

VARIATIONAL MONTE-CARLO APPROACH FOR ARTICULATED OBJECT TRACKING

Kartik Dwivedi¹, Harish Bhaskar² and Artur Loza²

¹Department of Electronics and Electrical Engineering, Indian Institute of Technology, Guwahati
k.dwivedi@iitg.ac.in

²Department of Electrical and Computer Engineering
Khalifa University, Abu Dhabi, U.A.E.
harish.bhaskar, artur.loza@kustar.ac.ae

ABSTRACT

In this paper, we describe a novel variational Monte Carlo approach for modeling and tracking body parts of articulated objects. An articulated object (human target) is represented as a dynamic Markov network of the different constituent parts. The proposed approach combines local information of individual body parts and other spatial constraints influenced by neighboring parts. The movement of the relative parts of the articulated body is modeled with local information of displacements from the Markov network and the global information from other neighboring parts. We explore the effect of certain model parameters (including the number of parts tracked; number of Monte-Carlo cycles, etc.) on system accuracy and show that our variational Monte Carlo approach achieves better efficiency and effectiveness compared to other methods on a number of real-time video datasets containing single targets.

KEYWORDS

Variational Inference, Articulated Object Tracking, People Tracking, Kullback-Liebler distance

1. INTRODUCTION AND RELATED WORK

Articulated object tracking is a central problem in Computer Vision particularly due to emerging applications within human computer interfaces, intelligent video surveillance, gait analysis, gesture analysis and video annotation. An "articulated" object is defined [1] as a multi-body system composed of at least two rigid components and at most six independent degrees of freedom between any components. A non-rigid, but constrained dependence exists between the components of an articulated object [2]. Examples include the human body, most animals, manipulation robots, long lorries with trailers and many others. The problem of tracking such articulating objects in video is particularly hard from a statistical computer vision perspective owing to the high degree of freedom for motion to the constituent parts of the object. Articulated object tracking is significantly different from multiple target tracking where the motion of each target is independent of the others. In the case of articulated object tracking, the physical links between the various parts of the articulating body impose physical constraints to their motion. For instance, [3] clearly distinguishes articulated object tracking from multiple target tracking by stating that a) if the parts of the articulating body are *independent* then the motion of each part

stays as a manifold orthogonal to the manifolds of the other parts thus allowing the use of multiple target approaches to be applied on them. However, in the case that the parts remain dependent as in typical articulated object tracking situations, such orthogonality and factorization properties cease to exist thus exponentially increasing the computational load due to the curse of dimensionality. There exist a number of research contributions within articulated object motion analysis. Typical methods in articulated object tracking involve three stages. The first stage concerns determining the structural model representation of the body parts. The body parts model can take two distinct forms such as the centralized model and decentralized model. Whilst the centralized model assumes body parts to be connected to each other at particular angular orientations, the decentralized model combines the independent motion of the different parts to one. Different techniques such as the card board person [4], the decentralized probabilistic model based on Markov networks, the loose-limbed model [5], [6], and tree structured model [7], have been proposed to use a decentralized structure representation for articulated objects. The second stage of the process, involves the application of inference methods for articulated motion analysis. These methods can be broadly summarized into deterministic and probabilistic types. According to [3], deterministic methods formulate articulated object tracking as a parameter estimation problem while the probabilistic schemes formulate it as a Bayesian inference problem. Though the deterministic methods are bound to get to an optimal solution, it is highly complex and essentially has high computational demands.

However, the probabilistic methods provide approximate estimation with a simple model under conditions of low time demands. Bayesian inference methods attempt to provide a solution by recovering the motion posterior sequentially at each time instant. In order to cope with non-Gaussian assumptions Monte Carlo simulation is generally used. Techniques have also been proposed to improve the efficiency for the probabilistic approach. For example, in multiple hypothesis tracking algorithms, the salient modes of the motion posteriors are retained for more efficient Monte Carlo simulation [8], [9]. Partitioned sampling is in the spirit of coordinate descent and performs the sampling in a hierarchical fashion [10], [11]. Non-parametric belief propagation [12] algorithm have been proposed in [6], [13] and applied for modeling articulated body motion.

