in Proc. IEEE Int'l Conf. on Computer Vision, Nice, France, 2003

Tracking Articulated Body by Dynamic Markov Network

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

Abstract

A new method for visual tracking of articulated objects is presented. Analyzing articulated motion is challenging because the dimensionality increase potentially demands tremendous increase of computation. To ease this problem, we propose an approach that analyzes subparts locally while reinforcing the structural constraints at the mean time. The computational model of the proposed approach is based on a dynamic Markov network, a generative model which characterizes the dynamics and the image observations of each individual subpart as well as the motion constraints among different subparts. Probabilistic variational analysis of the model reveals a mean field approximation to the posterior densities of each subparts given visual evidence, and provides a computationally efficient way for such a difficult Bayesian inference problem. In addition, we design mean field Monte Carlo (MFMC) algorithms, in which a set of low dimensional particle filters interact with each other and solve the high dimensional problem collaboratively. Extensive experiments on tracking human body parts demonstrate the effectiveness, significance and computational efficiency of the proposed method.

1 Introduction

Tracking articulated motion in images is an important problem, especially when the research of video-based human sensing has been advocated to achieve such emerging applications as perceptual interfaces [20], smart video surveillance [8] and automatic video footage [4], etc.

The problem involves the localization and identification of a set of linked but articulated subparts. Inheriting all the difficulties from single object tracking, the problem of tracking articulated body has to tackle some special challenges. One of these is the complexity incurred by the degrees of freedom of the articulated body.

Different from multiple target tracking where the motion of each subpart is independent of others, the physical links among different subparts reinforce motion constrains upon these articulated subparts. We can have an intuitive comparison of these two cases by the configuration space which is the joint motion space of the set of subparts. If the motion of subparts are independent, then configuration space will enjoy a nice property that the motion of each subpart stays in a linear subspace which is orthogonal to the subspaces corresponding to other subparts. Thus, independent trackers can be used to track independent multiple targets and the complexity is almost linear w.r.t. the number of targets. However, when the subparts are physically linked, the configuration space will not have such a nice orthogonality and factorization property of subspaces. Thus, the high dimensionality seems unavoidable, which is generally associated with the exponential increase of computation due to the curse of dimensionality.

Various approaches have been investigated to alleviate this challenge (see Section 2 for details), such as dynamic programming [18], annealed sampling [6], partitioned sampling [15, 16], eigenspace tracking [1], hybrid Monte Carlo filtering [5], etc. Different from these approaches, in this paper, we propose a novel solution based on a dynamic Markov network model and variational mean field approximations. The proposed dynamic Markov network embeds the subparts constraints in an undirected graphical model (i.e., a Markov network) associated with image observation processes, thus the model serves as a generative model for the articulated motion. Due to the dense connections in the graph, exact analysis is complicated and intractable. When we perform an analysis based on variational mean field method, tight approximation can be achieved while the computational complexity is significantly reduced. At each time instance, the mean field solution is achieved through efficient Monte Carlo algorithm. And based on that, we design a mean field sequential Monte Carlo for articulated body tracking. Extensive experiments show the effectiveness and efficiency of the proposed approach.

2 Related Work

There is a substantial literature on articulated motion analysis, and many different approaches have been investigated. For all these methods, three important issues should be addressed: the representations for articulated objects, the computational paradigms, and the way of reducing computation.

Basically, there could be two types of choices for articulated object representations. One employs joint angles [3, 13, 17, 16, 21], while the other uses the collection of the motion of all subparts. Of course, the first representation is non-redundant and reflects the degrees of freedom of the articulated motion directly, while the second one is highly redundant. However, due to the independence of the joint angles, the first method may suffer from an irreducible dilemma since the intrinsic dimensionality is probably reached. In this sense, the motion estimation problem can be posed as an unconstrained optimization in a high dimensional space (if we do not consider the natural motion constrain as in [21]). On the other hand, if the articulated motion is redundantly described by the individual motion of the subparts, each subpart may be solved individually, and then projected to the constrained space. Thus, it corresponds to a constrained optimization problem in a high dimensional space. By taking advantage of the structure of such a redundant representation, efficient solutions can be found as in this paper.

There are also different computational paradigms for articulated motion analysis. It could be deterministic or probabilistic. Deterministic methods generally formulate the problem as a parameter estimation problem based on nonlinear programming techniques, and then differential approach can be taken [17, 3, 13]. Thus, the difficulty of local minima exists. A probabilistic approach formulates the motion analysis problem as a Bayesian inference problem, where particle-based Monte Carlo strategies provide flexible but intensive computing frameworks. In general, the number of particles needed will increase exponentially with the increase of the dimensionality.

Therefore, it is crucial to reduce the computation. Different schemes have been investigated to reduce the number of particles. For example, the annealed particle filtering method performs a coarse-to-fine layered search [6], partitioned sampling is in the spirit of coordinate descent and preforms a hierarchical sampling [15, 16]. Both methods work with high dimensional probability spaces. Different from these methods, this paper presents a mean field Monte Carlo (MFMC) algorithm in which a set of low dimensional particle filters interact with each other to solve a high dimensional problem collaboratively.

