Collaborative Tracking of Multiple Targets

Ting Yu and Ying Wu
Department of Electrical & Computer Engineering
Northwestern University
2145 Sheridan Road, Evanston, IL 60208
{tingyu,yingwu}@ece.northwestern.edu

Abstract

Coalescence, meaning the tracker associates more than one trajectories to some targets while loses track for others, is a challenging problem for visual tracking of multiple targets, especially when similar targets move close or present occlusions. Existing approaches that are based on joint data association are confronted by the combinatorial complexity due to the concatenation of the state spaces of individual targets. This paper presents a novel collaborative approach with linear complexity to the coalescence problem. The basic idea is the collaborative inference mechanism, in which the estimate of an individual target is not only determined by its own observation and dynamics, but also through the interaction and collaboration with the estimates of its adjacent targets, which leads to a competition mechanism that enables different targets to compete for the common image observations. The theoretical foundation of the new approach is based on Markov networks. Variational analysis of this Markov network reveals a mean field approximation to the posterior density of each target, therefore provides a computationally efficient way for such a difficult inference problem. In addition, a mean field Monte Carlo (MFMC) algorithm is designed to achieve Bayesian inference by simulating the competition among a set of low dimensional particle filters. Compared with the existing solutions, the proposed new collaborative approach stands out by its effectiveness and low computational cost to the coalescence problem, as pronounced in the extensive experiments.

1 Introduction

Multiple target tracking in video is an important problem in many emerging applications, such as for intelligent video surveillance where tracking multiple targets is essential for action recognition and event detection, for sports video analysis where tracking multiple athletes can help coaches for decision making and performance analysis, and for video conferencing where effective low bit rate video communication requires the accurate localization and tracking of the attendees.

If targets are distinctive from each other, they can be tracked independently by using multiple independent trackers (M.i.T.) with least confusion. However, many real application scenarios prevent such a simple solution for two reasons: the targets may present more or less the same appearances, and it is infeasible to initialize specific trackers for different targets. For example, it is difficult to track multiple soccer players in a soccer field, since all soccer players are in uniform sports wear, which makes image observations less discriminative. In this circumstance, the tracker has to use the same target model and observation model to handle all the targets. In this sense, the tracker has to cope with "identical" targets simultaneously.

Due to the use of one single model for tracking multiple similar targets, both M.i.T. and the CONDENSATION algorithm [7, 2] can not work well. Since each individual tracker in M.i.T. tends to track the target that fits the model best, the targets that receive weaker image evidence are likely to be ignored, especially when the targets move close or present occlusions. As a single target tracker, CONDENSATION may also be used for multiple targets, since it can estimate the non-Gaussian posterior density of the targets which implies the presence of multiple targets. However, when the posterior distribution is propagating over frames by particles, it is likely that the targets that attract more particles will dominate the dispersion of particles, which will gradually reduce the number of particles for the targets that have weaker image observations, and finally lose track of them. In both cases, we observe the "coalescence" phenomenon, i.e. the multiple target tracker associates more than one trajectories to some targets while loses track for others.

A possible solution to this coalescence problem is based on joint data association that enumerates all the possible associations between targets and observations. Various methods have been developed (see Section 2 for details), such as joint probabilistic data association filter (JPDAF) [1, 12], sampling based multiple target tracking with background subtraction [15] or background model [8] and partitioned sampling [11]. The essence of their methods is the introduc-

1063-6919/04 $20.00 © 2004 IEEE

tion of the joint state space representation which concatenates together all the state spaces of the individual targets such that they can be jointly inferred based on the collected image observations. The coalescence problem may be correctly handled during the joint inference. However, these approaches are not scalable due to their nature of exponential complexity. For example, with the increasing number of targets, JPDAF-based methods suffer from the combinatorial complexity due to the exhaustive enumeration for data associations; and sampling based approaches are confronted by the exponential demand of the increase of particles.

In this paper we propose a new tracking algorithm to cope with the coalescence problem with linear complexity. The basic idea is a collaborative inference mechanism, where the state estimate of each target is not only determined by its own observation and dynamics, but also through the interaction and collaboration with the state estimates of its adjacent targets, which leads to a competition mechanism that enables different targets to compete for the common resources, i.e. image observations. The theoretical foundation of the new approach is based on Markov networks, in which each hidden node in the network represents the state of an individual target, and the links in the network correlate a target to those who compete image observations against it. The structure of the Markov network can change according to the spatial relations of the targets during the tracking process. We call it an *ad hoc Markov network*. Since such a Markov network is likely to contain loops, variational analysis is employed and reveals a mean field approximation to the posteriors of the targets, therefore it provides a computationally efficient way to this difficult inference problem. In addition, we design a mean field Monte Carlo (MFMC) algorithm that efficiently implements this mean field inference by simulating the competition among a set of low dimensional particle filters.