2. CONTRIBUTIONS AND STRUCTURE

The method proposed in this paper combines an evolving population sequential Monte Carlo technique with Variational inference for tracking articulated targets. One of the main novelty of our proposed method stems from simultaneously enforcing the likelihood constraints of the considered parts and the spatial coherence of their neighboring parts. We also integrate an evolving population Monte-Carlo filtering mechanism that allows particles to regenerate both in sampling and re-sampling steps. Our results suggest that the proposed integrated model can considerably improve performance when compared individually to any non-Variational or other sequential Monte-Carlo methods. We begin by describing our method, including the Variational framework and evolving population Monte-Carlo scheme in Section 3. We then perform experiments on some synthetic and real time datasets (Section 4) investigating the model and demonstrate the efficacy of the proposed model in comparison to other techniques. Finally, we present some concluding remarks in Section 5.

3. PROPOSED METHOD

An articulated body model can be represented using various forms of pictorial structures. In this paper, we restrict our discussion to modeling the articulated body as a collection of parts with connections between them. For example, the human body parts can correspond to the head, torso, arms and legs. The number of parts required to model the articulating object, directly relates to

the application under study, the level of accuracy expected and inversely to the computational expense incurred during processing. Throughout this paper, we assume that the silhouette of the articulating object is represented by a model of rectangular regions of specific dimensions as illustrated in 1. The simplest form of representing these parts as a structure is using a form of an undirected graph $G = (V ; E)$, where V is the set of vertices representing the M parts of the body and E the edge between the connected parts in the graph. In each time instance the articulating body in any scene can be expressed in terms of a configuration $X = \{x_1, x_2, \dots, x_M\}$ of different parts containing spatio-orientation parameters. The image observation associated with any part v_k with configuration x_k is denoted by z_k , which is a combination of the colour and edge features of the parts. The combined observation for the articulating body can be represented by $Z = \{z_1, z_2, \dots, z_M\}$. The goal of the tracking problem is to estimate the posterior $p(X|Z)$. As mentioned earlier, the different parts are not independent and each part distinctly interacts with a specified set of neighbors. This graphical representation of an articulated body is analogous to a Markov network.

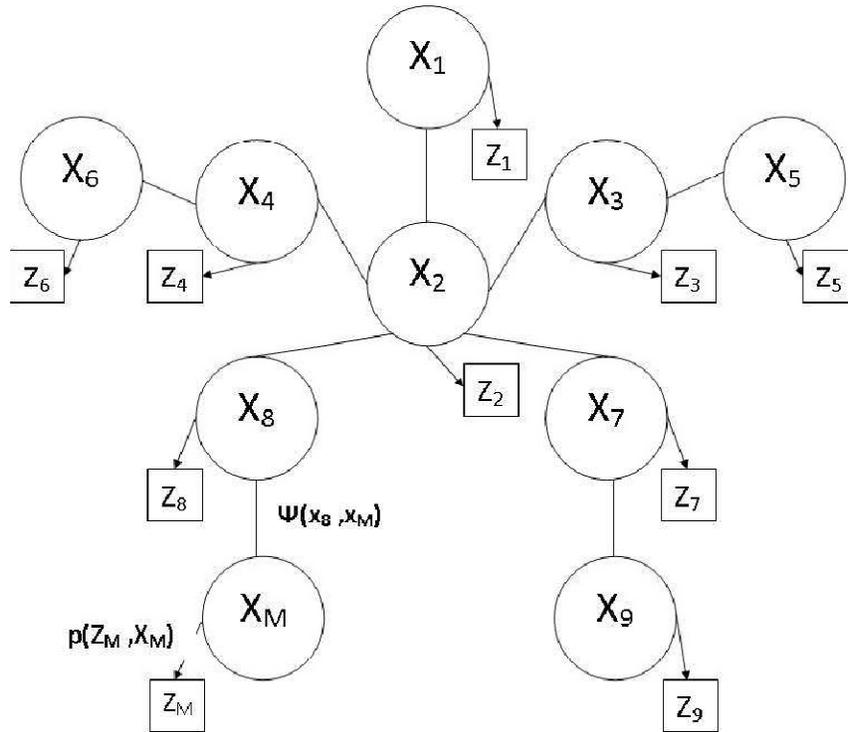


Figure 1. The Markov network for an articulated body

Given the undirected graph of the articulated body X , $p(X)$ can be factorized as:

$$p(X) = \frac{1}{Z_c} \prod_{c \in \zeta} \psi_c(X_c)$$

Where c is a clique in the set of cliques ζ of the undirected graph, X_c is the set of nodes associated with the clique and $\psi_c(X_c)$ is the potential function associated with this clique, and Z_c is a normalization function. Two different types of cliques are common in articulated objects. First order cliques ζ^1 , whose potential function ψ_i , where i denotes a part of the body model, provides the local prior for x_i and the second order clique, ζ^2 , has a potential function ψ_{ij} , where i and

j are neighbourhood parts, that denotes the spatial coherencebased on model constraints. Therefore, $p(X)$ reduces to:

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

The likelihood model denoted by $p(Z|X)$ can be factorized as:

$$p(Z|X) = \prod_{i=1}^N p_i(z_i|x_i)$$

According to the theory of Bayesian inference, the posterior $p(X|Z)$ can be estimated using the following relationship.

$$p(X|Z) \propto p(Z|X)p(X)$$

It is apparent that the posterior of each part $p(x_i|Z)$ will be affected by its local prior ψ_i , the likelihood and the constraints from its neighbors. Probabilistic reasoning consists in computing posterior distributions over hidden variables given visible variables, and estimating model parameters. There are different principled algorithms for approximate inference and learning in vision applications, including iterative conditional modes, Gibbs sampling, variational techniques, structural variational techniques, and belief propagation algorithm. Variational inference methods provide Bayesian inference to applications represented using graphical models. The main principle of behind the variational approximation is to estimate a variational distribution $Q(X)$ to approximate the posterior distribution $p(X|Z)$ such that the Kullback-Leibler (KL) divergence between the variational and the posterior distributions is minimized. Variational methods help in reducing an estimation problem to an optimization problem, that is,

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

In the case of articulate object tracking framework, it is effective to fully factorize the variational distribution $Q(X)$ among the M parts of the object as follows.

$$Q(X) = \prod_{i=1}^M Q_i(x_i)$$

Where $Q_i(x_i)$ only relies on the configuration x_i . According to [3], The KL divergence of each of the variational distribution Q_i can be simplified using the factorization of $p(X)$ and $p(Z|X)$. Therefore, $Q_i(x_i)$ can be reduced and shown that:

$$Q_i(x_i) = \frac{1}{Z_i} p_i(z_i|x_i) \psi_i(x_i) M_i(x_i)$$

Where,

$$M_i(x_i) = \exp \left\{ \sum_j \int_{x_j} Q_j(x_j) \log \psi_{ij}(x_i, x_j) \right\}$$

Where Z'_i is a constant and j refers to a part in the neighborhood of part i . The above Bayesian inference problem can be solved using two distinct approaches. The first involves assuming that all the distributions in the Markov network illustrating the articulated object are Gaussian. This allows obtaining a closed-form implementation of the equations above. However, video applications require the computation of features such as color, texture, edges etc. for the likelihood function, which are often contaminated by clutter thus breaking Gaussianity assumptions. This results in the second technique of Bayesian inference based on Monte Carlo methods. The authors of [3] present a detailed Mean Field Monte Carlo technique that helps in evaluating efficient solutions. In this paper, we propose a modified solution to the Monte Carlo based on population methods and likelihood combined variational inference. The proposed Monte-Carlo framework includes three main stages: 1) Initialisation, 2) Resampling and 3) Sampling. In the subsections below we present a detailed overview of the different stages of the proposed Evolving Monte-Carlo (EvMC) filtering mechanism.

3.1. Initialization

Let us begin by assuming the update for one body part i with configuration x_i , which further can be generalised to the other parts of the articulated object. The basic underlying principle behind EvMC filtering is to iteratively update the set of optimal variation distributions $Q_i(x_i)$ of that body part i and approximate it to the posterior density $p(x_{i,k}|Z_k)$ at any discrete time instant k , that is,

$$Q_{i,k} \sim \left\{ s_i^{(n),k}, w_i^{(n),k} \right\}_{n=1}^N$$

Where s_i denotes the samples, w_i their corresponding weights and N , the total number of samples. At the first time instant, i.e., $k = 1$ and iteration $t = 1$, samples $\left\{ s_{i,t}^{(n),k} \right\}_{n=1}^N$ are drawn from any proposal distribution $\pi_{i,k}$. Estimate the weights of samples using: $w_{i,t}^{(n),k} = p(z_{i,k}|s_{i,t}^{(n),k})$. The likelihood measurement $p(z_{i,k}|s_{i,t}^{(n),k})$ is computed using the color measurement cue. Other cues, including the texture, edge or motion and joint variations of any of them can be used. The measurement is extracted from the image as the color histogram q_t , computed inside the image regions specified by the states $s_{i,t}^{(n),k}$. The likelihood function is estimated as a Gaussian density of the Kullback Leibler distance between the colour histogram of the different channels in the RGB colour space between the reference histogram q^* of the part to be tracked and the histogram q_t computed from $z_{i,k}$ in regions specified by each sample $s_{i,t}^{(n),k}$, that is,

$$p(z_{i,k}|s_{i,t}^{(n),k}) \propto \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{[KL(q^*, q_t)]^2}{2\sigma^2} \right\}$$

The weights are then normalized as $\frac{w_{i,t}^{(n),k}}{\sum_{n=1}^N w_{i,t}^{(n),k}}$. Following the initialization step, iterative resampling and sampling is performed as below.