3 The Representation of an Articulated Body

We denote the motion of each individual subpart by $\mathbf{x}_i$, which can be the parameters of an affine motion. The motion of an articulated body is the concatenation $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M]$. Certainly, it is highly redundant. The image observation associated with $\mathbf{x}_k$ is denoted by $\mathbf{z}_k$, which could be the detected edges of the shape contour of the subpart, and the collective image observation of the entire articulated body is $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_M]$. An important task is to infer the posterior $p(\mathbf{X}|\mathbf{Z})$.

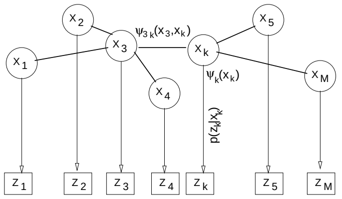

Figure 1: The Markov Network for an articulated body.

As shown in Figure 1, a mixture of undirected and directed graphical model can be used to characterize the generative process. The hidden layer is an undirected graph $G_x = \{V, E\}$, representing the relationship among different articulated parts. Obviously, different parts are not independent, and each individual part only interacts with its neighborhood parts. We denote the neighborhood parts of i by $\mathcal{N}(i)$. Clearly, it is a Markov network. In addition, each individual part is associated with its observation and the conditional likelihood distribution $p(\mathbf{z}_i|\mathbf{x}_i)$ is represented by a directed link.

Given the undirected graph of $\mathbf{X}$, $p(\mathbf{X})$ can be modelled as a Gibbs distribution and can be factorized as:

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

where c is a clique in the set of cliques $\mathcal{C}$ of the undirected graph, X_c is the set of hidden nodes associated with the clique and $\psi_c(X_c)$ is the probability of this clique, and Z_c is a normalization term or the partition function. Although Z_c is difficult to compute, we do not compute it, since a Monte Carlo method will be used as shown in later sections. The model accommodates two types of cliques: the first order clique, i.e., $i \in \mathcal{C}^1 = V$, and second order clique, i.e., $(i,j) \in \mathcal{C}^2 = E$, where $\mathcal{C} = \mathcal{C}^1 \bigcup \mathcal{C}^2$. The associated ψ is denoted by ψ_i and ψ_{ij}, respectively. Thus, Eq. 1 can also be written as:

$$p(\mathbf{X}) = \frac{1}{Z_c} \prod_{(i,j) \in \mathcal{C}^2} \psi_{ij}(\mathbf{x_i}, \mathbf{x_j}) \prod_{i \in \mathcal{C}^1} \psi_i(\mathbf{x_i}) \tag{2}$$

where $\psi_i(\mathbf{x_i})$ provides a local prior for $\mathbf{x}_i$, and $\psi_{ij}(\mathbf{x_i}, \mathbf{x_j})$ presents the constraints between the neighborhood nodes $\mathbf{x}_i$ and $\mathbf{x}_j$. In other words, $\psi_i(\mathbf{x_i})$ predicates a prior for the i-th part, while $\psi_{ij}(\mathbf{x_i}, \mathbf{x_j})$ reinforces the constraints between the i-th part and the j-th part. As a specific example, it can be modelled as:

$$\psi_{ij}(\mathbf{x_i}, \mathbf{x_j}) \propto e^{D(\mathbf{x}_i,\mathbf{x}_j)^T \Sigma^{-1} D(\mathbf{x}_i,\mathbf{x}_j)} \tag{3}$$

where $D(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{u}_i(\mathbf{x}_i) - \mathbf{u}_j(\mathbf{x}_j)$, and $\mathbf{u}_i(\mathbf{x}_i)$ and $\mathbf{u}_j(\mathbf{x}_j)$ are shown in Figure 2.

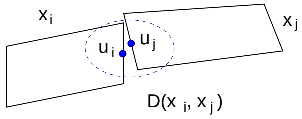

Figure 2: The constraint of two articulated parts.

Given a $\mathbf{x}_i$, its local observation $\mathbf{z}_i$ is independent of other articulated parts. Thus, we have:

$$p(\mathbf{Z}|\mathbf{X}) = \prod_{i=1}^{n} p_i(\mathbf{z_i}|\mathbf{x_i}). \qquad (4)$$

The problem of great interest is to infer the posterior $p(\mathbf{x}_i|\mathbf{Z})$. An intuition is that the posterior of $\mathbf{x}_i$ should be affected by three factors: its local prior ψ_i, its local evidence $\mathbf{z}_i$, and the constraints reinforced by its neighborhood through ψ_{ij}. This intuition will become clearer in Section 4. Since the exact analysis of such a model is complicated and involves heavy computation, it is more plausible to have an approximate but efficient solution.