With linear complexity in terms of the number of targets, the new approach cope with multiple target tracking in a distributed and collaborative fashion. The competition mechanism introduced by the collaborative inference mathematically incorporates the essence of joint data association where one single observation cannot support more than one target, therefore the coalescence problem can be naturally handled. Compared with the existing solutions, the new collaborative approach stands out by its effectiveness and low computational cost to the coalescence problem, as shown in the extensive experiments.

2 Related Work

Various multiple target tracking methods have been developed to handle the coalescence problem. They either employ background models [4, 15, 8] or not [12, 11]. The background models provide strong cues for target detection, and the extracted moving blobs greatly facilitate image observations for tracking. Recognizing the merging and splitting of the blobs is an important bottom-up clue to solve the coalescence problem. The limitation of these methods is the assumption of static cameras which make possible the modeling of backgrounds and thus confines its applications. In this paper, the proposed approach does not rely on background models, thus it can work for unknown and constantly changing backgrounds as shown in the experiments.

According to the implementation of tracking, the existing methods can be categorized into either parametric [1, 12] or non-parametric [15, 8, 5, 11]. The parametric methods extend Kalman filters to joint probabilistic data association filters (JPDAF) [1, 12] and handle the coalescence problem by the joint data association principle in which one image observation can only support a single target hypothesis and one target hypothesis can only occupy a single observation. Based on Monte Carlo techniques, non-parametric methods [15, 8, 5, 11] can obtain non-Gaussian Bayesian inference in a top-down process that generates and evaluates a large number of hypotheses. These methods generally handle the coalescence problem through the modeling of the priors in the joint state space of all the targets. However, the above approaches that deal with the joint state space directly are not scalable due to its nature of combinatorial or exponential complexity.

Different from the centralized joint state space representation, a distributed representation is proposed in this paper that leads to efficient solutions with linear complexity. Based on this new representation, a collaborative mean field Monte Carlo (MFMC) algorithm is proposed for multiple target tracking, in which a set of low dimensional particle filters compete against each other to solve the coalescence problem.

3 The Distributed Representation

We denote the state of an individual target by $\mathbf{x}_i$, the joint state by $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ for M targets, the image observation of $\mathbf{x}_i$ by $\mathbf{z}_i$, and the joint observation by $\mathbf{Z}$.

3.1 Conditional Dependency

When multiple targets move close or present occlusions, it is generally difficult to distinguish and segment these spatially adjacent targets from image observations, thus we can not simply factorize the joint image observation, i.e., $p(\mathbf{Z}) \neq \prod_i p(\mathbf{z}_i)$.

As a result, the image observations under this circumstance have to be treated as they are jointly produced by all these targets, i.e., we need to model the joint likelihood $p(\mathbf{Z}|\mathbf{x}_1, \ldots, \mathbf{x}_M)$. In this case, when the joint image observation is given, the posteriors of different targets are conditionally dependent, i.e.,

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_M|\mathbf{Z}) \neq \prod_i p(\mathbf{x}_i|\mathbf{Z}).$$

This conditional dependency of multiple targets is the root of the reason why M.i.T. and CONDENSATION can not cope with the coalescence problem. It also makes clear why the centralized methods that deal with the joint state space are confronted by the high dimensionality, since they have to model $p(\mathbf{Z}|\mathbf{X})$ as a centralized entity.

We present in the next sections a new distributed model to cope with this problem with linear complexity and a collaborative algorithm is developed for tracking multiple identical targets.

3.2 Our Formulation: *ad hoc* Markov Network

Since the motions of the multiple targets become dependent when they are spatially adjacent, we can consider to model the prior of the joint target states, i.e., $p(\mathbf{X})$. This prior can be very complicated due to the unknown correlations, but we can approximate it by a Gibbs distribution in general. Here we present a specific Gibbs model which leads to a theoretically plausible and practically efficient tracking algorithm.

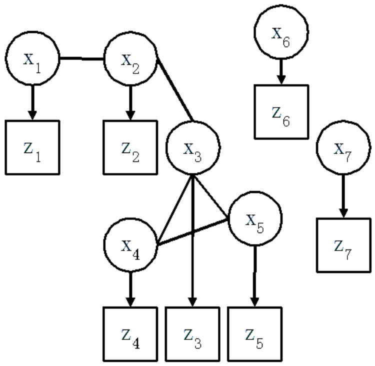

Figure 1: The Markov Network for multiple targets.