3.2. Resampling

The resampling step is performed in order to overcome the problem of degeneracy. However, degeneracy is commonly measured using the Effective Sample Size (ESS). ESS is estimated as $\left(\sum_{n=1}^N (w_{i,t}^{(n,k)})^2\right)^{-1}$ and compared against a threshold. If the ESS value is lower than the threshold, resampling is performed and the weights are updated to $\frac{1}{N}$. A number of resampling methods have been proposed in the literature including, multinomial, residual, stratified, systematic etc. The *multinomial re-sampling* procedure involves drawing uniform distributed samples and applying the inversion method [14]. *Residual resampling*, otherwise known as remainder resampling, is an efficient technique of decreasing the variance of the generated samples due to resampling [15]. *Stratified resampling* is based on ideas of survey sampling and involves pre-partitioning the uniform distribution interval into disjoint sets and drawing independently and henceforth applying multinomial resampling approach [15]. Finally, *systematic resampling* improves the stratified approach further by deterministically linking all the variables in the subintervals. Most of the techniques aforementioned are sensitive to the order in which the particles are presented, i.e., a simple permutation can change the resampling process. Though these methods are widely used in the literature, they lack thorough theoretical analysis of their behavior, apart some theoretical works [15].

3.3. Sampling

Following the resampling process, a transition prior model is applied to regenerate new samples using a specified Kernel function. This is usually done by applying a kernel function in the following way: $s_{i,t+1}^{(n,k)} = K(s_{i,t}^{(n,k)}, \cdot)$. The weights of samples are recomputed but now taking into account the influence of the neighbouring part j . This influence is characterized as a message function such as:

$$m_{ij}^{(n,k)} = w_{j,t}^{(n,k)} \cdot \log \psi_{ij}(s_{i,t+1}^{(n,k)}, s_{j,t}^{(n,k)})$$

The overall weight for part i is updated using

$$w_{i,t+1}^{(n,k)} = p(z_{i,t+1} | s_{i,t+1}^{(n,k)}) \exp \sum_j m_{ij}^{(n,k)}$$

Perform iteration over t and then proceed to the next time step $k+1$. The prior model is re-evaluated by computing the probability $p(x_{i,k+1} | x_{i,k})$ and is updated using, $p(x_{i,k+1} | x_{i,k}) \sim \hat{s}_{i,t}^{(n,k-1)}$

3.4. Genetic Kernel

In our proposed framework we employ a genetic kernel in the form of evolutionary cross-over, mutation and exchange steps to evolve a better population of samples at each cycle. In the following subsections we briefly describe the used genetic operators.

1) *CrossOver*: In the proposed framework the state vector is represented as a string of bits. The crossover point l_c is a random point on the string of bits of length l . The crossover operator cannot be applied to all parts of the state vector. Some parts of the state vector may not undergo any changes and thus for such components, the probability of crossover ρ_c is zero. This leaves the

crossover being operated only on components that are expected to undergo random changes. The crossover operator functions on two distinguished offsprings (paired particles), for example, $x_{s,\tau}$, $x_{q,\tau}$. The algorithm for crossover is described below:

- Draw a uniform random number u_c for every component $u_c \approx U(0, 1)$ and if $u_c \leq \rho_c$ then perform crossover
- Estimate a crossover point l_c based on a uniform random integer between 1 and the length of the component l
- Generate two offsprings:

$$x_{s,\tau} = (x_{1s}, x_{2s}, \dots, x_{(l_c-1)s}, x_{l_cq}, \dots, x_{dq})$$

$$x_{q,\tau} = (x_{1q}, x_{2q}, \dots, x_{(l_c-1)q}, x_{l_cs}, \dots, x_{ds})$$

where ds and dq refers to the length of the samples $x_{s,\tau}$ and $x_{q,\tau}$ respectively. Then in the crossover operation, performed with the two offsprings $x_{s,k}$ and $x_{q,k}$, the particle weight can be expressed in the form

$$w_k^{cr,(i)} = \frac{p(z_k | x_{s,k}^{(i)}) p(z_k | x_{q,k}^{(i)}) p(x_{s,k}^{(i)} | x_{s,k-1}^{(i)}) p(x_{q,k}^{(i)} | x_{q,k-1}^{(i)})}{q(x_{s,k-1}^{(i)} | z_{k-1}) q(x_{q,k-1}^{(i)} | z_{k-1})}$$