4 Mean Field Approximation

Variational analysis provides an approximate method for analyzing the model [12, 11]. The core idea of variational approximation is to find an optimal *variational distribution* $q(\mathbf{X})$ that approximates the posterior distribution $p(\mathbf{X}|\mathbf{Z})$, such that the Kullback-Leibler (KL) divergence of these two distributions is minimized, i.e.,

$$\begin{aligned} q^*(\mathbf{X}) &= \arg\min_q KL(q(\mathbf{X}||p(\mathbf{X}|\mathbf{Z})) \\ &= \arg\min_q \int_x q(\mathbf{X}) \log \frac{q(\mathbf{X})}{p(\mathbf{X}|\mathbf{Z})} \end{aligned}$$

Selecting a good class of variational distributions q would largely ease the difficulties, but it requires substantial creativity [12]. Here, we adopt a fully factorized form for simplicity:

$$q(\mathbf{X}) = \prod_i^M q_i(\mathbf{x}_i) \qquad (5)$$

where $q_i(\mathbf{x_i})$ is an independent distribution of the hidden node $\mathbf{x}_i$. Then, $H(q) = \sum_i H(q_i)$, where $H(q)$ is the entropy of $q(\mathbf{X})$, and $H(q_i)$ is the entropy of $q(\mathbf{x}_i)$. Then, the KL-divergence can be written as:

$$\begin{aligned} KL(q_i) = -H(q_i) - \int_{x_i} q_i(\mathbf{x}_i) E_q[\log p(\mathbf{X}, \mathbf{Z})|\mathbf{x}_i] \\ - \sum_{k \neq i} H(q_k) + \log p(\mathbf{Z}) \qquad (6) \end{aligned}$$

where $E_q[\cdot|\mathbf{x}_i]$ is the conditional expectation given $\mathbf{x}_i$ w.r.t. $q(\mathbf{X})$, and $\log p(\mathbf{Z})$ is the data likelihood, which is a constant. To search for a set of q_i to minimize Eq. 6, since each q_i is constrained to be a valid distribution function, we should construct a Lagrangian for each q_i:

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

Setting the derivative to zero, it is easy to see the solution to this constrained optimization problem is a set of fixed point equations:

$$q_i(\mathbf{x}_i) = \frac{1}{Z_i} e^{E_q[\log p(\mathbf{X},\mathbf{Z})|\mathbf{x}_i]} \qquad (8)$$

where $E_q[\log p(\mathbf{X}, \mathbf{Z})|\mathbf{x_i}]$ is the conditional expectation given $\mathbf{x}_i$, Z_i is a partition function for normalization, and $1 \leq i \leq M$. The iterative updating of $q_i(\mathbf{x}_i)$ will monotonically decrease the KL-divergence, and eventually reach an equilibrium. These updating equations are called *mean field equations*.

Moreover, the factorization of $p(\mathbf{X})$ in Eq. 2 enables further simplification of the mean field equations in Eq. 8. It is easy to show that:

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

$$M_i(\mathbf{x}_i) = \exp\{\sum_{k \in \mathcal{N}(i)} \int_{x_k} q_k(\mathbf{x}_k) \log \psi_{ik}(\mathbf{x}_i, \mathbf{x}_k)\}, \qquad (9)$$

where Z_i' is a constant, and $\mathcal{N}(i)$ is the neighborhood of the subpart i. From Eq. 9, the intuition stated at the end of Section 3 is more pronounced: the variational belief of a subpart $\mathbf{x}_i$ is determined by three factors: the local conditional likelihood $p_i(\mathbf{z}_i|\mathbf{x}_i)$, the local prior $\psi_i(\mathbf{x}_i)$, and the beliefs of the neighborhood subparts $\mathbf{x}_{\mathcal{N}(i)}$ (we call it neighborhood prior). This is illustrated in Figure 3.

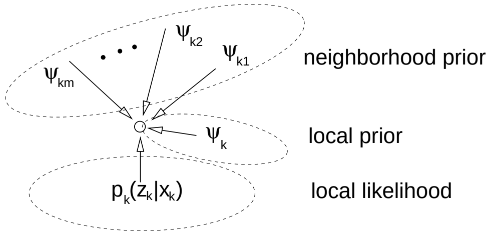

Figure 3: Three factors affect the updating of $q(\mathbf{x}_k)$.

Thus,we can treat the term $p_i(\mathbf{z}_i|\mathbf{x}_i)\psi_i(\mathbf{x}_i)$ as an analogue as the local belief, and treat the term $M_i(\mathbf{x}_i)$ as an analogue to the "message" [7] propagated through the nearby subpart of $\mathbf{x}_i$ in the belief propagation approach, but the computation of $M_i(\mathbf{x}_i)$ here is easier. In addition, we can clearly see from this equation that the computation is significantly reduced by avoiding multi-dimensional integrals, since Eq. 9 involves only one dimensional integrals.