The theoretical foundation of the new approach is based on Markov networks, as shown in Figure 1, which consists of two layers. The hidden layer is an undirected graph $G_x = \{V, E\}$ where each node represents the state or motion parameters (such as an affine motion) of a target $\mathbf{x}_i$, and the link between a pair of targets represents the motion correlation (of dependency) between them (as described below). In addition, the observable layer are nodes that represent the image observations and are individually associated with their corresponding hidden nodes. A directed link from the target $\mathbf{x}_i$ to its local image observation $\mathbf{z}_i$ represents the observation likelihood $p(\mathbf{z}_i|\mathbf{x}_i)$. Since the local observation $\mathbf{z}_i$ conditionally independent of others given $\mathbf{x}_i$, we have :

$$p(\mathbf{Z}|\mathbf{X}) = \prod_{i=1}^{n} p_i(\mathbf{z_i}|\mathbf{x_i}). \tag{1}$$

The core problem here is to infer the posterior $p(\mathbf{X}|\mathbf{Z})$.

The structure of the graph in the hidden layer depends on the spatial relations among the targets' states. The target that is not close to others is represented by an isolated vertex in the graph (such as $\mathbf{x}_6$ and $\mathbf{x}_7$ in Figure 1) . If two targets are close enough (in the sense the the specific image observer or detector used for tracking is unable to separate their image observations), there is an undirected link between them in the graph to represent their motion dependency (such a $\mathbf{x}_3$ and $\mathbf{x}_4$ in Figure 1), and a potential function is associated with this link to parameterize the motion correlation.

Since the targets are moving, their spatial relations change with time and the structure of the Markov network also change with time. Therefore, we name this type of graphical model as *ad hoc* Markov Network. Once the spatial relations of the targets are roughly determined, the structure of the network is fixed. The neighborhood of a target is those that are linked with it, and we denote the neighborhood of $\mathbf{x}_i$ by $\mathcal{N}(i)$.

In this formulation, the prior $p(\mathbf{X})$ is modelled as a Gibbs distribution and can be factorized as:

$$p(\mathbf{X}) = \frac{1}{Z_c} \prod_{c \in \mathcal{C}} \psi_c(X_c) \tag{2}$$

where c is a clique in the set of cliques $\mathcal{C}$ in the undirected graph, X_c is the set of hidden nodes associated with the clique and $\psi_c(X_c)$ is the potential function of this clique, and Z_c is a normalization term or the partition function. Our model allows two types of cliques: the first order clique, i.e., $i \in V$, and second order clique, i.e., $(i,j) \in E$, where $\mathcal{C} = V \bigcup E$. The associated potential function ψ_c is denoted by ψ_i and ψ_{ij}, respectively. Thus, Eq. 2 can also be written as:

$$p(\mathbf{X}) = \frac{1}{Z_c} \prod_{(i,j) \in E} \psi_{ij}(\mathbf{x_i}, \mathbf{x_j}) \prod_{i \in V} \psi_i(\mathbf{x_i}) \tag{3}$$

where $\psi_i(\mathbf{x_i})$ provides a local prior for $\mathbf{x}_i$ which can be the dynamics prior or the prior given by other modalities, and $\psi_{ij}(\mathbf{x_i}, \mathbf{x_j})$ presents the motion dependency between neighborhood nodes $\mathbf{x}_i$ and $\mathbf{x}_j$.

It is critical to model the motion dependency or correlation mentioned above. The motion of two targets become dependent *a posteriori* only because their image observations can not be separated. But when one target has been associated with part of the total image observations, the other target can only obtain the rest of the observations, since the same piece of image evidence can not support the existence of two different targets. Therefore, we can approximate the motion dependency of them by a *competition* correlation, i.e., targets compete against each other for the common image resources. In other words, if one target occupies a region in the state space, it will lower the probability of others to occupy the same region. As a specific example, the competition potential function can be modelled as:

$$\psi_{ij}(\mathbf{x_i}, \mathbf{x_j}) \propto 1 - e^{-d(\mathbf{x}_i,\mathbf{x}_j)^T \Sigma^{-1} d(\mathbf{x}_i,\mathbf{x}_j)} \tag{4}$$

where $d(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i - \mathbf{x}_j$ is the distance between the two targets in the state space, and Σ characterizes the size of competition region in the state space. Eq. 4 is like an upside-down Gaussian, which reduces the probability of the events where two targets occupy the same position in the state space. When competing for image resources, the target that is unlikely to win will be diffused around the winner.

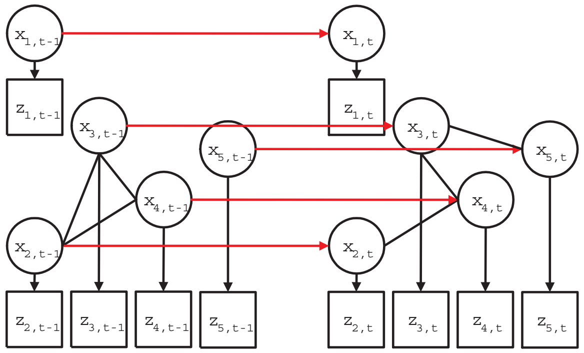

Figure 2: Dynamic Markov Network for multiple targets.

