompositing of video. In *IEEE International Conference on Computer Vision*, 2001.

[4] T. B. Dinh, N. Vo, and G. G. Medioni. Context tracker: Exploring supporters and distracters in unconstrained environments. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1177–1184, 2011.

[5] W. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97, 1970.

[6] C. Huang, Y. Li, and R. Nevatia. Multiple target tracking by learning-based hierarchical association of detection responses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(4):898–910, 2013.

[7] M. Isard and A. Blake. Condensation-conditional density propagation for visual tracking. *International journal of computer vision*, 29(1):5–28, 1998.

[8] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1409–1422, 2012.

[9] Z. Khan, T. Balch, and F. Dellaert. MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1805–1918, 2005.

[10] J. Kwon and K. M. Lee. Visual tracking decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1269–1276, 2010.

[11] M. Li, W. Chen, K. Huang, and T. Tan. Multi-Target Tracking by Learning Class-Specific and Instance-Specific Cues. In *Asian Conference on Computer Vision*, pages 67–81, 2011.

[12] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikainen, and S. Li. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1301–1306, 2010.

[13] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, et al. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087, 1953.

[14] K. Nummiaro, E. Koller-Meier, and L. Van Gool. An adaptive color-based particle filter. *Image and Vision Computing*, 21(1):99–110, 2003.

[15] S. Särkkä. *Recursive Bayesian inference on stochastic differential equations*. Helsinki University of Technology Laboratory of Computational Engineering Publications, 2006.

[16] A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle. Appearance models for occlusion handling. In *IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, 2001.

[17] J. Xing, H. Ai, and S. Lao. Multiple Human Tracking Based on Multi-view Upper-Body Detection and Discriminative Learning. In *IEEE International Conference on Pattern Recognition*, pages 1698–1701, 2010.

[18] A. R. Zamir, A. Dehghan, and M. Shah. Gmcp-tracker: Global multi-object tracking using generalized minimum clique graphs. In *European Conference on Computer Vision*, pages 343–356. Springer, 2012.

[19] Y. Zhou and H. Tao. A background layer model for object tracking through occlusion. In *IEEE International Conference on Computer Vision*, 2003.