5 Mean Field Monte Carlo (MFMC)

In this section, we propose a Monte Carlo method to implement the mean field updating as discussed in Section 4. We call this method *Mean Field Monte Carlo* (MFMC).

Once the mean field updating converges to a fixed point, then the set of optimal variational distributions $q(\mathbf{x}_i)$, where $i = 1, \ldots, M$, is obtained and can be treated as the optimal approximation to the posterior density $p(\mathbf{x}_i|\mathbf{Z})$.

To make the presentation clear, here we use a 2-link body as an example. W.l.g., we use i and j to index the two linked subparts, and we use k to index the mean field iteration. At the $k-1$-th iteration, for each subpart, a set of particle is maintained to represent the variational distribution, i.e.,

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

where s and π denote the sample and the weight respectively. Then at the next iteration, we perform the following steps according to Eq. 9:

1. Sampling local prior $\psi_i(\mathbf{x}_i)$ for $\{s_i^{(n)}(k), 1\}_{n=1}^N$;
2. calculating the "message" from j:
$$m_i^{(n)} = \sum_{t=1}^N \pi_j^{(n)}(k-1) \log \psi_{ij}(s_i^{(n)}(k), s_j^{(t)}(k-1)).$$
3. Performing observation for each particle $s_i^{(n)}(k)$,
$$w_i^{(n)} = p(z_i|s_i^{(n)}(k)).$$
4. Re-weighting the particles by:
$$\pi_i^{(n)}(k) = e^{m_i^{(n)}} \times w_i^{(n)}.$$
and normalize to produce $\{s_i^{(n)}(k), \pi_i^{(n)}(k)\}$.
5. Performing the same steps for j according to Eq. 9. And then increase k for next mean field updating.

After the k-th iteration, we end up with:

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

After several iterations, the distribution will reach an equilibrium. For a subpart which is linked to multiple subparts, the only difference is in the 2nd step of calculating "messages",

$$m_i^{(n)} = \sum_{j \in \mathcal{N}(i)} \sum_{t=1}^N \pi_j^{(n)}(k-1) \log \psi_{ij}(s_i^{(n)}(k), s_j^{(t)}(k-1)).$$

which sums over all "messages" passed from the neighbors $\mathcal{N}(i)$ (i.e., the Markov blanket) of $\mathbf{x}_i$.

Since $\mathbf{x}_i$ describes the motion of a subpart, its image observation $\mathbf{z}_i$ should be a function of $\mathbf{x}_i$, i.e., $p(\mathbf{z}_i|\mathbf{x}_i)$ is in fact $p(\mathbf{z}_i(\mathbf{x}_i)|\mathbf{x}_i)$. Since $p(\mathbf{z}_i|\mathbf{x}_i)$ will be used to re-weight the belief (or the posterior density) of $\mathbf{x}_i$, the locations of the particles $\{s_i^{(n)}\}$ will affect the faith of approximating the belief by the set of particles, if the ratio of valid particles is not satisfactory (meaning that a small portion of the particles dominates the re-weighting). To enhance the ratio of valid particles, we use importance sampling technique [14] to place the particles to "better" locations.

The only modification on the above mean field Monte Carlo (MFMC) is on the first step: instead of sampling the local prior $\psi(\mathbf{x}_i)$ directly to produce $\{s_i^{(n)}, 1\}_{n=1}^N$, we draw samples $\{s_i^{(n)}, 1\}_{n=1}^N$ from an importance density $g(\mathbf{x}_i)$. After weight compensation, the set of re-weighted particle still a properly weighted set for the density $\psi(\mathbf{x}_i)$, i.e.,

$$\psi(\mathbf{x}_i) \sim \{s_i^{(n)}, \frac{\psi(s_i^{(n)})}{g(s_i^{(n)})}\}_{n=1}^N.$$

The selection of importance density can be arbitrary. Here we give an specific example by using a two-link (where i and j are connected subparts). To generate samples for $\psi_i(\mathbf{x}_i)$, we find the means $\bar{s}_i$ and $\bar{s}_j$ from the two particle sets. After identifying the point $\bar{u}_j$ on $\bar{s}_j$ and the median axis $\bar{L}_i$ of $\bar{s}_i$ (see Figure 4), we sample $u_i^{(n)}$ from $\mathcal{G}(u_i : \bar{u}_j, \Sigma_u)$, and $L_i^{(n)}$ from $\mathcal{G}(L_i : \bar{L}_i, \Sigma_L)$, where $\mathcal{G}$ is Gaussian.

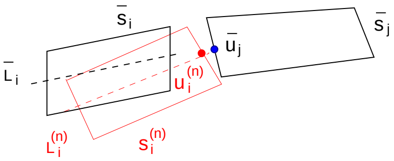

Figure 4: Importance density.