Putting the above Markov network in the temporal context by accommodating the dynamics model $p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1})$ for each target, we can model the visual dynamics of multiple targets in a more complicated graphical model, which can be called as a dynamic *ad hoc* Markov Network, as shown in Figure 2. In this figure, the structures of the Markov networks in two consecutive time frames are a little different, which illustrates the changes of motion correlations among the targets, due to the change of the spatial relations among them.

In all the notations, the subscript t represents the time index. In addition, we denote the collection of all the image observation up to time t by $\underline{\mathbf{Z}}_t = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_t\}$. In this formulation, the multiple target tracking problem is to infer the posterior of each target $p(\mathbf{x}_{i,t}|\underline{\mathbf{Z}}_t)$, which will be solved in the following sections.

4 Mean Field Inference

Belief propagation [3] is generally used to obtain exact Bayesian inference for non-loopy Markov networks. However, this method may not be appropriate for analyzing the Markov networks introduced in the previous section for multiple target tracking, because these Markov networks are likely to contain loops when three or more targets are linked together. In contrast to belief propagation, variational analysis methods [10, 9, 16] are more flexible to the structure of the network. Although only the approximate inference can be obtained, they provide lower bounds of the approximation as a theoretical benefit. Thus we perform variational analysis for the above Markov networks in this section. For clarity, we first analyze the static Markov network and then generalize the results to the dynamic Markov network.

The fundamental idea of the probabilistic variational method is the employment of a variational distribution $Q(\mathbf{X})$ with variational parameters as a variation of the density we want to infer, e.g., the posterior $p(\mathbf{X}|\mathbf{Z})$ in our case. Variational analysis aims at finding the optimal variational distribution $Q^*(\mathbf{X})$ that minimizes the Kullback-Leibler (KL) divergence between them, i.e.,

$$Q^*(\mathbf{X}) = \arg\min_Q KL(Q(\mathbf{X})||p(\mathbf{X}|\mathbf{Z})) \tag{5}$$

This is feasible when the appropriate forms of the variational densities are adopted. For simplicity, a fully factorized form is usually employed, i.e.,

$$Q(\mathbf{X}) = \prod_i^M Q_i(\mathbf{x}_i) \tag{6}$$

where $Q_i(\mathbf{x_i})$ is an independent distribution of the hidden node $\mathbf{x}_i$. Since Q_i has to be a probability density function, this becomes a constrained optimization problem with the following Lagrangian for each Q_i:

$$L(Q_i) = KL(Q_i) + \lambda(\int_{x_i} Q_i - 1) \tag{7}$$

When using the Gibbs model for $p(\mathbf{X})$ in Eq. 3, it is easy to show the solution is a set of fixed point equations [16]:

$$Q_i(\mathbf{x}_i) \longleftarrow \frac{1}{Z_i'} p_i(\mathbf{z}_i|\mathbf{x}_i)\psi_i(\mathbf{x}_i)M_i(\mathbf{x}_i), \qquad \text{where}$$

$$M_i(\mathbf{x}_i) = \exp\{\sum_{k\in\mathcal{N}(i)} \int_{x_k} Q_k(\mathbf{x}_k) \log \psi_{ik}(\mathbf{x}_i, \mathbf{x}_k)\}, \tag{8}$$

where Z_i' is a constant, and $\mathcal{N}(i)$ is the neighborhood of the subpart i, and $i = \{1, \ldots, M\}$. The iterative updating of $Q_i(\mathbf{x}_i)$ decreases the KL-divergence and reaches an equilibrium. These fixed point equations are called *mean field equations*.

The same procedure can also be applied to the dynamic Markov network, and the mean field equations can be derived:

$$\begin{aligned} Q_{i,t}(\mathbf{x}_{i,t}) \quad &\longleftarrow \quad \frac{1}{Z_i'} p_i(\mathbf{z}_{i,t}|\mathbf{x}_{i,t}) \\ &\times \quad \int p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1}) Q_{i,t-1}(\mathbf{x}_i) \\ &\times \quad M_{i,t}(\mathbf{x}_{i,t}) \end{aligned} \qquad (9)$$

where the second term $\int p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1})Q_{i,t-1}(\mathbf{x}_i)$ is actually very similar to the dynamics prediction prior, i.e., $p(\mathbf{x}_{i,t}|\underline{\mathbf{Z}}_{t-1})$.