Then the recursive weights can be written as

$$w_k^{cr,(i)} = w_{s,k-1}^{cr,(i)} w_{q,k-1}^{cr,(i)} \mathcal{L}_s(z_k, x_{s,k}^{(i)}) \mathcal{L}_q(z_k, x_{q,k}^{(i)})$$

Here $\mathcal{L}_s^{(i)}(z_k, x_{s,k}^{(i)}) = p(z_k | x_{s,k}^{(i)})$ is the likelihood function for the s^{th} offspring, $\mathcal{L}_q(z_k, x_{q,k}^{(i)}) = p(z_k | x_{q,k}^{(i)})$ is the likelihood function of the q^{th} offspring; $w_{s,k-1}^{cr,(i)}$ and $w_{q,k-1}^{cr,(i)}$ are the weights at $(k-1)^{th}$ time instant, for the s^{th} and q^{th} offspring, respectively.

2) *Mutation*: A probability of mutation ρ_m is initially defined for each component. Such a probability is chosen in order to make sure that components that need no stochastic fluctuations could be prohibited from undergoing mutation operation. For such components, the probability of mutation ρ_m is considered zero. The components are assumed as a vector string of binary units. According to the proposed mutation mechanism,

- Draw a uniform random number u_m for every component $u_m \approx U(0, 1)$ and if $u_m \leq \rho_m$ then perform mutation
- Estimate a mutation point l_m based on a uniform random integer between 1 and length of the component l
- Flip the mutation point l_m The weights are of the form:

$$w_k^{mutation,(i)} = \frac{p(x_k^{(i)} | x_{k-1}^{(i)})}{U q(x_k^{(i)} | x_{k-1}^{(i)})}$$

Where, U is a uniform random number. During mutation, samples that undergo mutation are mutually independent. Therefore, the updated proposal distribution at time k is a factor of the proposal distribution at the previous iteration $k-1$.

3) *Exchange*: Consider two independent chains of samples, for example, $x_{s,k}$ and $x_{q,k}$. For their target distributions π_s and π_q , respectively, the swap of information between these two chains can be performed with a Metropolis-Hastings step. The swap occurs with probability $\min\{1, A\}$, where

$$A = \frac{\pi_s(x_{q,k})\pi_q(x_{s,k})}{\pi_s(x_{s,k})\pi_q(x_{q,k})}$$

4. RESULTS AND ANALYSIS

To track articulated objects we consider people (together with their multiple body parts) in video from the CAVIAR [16] data set. The CAVIAR dataset contains over 80 video sequences with one or more targets moving in real-time scenarios. We have chosen 10 videos from the data set at varying levels of complexity in terms of the motion characteristics, occlusion and clutter (due to illumination changes). We further divided the 10 selected sequences in 20 short clips with a single target. The tracked people have been manually annotated in these clips using a maximum 10 point model consisting of the centers of the head (1 point), torso (1 point), arms (2 points for each arm) and legs (2 points for each leg). We would like to particular highlight that the frames are of resolution 384 x 288 captured at 25 frames per second. The scales of the human targets are fairly small in comparison with the resolution of the image and thus in some of the images a 10 point model gives a dense set of labeled ground-truth which is unusable; therefore in such sequences we use a 6 point model. We use the method of [17], [18] to initialize the multiple body parts composed multiple targets in first frame of each video using the pictorial structure model. For all our experiment described below unless mentioned otherwise, the number of Monte Carlocycles is fixed to 100 and the number of particles used is 500. All our experiments are conducted on an Intel Duo Core processor with 4GB RAM. To evaluate the performance of the models, we compute the root mean square error distance (RMSE) between the estimated center point of every part and its manually labeled counterpart.

4.1. Comparison of Proposed with Related Techniques

One of the main novelties of the proposed method is the integration of the variational inference based on measuring the KullbackLiebler divergence into the evolving population Monte Carlo framework. In the following experiment we compare our method to the conventional evolving population MCMC filter as in [19] to study the effect of variational inference on tracking multiple body parts of human targets and in addition compare our method to a non-MCMC technique in the form of the generic particle filter with joint probabilistic data association filter (JPDA PF) proposed in [20], [21]. We summarize the performance of our method in the Table I. In terms of the computational demand, our algorithm takes 3742 msec on an average per image frame as against the combined EPMCMC and PDA methods that take 2169 msec(implemented in MATLAB).

