Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering 16 (2011) 291 – 297
Procedia Engineering
www.elsevier.com/locate/procedia

International Workshop on Automobile, Power and Energy Engineering

Behavior Extraction from Examples Using Federate MCMC-Based Particle Filtering

Meng Li[a], Jia-Hong Liang[b],a*

[a,b]*College of Mechanical Engineering and Automation, National University of Defence technology, Changsha 410073 , P.R.China*

Abstract

Data-driven methods of simulating a crowd of virtual humans that exhibit behaviors imitating real human crowds play an important role in crowd simulation. In this paper, we propose a Bayesian framework for the extraction of real human's behaviors which exhibit interactions in their daily life using multiple fixed cameras. The described Markov chain Monte Carlo particle filter can effectively deals with interacting targets which are influenced by the proximity and behaviors of other targets. In this paper, we use a Markov random field motion prior combing with a federate filter algorithm which treats the observations discriminatorily to substantially improve the tracking of a fixed number of interacting targets. Simultaneously, we replace the traditional importance sampling step with MCMC sampling step to get over the vast computational requirements for large numbers of targets. i.e., we focus on the data fusion and the behavior recognition process. Finally, experimental results demonstrate that the proposed Bayesian framework deals efficiently and effectively with extractions of interacting behavior.

Keywords: behavior extraction, MCMC-based particle filter, Markov random field motion prior, federate filter

1. Introduction

The behaviors of human crowds in the real world characterized with trajectory, velocity, and interaction vary significantly depending on time, place, and many other factors [1]. This work is concerned with the problem of behavior extraction from sequential images recorded by multiple cameras. The extracted behaviors can be used to produce many desired styles in data-driven methods of simulating the crowd. Learning human behaviors from videos is a challenging problem because the motion of each individual is influenced by many factors the most important one of which is the interactions between proximate targets. We assume that each target motion in the video is Markovian and targets do not overlap with each other vertically. Our objective is to correctly detect entering and leaving targets and obtain a

* *E-mail address*: mengshuqin1984@163.com

1877-7058 © 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2011.08.1086

record of the trajectories of targets over time, maintaining a correct, unique identification for each target throughout. Tracking multiple targets which are identical in appearance becomes significantly more challenging when the targets interact. Methods that appropriately deal with this difficult issue are useful for applications where many interacting targets need to be tracked over time. In particular, they have important implications for vision-based tracking of human behaviors which can be used to improve the realism of crowd simulation [1,2].

There is a great potential to extract the real people's behaviors from video sequences by use of computer vision algorithms. A number of classical approaches have been proposed to the problem of multiple hypothesis tracker (MHT) and the joint probabilistic data association filter (JPDAF) [2]. These multi-target tracking algorithms which including nearest neighbor tracking in [3], the MHT in [4], and the JPDAF in [5] and [6] have been extensively used in the context of computer vision. However, the foregoing methods do not take into account targets' interactions. There is an abundance of literature devoted to the PF approach to multi-object tracking (MOT). Methods using a single-object state-space model are usually computationally inexpensive. A shortcoming of this approach is that identities and interactions between objects can not be easily modeled in formal (and algorithmic) terms [7]. Recently, A joint state-space model was proposed to efficiently track a fixed number of interacting objects using a sampling method that combines a PF formulation with MCMC sampling in [2]. This approach addressed the problem of interaction as well, by defining a pair-wise Markov Random Field (MRF) prior in the dynamical model, which is more computationally tractable than other methods [8]. However, the aforementioned approaches only use a fixed camera to extract the features of targets.

Although multiple object tracking has been extensively studied in computer vision, automatic behavior extraction is still a challenge in practice. In this paper we address the problem of interacting targets by employing federate PF on the basis of previous studies. The federated filter fused the outputs of local filters to generate the global estimation. The behaviors of interacting targets are modeled by a motion model based on MRF. Considering the computational complexity of federate filter, we use MCMC-based sampling step to replace traditional importance sampling. And the observation model is represented by binary information picked-up from background subtraction, together with the foreground and background color features. A specific disposal to the motion model and observation model is proposed to solve the problem of information fusion. In our algorithm, the MCMC-based filter combing MRF dynamical model serves as the local filter, the federated filter is used to fuse outputs of all local filters, and the global filter result is obtained. The advantages of our approach are obvious that it combines the benefits of MCMC-based filter and federate filter, the most valuable one of which is our approach's powerful fault detection, isolation and recovery abilities.