The mean field equations are very meaningful, since they reveal a collaborative solution to the very difficult Bayesian inference problem: the posterior of a target $\mathbf{x}_i$ is not only determined by its local prior $\psi_i(\mathbf{x}_i)$ (such as the dynamics prediction prior) and its local image likelihood $p_i(\mathbf{z}_i|\mathbf{x}_i)$, but also the beliefs of its neighborhood targets that compete image resources against it. The influence of the competition is summarized in the "message" term, as defined in Eq. 8, that is passed to $\mathbf{x}_i$ during the mean field iterations.

Based on this mean field iteration, it is clear that the computational complexity of the collaborative tracker is linear with respect to the number of targets and the number of iterations, which is a significant improvement of the methods that deal with the joint state space directly.

During collaborative tracking, when the competition mechanism takes place, the distribution of the targets that are unlikely to win the competition will be diffused around the target that is likely to win, until other image observations become available in the future. Once some targets do not compete, i.e., without motion correlation, their image observations can be readily separated and thus these targets and be tracked independently. At this time, the collaborative tracker acts as the same as M.i.T..

5 Mean Field Monte Carlo

Since the image observation likelihoods are generally non-Gaussian due to the presence of clutters for example, it is not plausible to express the mean field equations in parametric forms by assuming all the densities are Gaussian. Thus, we describe in this section a non-parametric implementation of the mean field inference, called *Mean Field Monte Carlo* (MFMC).

In MFMC, a set of particle is employed to represent the variational density $Q_i(\mathbf{x}_i)$ for each target $\mathbf{x}_i$, i.e.,

$$q_i^k(\mathbf{x}_i) \sim \{s_i^{(n)}(k), \pi_i^{(n)}(k)\}_{n=1}^N$$

where s and π denote the sample and its weight and N is the number of samples. Based on Eq. 9, the Monte Carlo can be summarized as in Figure 3.

```
1. Set k=0, sample $Q_i(\mathbf{x}_{i,t-1})$ for $\{\widetilde{s}_{i,t-1}^{(n)}(k), 1\}_{n=1}^N$.
2. $\forall\, \widetilde{s}_{i,t-1}^{(n)(k)}$, sample $s_{i,t}^{(n)}(k)$ from $p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1})$.
3. Iteration: $k = k + 1$;
    (a) calculate the "message" from neighbors:
        $m_{i,t}^{(n)} = \sum_{j \in \mathcal{N}(i)} \sum_{m=1}^{N} \pi_{j,t}^{(m)}(k-1) \log \psi_{ij}(s_{i,t}^{(n)}(k), s_{j,t}^{(m)}(k-1))$.
    (b) Perform observation for each $s_i^{(n)}(k)$,
        $w_{i,t}^{(n)} = p(z_{i,t}|s_{i,t}^{(n)}(k))$.
    (c) Re-weight the particles by:
        $\pi_{i,t}^{(n)}(k) = e^{m_{i,t}^{(n)}} \times w_{i,t}^{(n)}$.
    (d) normalize to obtain
        $Q_{i,t}^k(\mathbf{x}_{i,t}) \sim \{s_{i,t}^{(n)}(k), \pi_{i,t}^{(n)}(k)\}$
```

Figure 3: The mean field Monte Carlo (MFMC) algorithm.

An equilibrium will be reached after several iterations. Then the optimal variational distributions $Q_{i,t}(\mathbf{x}_{i,t})$ can be treated as the approximation to the posterior $p(\mathbf{x}_{i,t}|\underline{\mathbf{Z}_t})$. In general, mean field equations converge very quickly due to the nature of the fixed point. Although we have not obtained the rigorous results on the convergence rate, we always observe the convergence in less than five iterations in our experiments.

The significance of the above tracking algorithm is its distributed and collaborative mechanism, where each individual target is associated with a particle filter. These particle filters are not independent but competitive through message passing and mean field iterations.

Most recently, Sudderth *et al* [14], Isard [6] and Sigal *et al* [13] have developed algorithms for the interactions among multiple particle sets. These algorithms are based on belief propagation, while the above MFMC algorithm is based on probabilistic variational analysis. Although belief propagation and mean field iteration share the same paradigm of message passing, the difference between them are the contents of the "messages" and the theoretical analysis for the case of loopy graphs. Theoretically, MFMC is a very good choice for tracking multiple targets as described in the previous sections, which is also supported by the extensive and very promising experiment results as will be reported in next section.