Then the sample $s_i^{(n)}$ is produced by $(L_i^{(n)}, u_j^{(n)})$, and the importance density is:

$$g(\mathbf{x}_i) = \mathcal{G}(u_i : \bar{u}_j, \Sigma_u)\mathcal{G}(L_i : \bar{L}_i, \Sigma_L).$$

Similar importance densities can be easily constructed for a subpart which is linked to multiple subparts. The use of importance sampling techniques greatly enhances the robustness of the mean field Monte Carlo algorithms.

Sudderth *et al* [19] and Isard [9] have independently developed algorithms for the interactions among multiple particle sets. Their methods are based on belief propagation, while our method on probabilistic variational analysis and mean field iterations.

6 Dynamic Markov Network and Sequential Mean Field Monte Carlo

Section 4 and Section 5 describe the mean field approximation and mean field Monte Carlo at one time instance. They can be easily modified for tracking. When considering multiple time instances, the model becomes a dynamic Markov network, as shown in Figure 5. Denote the collection of observations by $\underline{\mathbf{Z}}_t = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_t\}$. Tracking algorithms aim at inferring $p(\mathbf{X}_t|\underline{\mathbf{Z}}_t)$ by knowing $p(\mathbf{X}_{t-1}|\underline{\mathbf{Z}}_{t-1})$. It involves a density propagation process [10]:

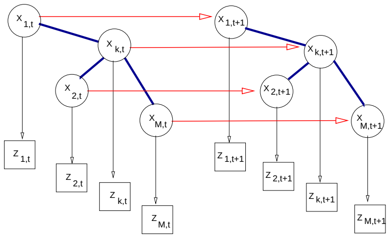

Figure 5: Dynamic Markov Network

$$p(\mathbf{X}_t|\underline{\mathbf{Z}}_t) \propto p(\mathbf{Z}_t|\mathbf{X}_t) \int_{x_{t-1}} p(\mathbf{X}_t|\mathbf{X}_{t-1}) p(\mathbf{X}_{t-1}|\underline{\mathbf{Z}}_{t-1})$$

Once $\mathbf{X}$ consists of a number of articulated parts, the increase of dimensionality will incur exponential increase of computation. The advantage of mean field approximation is that it decouples different parts, and transforms the problem of exponential complexity to a simpler problem close to linear complexity. The constraint reinforcement needs some computation as a cost, but it is not significant.

At time instance t, mean field approximation finds a variational distribution $q_{i,t}(\mathbf{x}_i)$ to approximate $p(\mathbf{x}_{i,t}|\underline{\mathbf{Z}}_t)$ for the i-th subpart. The mean field equation can be written as:

$$q_{i,t}(\mathbf{x}_{i,t}) = \frac{1}{Z_i'} p_i(\mathbf{z}_{i,t}|\mathbf{x}_{i,t}) \times \int p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1}) q_{i,t-1}(\mathbf{x}_i) \times \exp\{\sum_{k \in \mathcal{N}(i)} \int_{x_{k,t}} q_{k,t}(\mathbf{x}_k) \log \psi_{ik}(\mathbf{x}_{i,t}, \mathbf{x}_{k,t})\} \quad (10)$$

Comparing Eq. 10 to Eq. 9, we clearly see that the predication density $\int p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1}) q_{i,t-1}(\mathbf{x}_i)$ in Eq. 10 of a dynamic Markov network plays the same role as $\psi_i(\mathbf{x}_i)$ in Eq. 9. Thus, at time instance t, the variational belief of the i-th subpart is also determined by three factors: the local evidence, the predication prior from previous time frame, and the belief of the neighborhood subparts.

Therefore, the sequential mean field Monte Carlo can be obtained by modifying the mean field Monte Carlo algorithm in Section 5. At the first step, instead of sampling from $\psi_i(\mathbf{x}_i)$, we should sample the prediction prior instead. Suppose at $t-1$, $q_{i,t-1}(\mathbf{x}_i)$ is represented by:

$$q_{i,t-1}(\mathbf{x}_i) \sim \{s_{i,t-1}^{(n)}, \pi_{i,t-1}^{(n)}\}_{n=1}^N.$$

The, we can use the following steps to replace the 1st step in the mean field Monte Carlo algorithm in Section 5:

1.a Re-sampling from $q_{i,t-1}(\mathbf{x}_i)$ for $\{\tilde{s}_{i,t-1}^{(n)}, 1\}_{n=1}^N$.

1.b $\forall\ \tilde{s}_{i,t}^{(n)}$, sampling $s_{i,t}^{(n)}$ from $p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1})$.

Impressive results have been achieved and reported in Section 7.

We have a rough comparison on the computational complexity of the proposed approach with the original CONDENSATION algorithm with joint angle representation. Assume the articulated body consists of M subparts, each of which contribute one DoF, and assume a number of T particles are needed for tracking one subpart. In addition, we assume when one more DoF is added, CONDENSATION needs $P \times T$ particles to work. Through our experiments, 10 is reasonable for P. In our mean field Monte Carlo, we denote the number of mean field iteration by K, which is 5 in our experiments. In both methods, the most intensive computation is on calculating image observation, while the extra computation induced by $M_i(\mathbf{x}_i)$ in Eq. 9 is negligible. Thus, the complexity of our method is $O(TKM)$, while CONDENSATION has $O(TP^{M-1})$ which is much larger than the proposed mean field Monte Carlo algorithm.