4.2. Effect of the number of parts tracked

In this subsection, we investigate the effect of increasing the number of parts of each target that is used for tracking. In the lower levels of the model, we have a smaller collection of more salient parts representing the target followed by increasing number of lesser important parts. In the following experiment, we examine the impact of increasing the number of parts representing the target by comparing the model with level one (1 part: torso), level two (2 parts: head and torso), level three (6 parts: head, torso, two hands, and two legs), and level four (10 parts-head, torso, four for hands, and four for legs). In order to demonstrate the effectiveness of the variational approach in combining the effect of neighboring body parts

Video	Proposed	Baseline	JPDAPF
B2	4.3341 5.1123	6.9133 7.2309	4.6612 5.4313
BWW1	5.0098	7.0384	6.5367
FC	4.9833	6.7847	5.9749
LB	5.7683 4.9987	6.2053 5.7574	5.9930 6.0393
LBx	6.0139	6.8932	7.1938
RFF	5.1342	8.8402	7.1305

Table 1. Tabular description of the chosen video clips (B2- browse2, BWW1- Browse While Walking1, FC- Fight Chase, FR1- Fight Runway1, LB- Left Bag, LBx- Left Box, RFF- Rest Fallen Floor) and comparison of combined RMSE between variational model (proposed), EPMCMC model (baseline) and generic particle filter framework (JPDA PF)

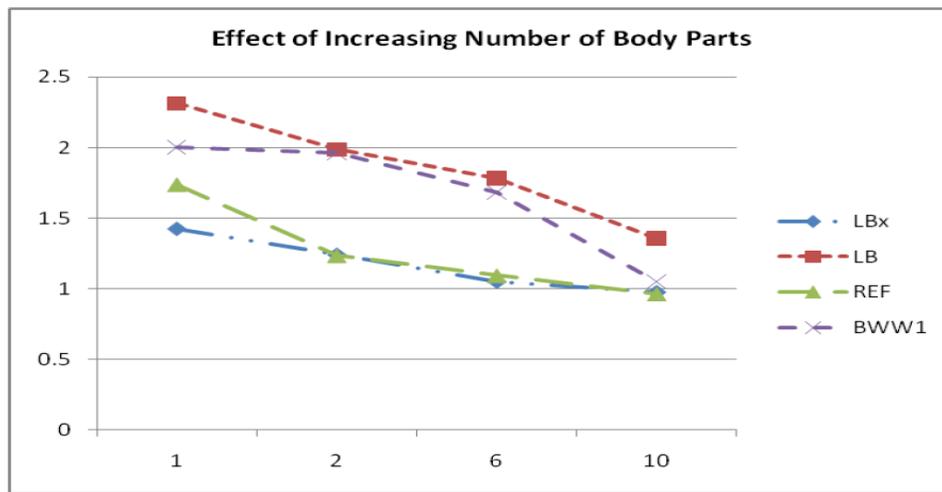


Figure 2. RMSE of the Torso (y-axis) versus the Number of Body Parts (x-axis) on Different Sequences

4.3. Effect of the number of Monte Carlo cycles

Finally, we also explore the effect of increasing the number of Monte-Carlo cycles on the accuracy of our tracking procedure. We record our finding as a plot of the increasing number of Monte-Carlo cycles against cumulative RMSE error in Figure 3 and against time in Figure 4. Our results suggest that with increase in the number of Monte-Carlo cycles the accuracy increases but remain nearly constant after 200 cycles. As it can be clearly visualized, with increasing Monte-Carlo runs, the time complexity of the algorithm also increases. Therefore, a good trade-off needs to be set for balancing the increased accuracy against computational demand.

5. CONCLUSIONS

We have proposed an innovative method for variational inference combined with evolving population Monte-Carlo for robust and accurate target tracking. The evolving population filter introduces variety in the population of particles by combining them in both the sampling and resampling steps using constrained genetic operations. We measure the sensitivity and robustness of our proposed framework against key system parameters such the the number of body parts tracked and increasing number of Monte-Carlo cycles and have reported results that outperform

other baseline methods on standard datasets. Our future work will focus on combining the proposed methodology with appropriate data association schemes for reliable multiple articulated target tracking.