6 Experiments

Extensive and comparative experiments on both synthetic and real data are reported in this section. In all these experiments, the individual tracker is a 2D appearance-based region tracker, in which the target state $\mathbf{x}_i$ is modelled by 2D affine parameters, the dynamic model $p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1})$ is a 2nd order dynamic model, the likelihood function $p(\mathbf{z}_i|\mathbf{x}_i)$ is calculated by matching a PCA-based appearance model which is trained in advance, and 200 particles are used to represent the posterior of each target state. In all these experiments, our collaborative tracker runs comfortablely at 15-20 fps on a PIV 2GHz PC.

6.1 Proof-of-concept

To clearly demonstrate the basic idea and the correctness of our approach, we firstly test our algorithm by a synthetic video sequence, in which five identical and moving tennis are casted into a real dynamic scene. This synthetic testing case prevents the subtraction of the background to obtain easy detection of the targets. Each tennis presents an independent const velocity motion and is bounced by the image borders. This sequence challenges many existing method due to the frequent present of occlusions.

Equipped with the competition mechanism, our collaborative tracker performs excellently. Sample frames from the results are shown in Figure 4 and details can be seen in the video `"Tennis1_MFMC.avi"`. We use different colored rectangles to display the estimated target positions. An index is also attached to each rectangle to identify these tennis uniquely. The blue lines in Figure 4 that link the different targets are the visual illustration of the structure of the *ad hoc* Markov network. Therefore, by observing the changing structure of the network over frames, we can clearly learn that which tennis are subject to the collaborative inference and which are simply being tracked by an individual tracker.

Although our collaborative tracker does not deliberately address the identity switching problem, we find in our experiments that our approach seems to have such a capability to nicely handle this problem when combined with motion coherence and dynamic predictions. This can be easily validated by the subjective evaluation on the tracking sequence.

We also compare our results with those obtained by the multiple independent trackers, as shown in Figure 5. The number of particles for each target in the M.i.T. algorithm is the same as in our algorithm. However, M.i.T. can not produce satisfactory results, where the coalescence problems always happen during the tracking. The submitted video for this M.i.T. implementation is `"Tennis1_MiT.avi"`.

6.2 Lab Environments

The second and third test sequences are taken in the laboratory environment. In the second sequence, a person is moving a tennis to cross behind a row of other 4 tennis that act as several identical camouflages. Obviously, the occluding tennis increase the burden of correct tracking of the occluded tennis. As expected, our collaborative framework can still effectively handle the difficulty and lead to a very robust tracking to the occluded target, even successfully keeping the identity of those five tennis. Sample frames of the results are shown in Figure 6 and details can be obtained from the video `"Tennis2_MFMC.avi"`.

The third sequence contains 2 moving tennis and 3 still tennis, where different configuration of the structure of ad hoc Markov network is intentionally exploited by changing the positions of the two movable tennis. Once again, our collaborative framework successfully keeps tracking those five tennis. Sample frames are shown in Figure 7 and the video is `"Tennis3_MFMC.avi"`.

6.3 Real Scenarios

Both our collaborative tracking framework and M.i.T. trackers have been tested on real scenarios. In the first scenario, two persons are walking around in the scene and occlusion continuously happens between these two persons. It is easy for our collaborative tracker to obtain very robust results, as shown in Figure 8 and the video `"TwoHumans_MFMC.avi"`, while M.i.T. can not work well as shown in Figure 9 and video `"TwoHumans_MiT.avi"`.

Finally, a sequence that contains three women soccer players drilling in a field is tested. As expected, our new method provides robust and stable results, as can be seen in Figure 10 and video `"ThreeHumans_MFMC.avi"`.

7 Discussion and Conclusions

Coalescence that means the tracker associates more than one trajectories to some targets while loses track for others is a challenging problem in multiple targets tracking. In this paper, we present a novel collaborative approach with linear complexity to this problem. The basic idea is the collaborative inference mechanism, in which the estimate of an individual target is not only determined by its own observation and dynamics, but also through the interaction and collaboration with the estimates of its adjacent targets, which leads to a competition mechanism that enables different targets to compete for the common resources, i.e. image observations. The theoretical foundation of the new approach is based on Markov networks, in which the links of the network introduce the competition for image resources among targets. Variational analysis of this Markov network reveals a mean field approximation to the posterior densities of each targets. Therefore a mean field Monte Carlo (MFMC) algorithm is designed to efficiently implement this mean field approximation inference by simulating the competition among a set of low dimensional particle filters. The

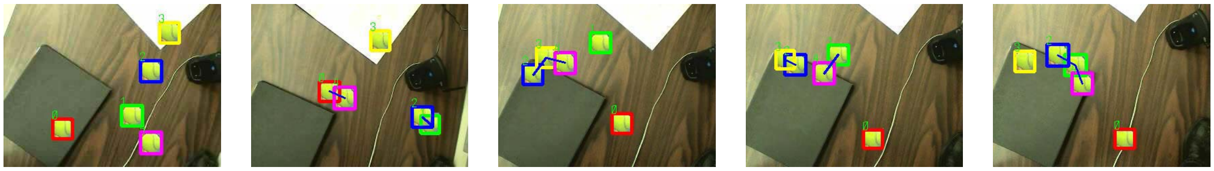

Figure 4: MFMC tracker: 5 tennis in a synthetic video. The blue links among the targets illustrate the structure of the *ad hoc* Markov network. Details please see `Tennis1_MFMC.avi`.