In addition, MFMC is different from the *partitioned sampling* (PS) method [15, 16], although both can reduce the exponential complexity. (1) PS applies to centralized models with independent dimensions, while MFMC can handle various HDMs including articulation, deformation and multi-motion; (2) PS uses one high-dimensional particle filter, while MFMC use a network of low-dimensional but collaborative particle filters; (3) PS is hierarchical and uni-directional, while MFMC is networked and multi-directional.

7 Results

We performed extensive experiments on articulated body with different DoFs, and obtained impressive results as reported in this section.

7.1 Experimental Setup

Our experiments mainly concerned about 2D tracking. Thus we adopted a cardboard model where each subpart in the articulated body is represented by a planar object, and thus the state of $\mathbf{x}_i$ is the parameters of a 2D affine transform. The motion model $p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1})$ is a standard 2nd order const acceleration model for each subpart. Although the motion model can be learned, we preset the parameters for simplicity.

The observation model $p(\mathbf{z}_i|\mathbf{x}_i)$ is also an important factor in tracking. We used two types of visual cues: edge and intensity. We adopted the same method in CONDENSATION [10, 2] for edge observation, where a set of independent measurement lines were used to measure the likelihood of detected edge points. In addition, since the articulated targets were human body parts and the skin or clothes on the body parts are similar, we also used the intensity clue and assumed the distribution of the intensity of a subpart be a Gaussian distribution. The mean and variance of the Gaussian density was trained for each individual subpart.

7.2 Results of MFMC Iteration

To verify if the mean field updating does converge and to check if it is functioning as expected, we collected the intermediate results on MFMC iteration. Two examples are shown in Figure 6 and Figure 7. The upper half shows an example of a 2-part arm, and the lower half 3-part finger. In both cases, the estimates of the first five iterations are shown. Before the iteration, the initial status was quite unpleasant. But after a couple of mean field iteration, the estimates settled down on the right positions as expected. From our experiments, most iterations converged in less than five times.

7.3 Various Articulated Objects

To demonstrate the effectiveness, efficiency and scalability, we performed experiments on various articulated objects of difference DoFs, including a 2-part arm, 3-part finger, 6-part upper body, and 10-part full body [1].

The first test sequence is a 2-part arm, which consists of two subparts: upper arm and lower arm. The sequence consists of 441 frames. The lower arm presents larger motion than upper arm in the testing sequence. The MFMC algorithm performed excellently due to the constraint reinforcement. Sample frames are shown in Figure 8.

We compared the results from MFMC with multiple independent trackers (MiT). Although there are only two subparts, MiT did not produce satisfactory results, since either one had risks to lose track and there were no other constraints to get it back except image observations, and MiT hardly produced plausible results satisfying the physical link constraints. Some frames of MiT are shown in Figure 9.

The second test sequence is on a 3-part finger and consists of 182 frames. As expected, MFMC produced very robust and stable result. Sample frames are shown in Figure 10.

The third test sequence is on a 6-part upper body, where complex arm motions present as well as global movement of the torso and head. The sequence consists of 834 frames. Although the articulation is quite complicated, it did not fail MFMC. Sample frames are shown in Figure 11.

The most complicated test sequence we have experimented is the 10-part full body motion, and sequence has 368 frames. Arms and legs are the most articulated body parts, and they present significant motion. None of our run of MiT succeeded, because a leg was easy to get lost and never be able to come back. Sample frames of MiT are shown in Figure 12.

When MFMC was applied, the tracking result was still very stable unlike MiT. Through subjective evaluation, the tracking quality did not decrease due to the increase of the complexity of the articulation. Sample frames are shown in Figure 13.

The MFMC algorithm runs on a single processor PC of 2.0GHz running WindowXP. We did not perform code optimization. For all these experiments, the number of mean field iteration was set to 5. The number of particles for each part and the frame rates are shown in Table 1.

experiments	2-part	3-part	6-part	10-part
particles/part	200	200	200	200
frame/second	2.02	1.28	0.94	0.56

Table 1: A comparison of the computation of different articulated objects. The exponential requirement for computation is overcome as expected.

8 Discussion and Conclusions

Tracking articulated objects is a challenging problem, since the increase of number of subparts and the physical connection constraints of them would potentially incur high dimensionality, and fail tracking algorithms developed for single target. Thus, algorithms with close to linear complexity would have much better scalability. In this paper, we propose a collaborative approach to achieve such a goal. Instead of using the joint angle representation which is irreducible, we adopt a highly redundant representation for articulated body, i.e., represent individual subpart by its own motion parameters, but reinforce the constraints of different subparts by a Markov network. Variational analysis is performed for approximated analysis of this graphical model. Interestingly, a set of fixed point equations (i.e., the mean field equations) is found, which suggests a collaborative solution to the problem through interaction with neighborhood subparts and through iterations. Then a mean field Monte Carlo (MFMC) algorithm is designed to achieve effective computation. Considering motion, we propose a dynamic Markov network model and MFMC is extended to a sequential MFMC algorithm for tracking. Extensive experiments demonstrate the applicability of the proposed methods.