The remainder of this paper is organized as follows. Section 2 reviews the work related to Bayesian multi-object tracking combing with the dynamic model of interacting targets. Section 3 discusses the MCMC-based sampling step. Section 4 covers, respectively, the fusion of data from multiple cameras. Section 5 provide some remarkable experimental validations and conclusions

2. Bayesian multi-object tracking and interactive model

2.1. traditional independent PF

Bayes filter is a classical approach in tracking multiple targets. The posterior distribution $P\left(X_{i,t} \mid Z_j^t\right) i \in 1...n;\ j \in 1...m$ is recursively updated give all observations $Z_j^t = \left\{Z_j^1..Z_j^t\right\}$ up to and including time t according to (1). *n* is the number of targets, and *m* is the number of cameras.

$$P\left(X_{i,t} \mid Z_j^t\right) = kP\left(Z_j^t \mid X_{i,t}\right)\int_{X_{i,t-1}} P\left(X_{i,t} \mid X_{i,t-1}\right)P\left(X_{i,t-1} \mid Z_j^{t-1}\right) \quad (1)$$

where $P\left(X_{i,t} \mid X_{i,t-1}\right)$ is the dynamical motion model that predicts the state $X_{i,t}$ of the *i*-th targets at time t given the previous state $X_{i,t-1}$.

The likelihood $P\left(Z_j^t \mid X_{i,t}\right)$ denotes the observation likelihood, the probability we observed the *j*-th observation given the state $X_{i,t-1}$ at t. And *k* is the normalization constant.

Behavior extraction is a non-linear, non-Gaussian problem, in which PFs are accomplished. The essence of filtering distribution is that the posterior at the previous time step is approximated by a set of weighted particles $\left\{\left(X_{i,t-1}^{(r)}, w_{i,t-1}^{(r)}\right), r = 1,...N\right\}$, where $X_{i,t}^{(r)}$ and $w_{i,t}^{(r)}$ denote the *r*-th sample and its associated weight at each time-step. We can rewrite (1) as:

$$P\left(X_{i,t} \mid Z_j^t\right) \approx kP\left(Z_j^t \mid X_{i,t}\right)\sum_r w_{i,t-1}^{(r)} P\left(X_{i,t} \mid X_{i,t-1}^{(r)}\right) \quad (2)$$

using importance sampling. And for the current time-step we draw *N* samples $X_{i,t}^{(s)}$ from a proposal distribution

$$X_{i,t}^{(s)} \sqcup q\left(X_{i,t}\right) = \sum_r w_{i,t-1}^{(r)} P\left(X_{i,t} \mid X_{i,t-1}^{(r)}\right) \quad (3)$$

The weights are then computed as $w_{i,t}^{(r)} \propto P\left(Z_j^t \mid X_{i,t}^{(r)}\right)$

2.2. Dynamic model for interactive targets

Our approach to model interactive targets introduces a capable motion model, based on MRFs. The MRF is defined on an undirected graph (*V, E*), of which interactive targets define the nodes. Interactions are defined on contiguity cliques. The state of one target is influenced by the constrains placed by the MRF prior (Fig. 1). The pair-wise interaction between two neighbours can be writed as:

$$P\left(X_t \mid X_{t-1}\right) \propto \prod_i P\left(X_{i,t} \mid X_{i,t-1}\right)\prod_{i,j\in E} \phi\left(X_{i,t} \mid X_{j,t}\right) \quad (4)$$

Where $\phi\left(X_{i,t} \mid X_{j,t}\right)$ are the pair-wise interaction potentials. Since it is easier to specify the interaction potential in the log domain, $\phi\left(X_{i,t} \mid X_{j,t}\right)$ is expressed by means of the Gibbs distribution [2]:

$$\phi\left(X_{i,t} \mid X_{j,t}\right) \propto \exp\left(-g(X_{i,t}, X_{j,t})\right) \quad (5)$$

where $g(X_{i,t}, X_{j,t})$ is a penalty function. In the application of behaviour extraction, $g(X_{i,t}, X_{j,t})$ depends only on the number of pixel overlap between the targets' body rectangle (Fig. 2).