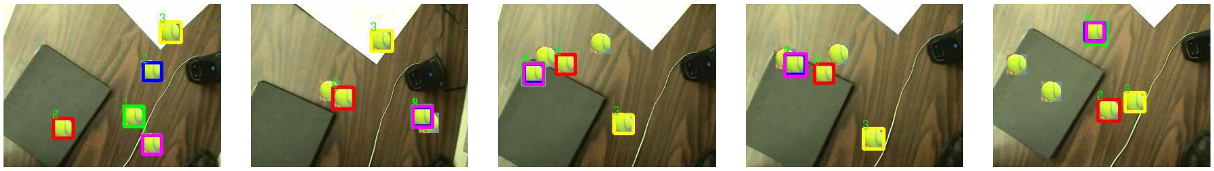

Figure 5: M.i.T. tracker: 5 tennis in a synthetic video. Details please see `Tennis1_MiT.avi`.

effectiveness of handling coalescence and the great computational efficiency have been demonstrated by extensive experiments on various scenarios.

Since in our current framework the addition and deletion of the targets have not be implemented yet, one possible future work is to extend the algorithm to handle it. Because the current Markov network in our approach does have the capability to change the structure configuration, so there should be no theoretical obstacle to prevent us from solving the problem of targets addition and deletion in our framework.

Acknowledgments

This work was supported in part by National Science Foundation Grants IIS-0347877, IIS-0308222, Northwestern faculty startup funds and Murphy Fellowship.

References

[1] Y. Bar-Shalom and T. Fortmann. *Tracking and Data Association*. Academic Press, Orlando, FL, 1988.

[2] A. Blake and M. Isard. *Active Contours*. Springer-Verlag, London, 1998.

[3] W. Freeman, E. Pasztor, and O. Carmichael. Learning low-level vision. *Int'l Journal of Computer Vision*, 40:25–47, 2000.

[4] I. Haritaoglu, D. Harwood, and L. Davis. W4: Who? when? where? what? a real time system for detecting and tracking people. In *Proc. IEEE Int'l Conf. on Face and Gesture Recognition*, Nara, Japan, April 1998.

[5] C. Hue, J. Cadre, and P. Perez. Tracking multiple objects with particle filtering. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):791–812, 2002.

[6] M. Isard. PAMPAS: Real-valued graphical models for computer vision. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume I, pages 613–620, Madison, WI, June 2003.

[7] M. Isard and A. Blake. Contour tracking by stochastic propagation of conditional density. In *Proc. of European Conf. on Computer Vision*, pages 343–356, Cambridge, UK, 1996.

[8] M. Isard and J. MacCormick. BraMBLe: A bayesian multiple-blob tracker. In *Proc. IEEE Int'l Conf. on Computer Vision*, pages 34–41, Vancouver, Canada, 2001.

[9] T. S. Jaakkola. Tutorial on variational approximation methods. MIT AI Lab TR, 2000.

[10] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37:183–233, 2000.

[11] J. MacCormick and A. Blake. A probabilistic exclusion principle for tracking multiple objects. In *Proc. IEEE Int'l Conf. on Computer Vision*, pages 572–578, Greece, 1999.

[12] C. Rasmussen and G. Hager. Probabilistic data association methods for tracking complex visual objects. *IEEE T-PAMI*, pages 560–576, Jun. 2001.

[13] L. Sigal, M. Isard, B. Sigelman, and M. Black. Attractive people: Assembling loose-limbed models using nonparametric belief propagation. In *NIPS*, 2004.

[14] E. Sudderth, A. Ihler, W. Freeman, and A. Willsky. Nonparametric belief propagation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume I, pages 605–612, Madison, WI, June 2003.

[15] H. Tao, H. Sawhney, and R. Kumar. A sampling algorithm for detecting and tracking multiple objects. In *Proc. ICCV'99 Workshop on Vision Algorithm*, Corfu, Greece, 1999.

[16] Y. Wu, G. Hua, and T. Yu. Tracking articulated body by dynamic Markov network. In *Proc. IEEE Int'l Conf. on Computer Vision*, pages 1094–1101, Nice, France, Oct. 2003.