[1] Demo sequences are at `http://www.ece.nwu.edu/~yingwu`.

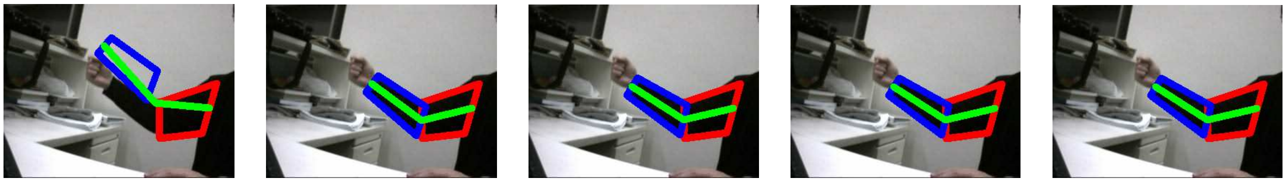

Figure 6: The first five iterations of MFMC on the (2-part) Arm sequence.

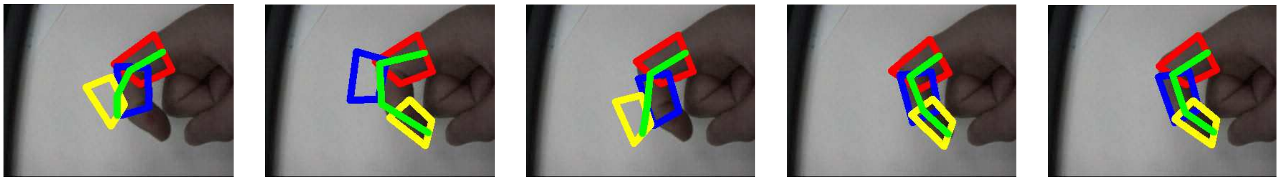

Figure 7: The first five iterations of MFMC on the (3-part) Finger sequence.

One of the future work is to extend the algorithm to 3D. Since self-occlusion seems a severe issue for articulated motion, another possible direction is to design collaborative algorithms for solving the occlusion problem.

References

[1] M. Black and A. Jepson. Eigentracking: Robust matching and tracking of articulated object using a view-based representation. In *Proc. European Conf. Computer Vision*, volume 1, pages 343–356, Cambridge, UK, 1996.

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

[3] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 8–15, Santa Barbara, CA, June 1998.

[4] T.-J. Cham and J. Rehg. A multiple hypothesis approach to figure tracking. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 239–244, 1999.

[5] K. Choo and D. Fleet. People tracking using hybrid Monte Carlo filtering. In *Proc. IEEE Int'l Conf. on Computer Vision*, volume II, pages 321–328, Vancouver, Canada, July 2001.

[6] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle filtering. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume II, pages 126–133, Hilton Head Island, South Carolina, 2000.

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

[8] 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.

[9] 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.

[10] 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.

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

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

[13] S. Ju, M. Black, and Y. Yacoob. Cardboard people: A parametrized model of articulated motion. In *Proc. Int'l Conf. on Automatic Face and Gesture Recognition*, pages 38–44, Killington, Vermont, Oct. 1996.

[14] J. Liu, R. Chen, and T. Logvinenko. A theoretical framework for sequential importance sampling and resampling. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo in Practice*. Springer-Verlag, New York, 2000.

[15] 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.

[16] J. MacCormick and M. Isard. Partitioned sampling, articulated objects, and interface-quality hand tracking. In *Proc. of European Conf. on Computer Vision*, volume 2, pages 3–19, 2000.

[17] J. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. In *Proc. of IEEE Int'l Conf. Computer Vision*, pages 612–617, 1995.

[18] Y. Song, X. Feng, and P. Perona. Towards detection of human motion. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Hilton Head Island, SC, June 2000.

[19] 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.

[20] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9:780–785, July 1997.

[21] Y. Wu, J. Lin, and T. S. Huang. Capturing natural hand articulation. In *Proc. IEEE Int'l Conference on Computer Vision*, volume II, pages 426–432, Vancouver, July 2001.

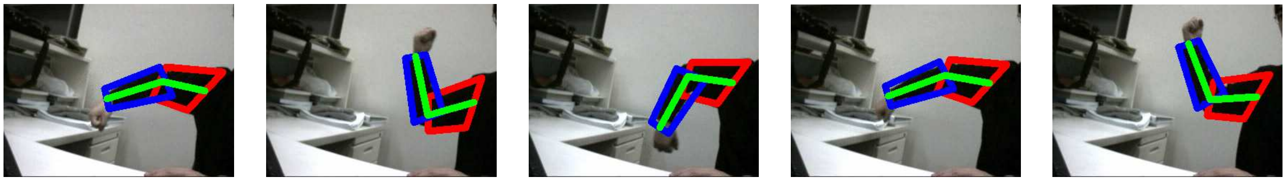

Figure 8: Mean field Monte Carlo (MFMC): tracking 2-part arm.