Figure 1 Interaction potentials. (Left : observe from horizontal camera) A pair-wise MRF is built among individual pairs. Proximate individuals influence each other stronger than distant individuals. (Right : observe from vertical camera)

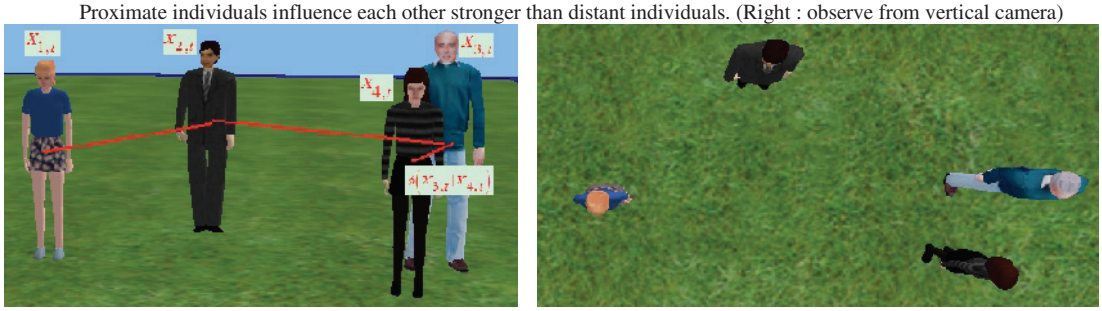

Figure 2 Overlapping individuals. (Left : observe from horizontal camera) A pair-wise MRF is built among individual pairs. Individuals ($x_{3,t}$, $x_{4,t}$) overlapping with each other are penalized. (Right : observe from vertical camera)

In [7], the observations only come from horizontal cameras. It will bring misjudgements on estimating the overlapping region when two individuals overlap from horizontal angle of view, but the distance between them goes beyond each other's influence area. In [2], although the observations come from vertical cameras, it aims at ants' behaviors. Because the ants can overlap with each other vertically, this phenomenon rarely happens in human beings. This paper estimates the penalty function depending on vertical and horizontal observations, e.g., the vertical observations revise the judgement. So when we design the federate filters, vertical observations and its' observation model are regarded as the master filter.

3. An MCMC-based MRF particle filter

MCMC-based PFs generate a sequence of samples from a Markov chain of which the stationary distribution corresponds to the target distribution after an adequate time [7]. In fact, this section pays attention to replacing the inefficient importance sampling step of a straightforward PF implementation by a more efficient MCMC sampling step, which drastically improves tracking results and reduces the computation burden. Instead of using weighted samples obtained using importance sampling, the posterior is represented as a set of unweighted samples. Thereby, on the basis of (2) ~(4), we obtain (6):

$$P\left(X_t \mid Z_j^t\right) \approx kP\left(Z_j^t \mid X_{i,t}\right) \prod_{i,j\in E} \exp\left(-g(X_{i,t}, X_{j,t})\right) \sum_r \prod_i P\left(X_{i,t} \mid X_{i,t-1}^{(r)}\right) \tag{6}$$

A Metropolis-Hastings (MH) sampler is used to sample in MCMC techniques at each time step to generate a set of samples from a proposal distribution density $q\left(X^* \mid X\right)$, where X and X^* denote the current and proposed states. The acceptance ration for the *j*-th observation is computed as:

$$\alpha = \min\left(1, \frac{p\left(X_t^* \mid Z_j^t\right) q\left(X_t \mid X_t^*\right)}{p\left(X_t \mid Z_j^t\right) q\left(X_t^* \mid X_t\right)}\right) \tag{7}$$

In MH algorithm, if $\alpha \geq 1$, $X_{i,t}^*$ is accepted as the new particle. Otherwise, X^* is only accepted with probability α, and reject otherwise. In the latter case, $X_{i,t}$ is accepted as the new particle. The proposal density q is computed as [2]:

$$q\left(X_t^* \mid X_t\right) \overset{\Delta}{=} \begin{cases} \frac{1}{n} q\left(X_{i,t}^* \mid X_{i,t}\right) & \text{if } X_{-i,t}^* = X_{-i,t} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

The step of MCMC-based PF which is familiar to many related researchers is not list in this paper.

4. Framework of data fusion

The structure of federate MCMC-based MRF particle filtering is show in Fig.3, in which vertical observations and horizontal ones are treated discriminatorily.