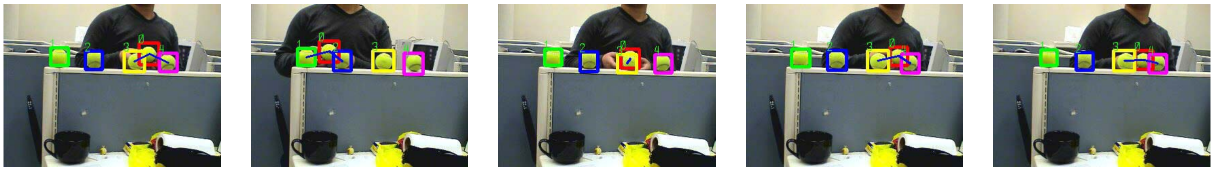

Figure 6: MFMC tracker: a tennis moving behind a row of 4 tennis. The blue links among the targets illustrate the structure of the *ad hoc* Markov network. Details please see "Tennis2_MFMC.avi".

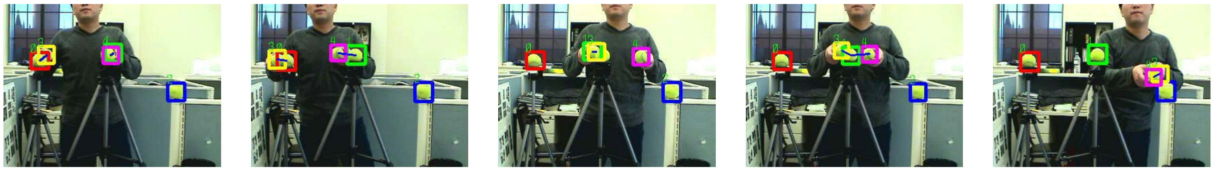

Figure 7: MFMC tracker: 2 tennis moving around 3 static tennis. The blue links among the targets illustrate the structure of the *ad hoc* Markov network. Details please see "Tennis3_MFMC.avi".

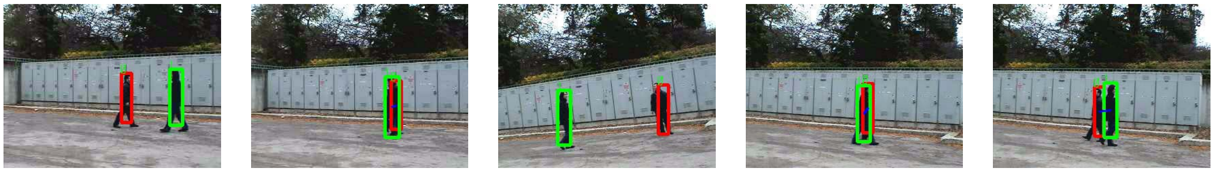

Figure 8: MFMC tracker: two people walking. The blue links among the targets illustrate the structure of the *ad hoc* Markov network. Details please see "TwoHumans_MFMC.avi".

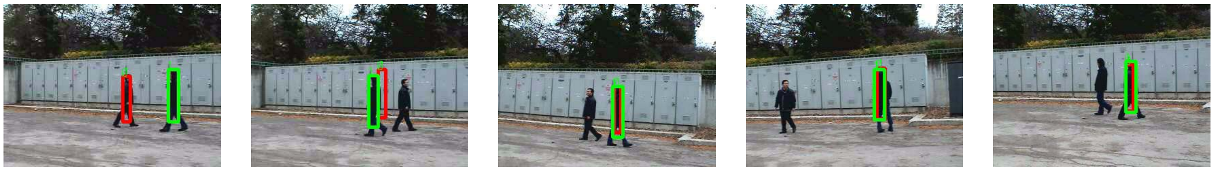

Figure 9: M.i.T. tracker: two people walking. Details please see "TwoHumans_MiT.avi" .

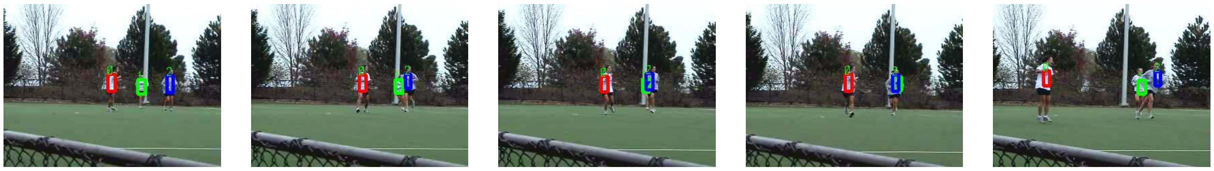

Figure 10: MFMC tracker: three women soccer players drilling. The blue links among the targets illustrate the structure of the *ad hoc* Markov network. Details please see "ThreeHumans_MFMC.avi".