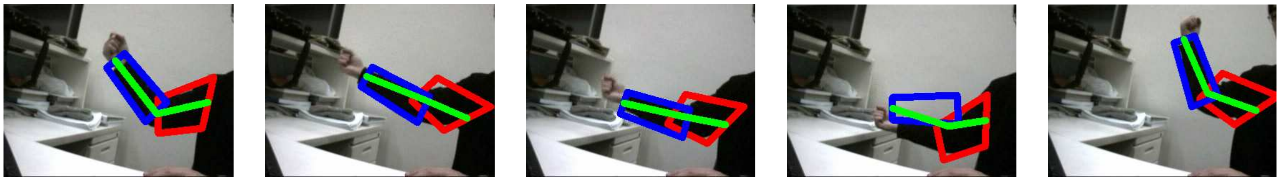

Figure 9: Multiple independent tracker (MiT): tracking 2-part arm.

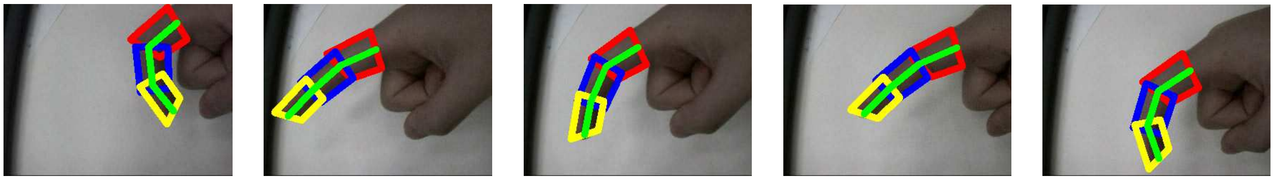

Figure 10: Mean field Monte Carlo (MFMC): tracking 3-part finger.

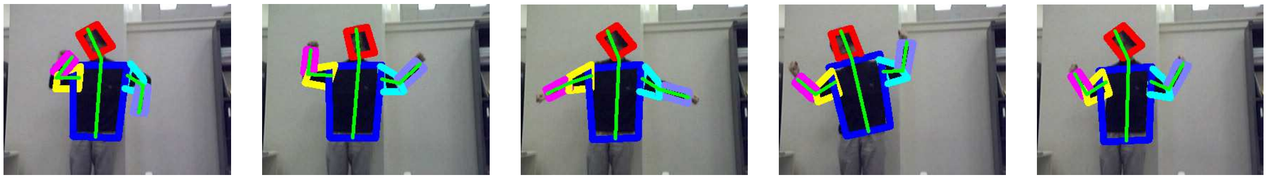

Figure 11: Mean field Monte Carlo (MFMC): tracking 6-part upper body.

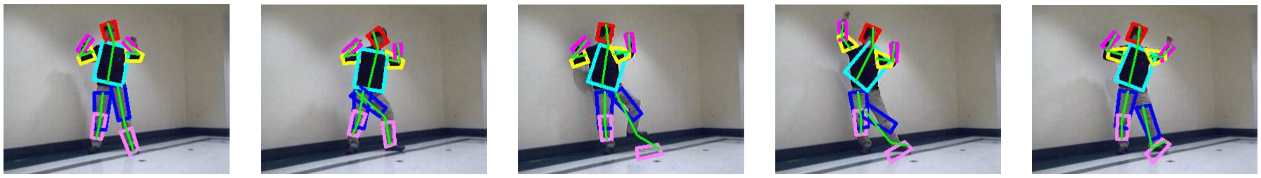

Figure 12: Multiple independent tracker (MiT): tracking 10-part full body.

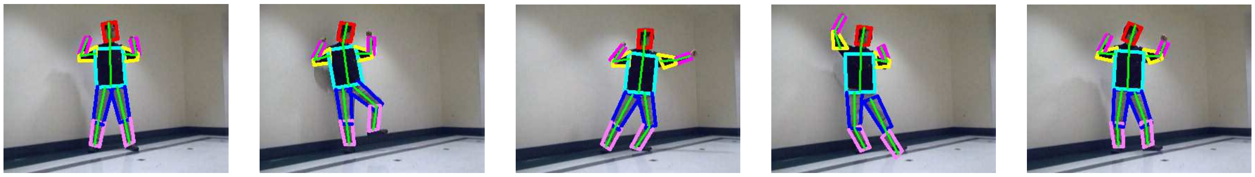

Figure 13: Mean field Monte Carlo (MFMC): tracking 10-part full body.