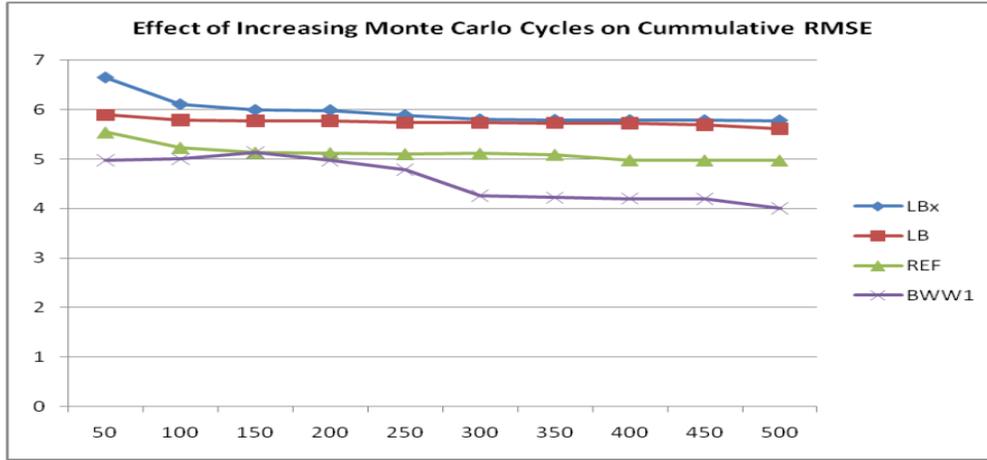


Figure 3. Combined RMSE of the Body Parts (y-axis) versus the Number of Monte Carlo Cycles (x-axis) on Different Sequences

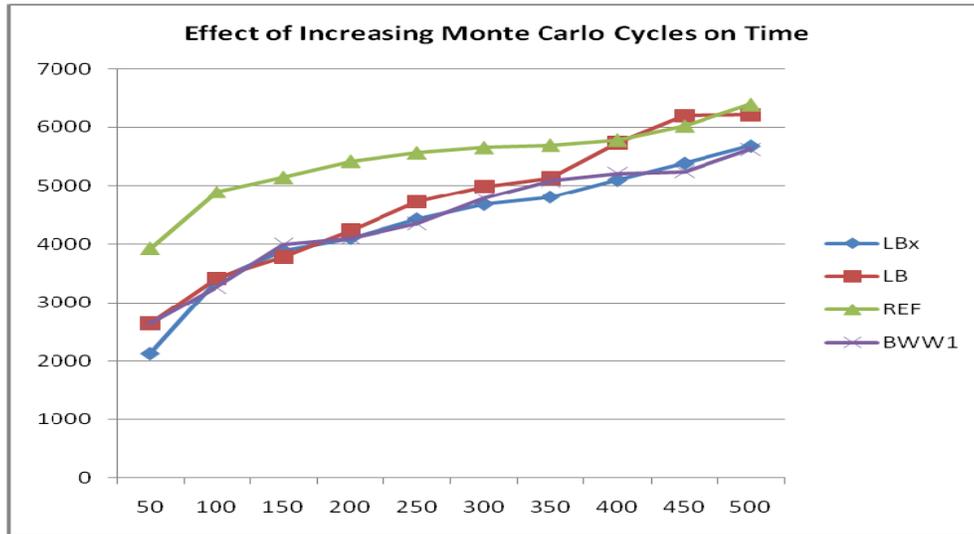


Figure 4. Mean Time for Tracking (y-axis) versus the Number of Monte Carlo Cycles (x-axis) on Different Sequences

REFERENCES

- [1] A. Comport, E. Marchand, and F. Chaumette, "Complex articulated object tracking," in LNCS from the International Workshop on Articulated Motion and Deformable Objects, Palma de Mallorca, Spain, Sept. 2004.
- [2] L. Sigal and M. Black, "Predicting 3D people from 2d pictures," in Proc. of the IVth Conf. on Articulated Motion and Deformable Objects, 2006.
- [3] G. Hua and Y. Wu, "A decentralized probabilistic approach to articulated body tracking," *Computer Vision and Image Understanding*, Special issue on spatial coherence for visual motion analysis, vol. 108, no. 3, pp. 272–283, 2007.
- [4] S. Ju, M. Black, and Y. Yacoob, "Cardboard people: a parameterized model of articulated image motion," Proc. of the IEEE International Conference on Face and Gesture Recognition, 1996, pp. 38–44.
- [5] A. O. Balan and M. J. Black, "An adaptive appearance model approach for model-based articulated object tracking," in Proc. of the CVPR, 2006.
- [6] L. Sigal, M. Isard, B. Sigelman, and M. Black, "Attractive people: Assembling loose-limbed models using non-parametric belief propagation," in *Advances in Neural Information Processing Systems (NIPS)*, 2003.
- [7] P. Felzenswalb and D. Huttenlocher, "Efficient matching of pictorial structures," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2000.
- [8] R. Rubinstein, *Simulation and the Monte Carlo Method*. New York, Wiley, 1981.
- [9] E. Poon and D. Fleet, "Hybrid Monte Carlo filtering: Edge-based tracking people," in Proc. of the IEEE Workshop on Motion and Video Computing, Orlando, Florida, Dec. 2002, pp. 151–158.
- [10] T. Cham and J. Rehg, "A multiple hypothesis approach to figure tracking," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2, 1999, pp. 239–245.
- [11] J. MacCormick and M. Isard, "Partitioned sampling, articulated objects, and interface-quality hand tracking," in ProcEurop. Conf. Comp. Vision, 2000.
- [12] X. Lan, S. Roth, D. Huttenlocher, and M. Black, "Efficient belief propagation with learned higher-order Markov random fields," in Proc. of the ECCV, 2006.
- [13] E. Sudderth and W. Freeman, "Dsp applications [Signal and Image Processing with Belief Propagation]," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 114 – 141, 2008.
- [14] J. Hol, T. Sh`on, and F. Gustaffsson, "On resampling algorithms for particle filters," in Proc. Nonlinear Statistical Signal Processing Workshop, Cambridge, United Kingdom, September, 2006.
- [15] A. Douc, E. Moulines, and O. Capp`e, "Comparison of resampling schemes for particle filtering," in Proc. of the 4th International Symp. on Image and Signal Processing and Analysis, 2005, pp. 64–69.
- [16] "CAVIAR test case scenarios, available at <http://homepages.inf.ed.ac.uk/rbf/caviardata1/>," 2005. [Online]. Available:<http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/>
- [17] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Tracking people by learning their appearance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 65–81, 2007.
- [18] D. Ramanan and D. Forsyth, "Finding and tracking people from the bottom up," in Proc. of CVPR, 2003. [Online]. Available:citeseer.ist.psu.edu/ramanan03finding.html
- [19] H. Bhaskar, L. Mihaylova, S. Maskell, and S. Godsill, "Evolving population Markov Chain Monte Carlo particle filtering," Journal paper (in preparation).
- [20] M. Jaward, L. Mihaylova, N. Canagarajah, and B. D., "Multiple objects tracking using particle filters in video sequences," in Proc. of the IEEE Aerospace Conf. Big Sky, MT, USA, 2006.
- [21] Jaward, M., Mihaylova, L., Canagarajah, N., and Bull, D., "A data association algorithm for multiple objects tracking in video sequences," in Proc. of the IEE Seminar on Target Tracking: Algorithms and Applications, Birmingham, UK, 2006, pp. 131–136.

AUTHORS

Kartik Dwivedi is a Senior Undergraduate student at the department of Electronics and Electrical Engineering, Indian Institute of Technology, Guwahati. He was a visiting research intern at Khalifa University, UAE during summer 2013. His current research interests include computer vision and machine learning.



Harish Bhaskar received Dr. Harish Bhaskar is a Chief Engineer at the Advanced Technology Group at Samsung Research Institute Noida. After a successful academic career as an assistant professor of computer engineering at Khalifa University of Science, Technology and Research, since 2009, Dr. Bhaskar joined Samsung India in the mid-2013. Before moving to the UAE for an academic career, Dr. Bhaskar worked as a Researcher at Manchester and Lancaster Universities, U.K. Dr. Bhaskar has been actively associated with several European Research institutes and the Ministry of Defense UK. His research interests are in the field of computer vision, image processing, data mining, visual cryptography, artificial intelligence and robotics.



Artur Loza received the M.Sc. degree in electrical engineering from Wroclaw University of Technology, Wroclaw, Poland, in 2000, and the Ph.D. degree from the University of Bristol, Bristol, U.K., in 2006. He has also been a Research Associate at the University of Bristol, UK and Khalifa University, UAE. In 2010, he was awarded an EU-funded two-year scholarship to conduct his research at Shanghai JiaoTong University, Shanghai, China. His current research interests include statistical and multiresolution image processing, including image fusion and enhancement, and video object tracking.