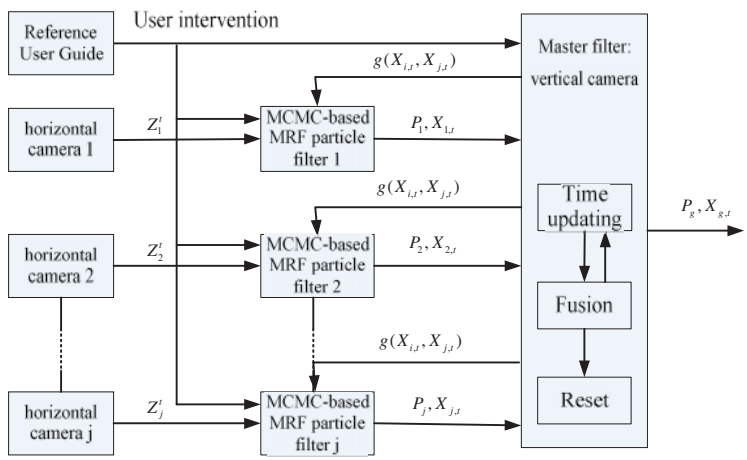

Fig.3 The structure of federate MCMC-based MRF PF

Federate MCMC-based MRF PF which is composed of MCMC-based MRF PF and the federated filter is based on the nonlinear state model and the information fusion theory. First of all, MCMC-based MRF PF is employed to estimate local filters and their estimation results are acquired. Then, local filters' estimations are fused to acquire the system's global optimal estimate by the master filter [9].

In the structure of Federate MCMC-based MRF PF, local filters run in parallel and accept external sensor data from horizontal cameras. Outputs of local filters are fused via master filter which provides feedbacks $g(X_{i,t}, X_{j,t})$ to local filters. The feedbacks are closely related to the observations from vertical cameras. Estimates yielded by the federated filter X_g are globally optimal or conservatively suboptimal [10].

In addition, our system offers a chance for user interventions which provide a putting right capacity from users. Our user guide allows the user to refine the (possibly flawed) tracking result, for example, the "hijacking" problem.

5. Experimental Validation and Conclusion

We adopt observations which are binary and color measurements proposed in [7], $Z_j^t = \left(Z_{j,t}^b, Z_{j,t}^c\right)$. Binary observations $Z_{j,t}^b$ are extracted using background removal techniques. Accordingly, the color observations $Z_{j,t}^c$ are made in HS space. The observation likelihood is defined as:

$$p\left(Z_j^t \mid X_t\right) = p\left(Z_{j,t}^c \mid Z_{j,t}^b, X_t\right) p(Z_{j,t}^b \mid X_t) \tag{9}$$

For an image sequence divided into foreground and background pixels, the observations can be expressed as $Z_{j,t} = \left(Z_{j,t}^F, Z_{j,t}^B\right)$, then $p\left(Z_{j,t}^c \mid Z_{j,t}^b, X_t\right) = p\left(Z_{j,t}^{c,F} \mid Z_{j,t}^{b,F}, X_t\right) p\left(Z_{j,t}^{c,B} \mid Z_{j,t}^{b,B}, X_t\right)$ and $p(Z_{j,t}^b \mid X_t) = p(Z_{j,t}^{b,F} \mid X_t) p(Z_{j,t}^{b,B} \mid X_t)$.

Table 1 is the experimental results which are compared with MCMC-based MRF PF algorithms.

Table 1. experimental results

Tracker	samples	failures	Per target error in pixel
MCMC	1000	29	2.01 ± 0.82
MCMC	2000	26	1.86 ± 0.94
federate MCMC-based MRF PF	1000	20	1.95 ± 0.73
federate MCMC-based MRF PF	2000	15	1.29 ± 0.81

The experimental results show that federate MCMC-based MRF PF not only tracks the targets in low errors, but has the virtue of fault detection, isolation and recovery capability. So it's feasible to solve the problem of behavior extraction of interacting pedestrians by way of vertical and horizontal cameras..

After successfully tracking the moving pedestrian from one frame to another in video sequences, the problem of understanding and describing people's behaviors, especially the group interaction behaviors, follows naturally. Behavior understanding involves the analysis and recognition of motion patterns. The high-level description of behaviors like actions and interactions involves the methods of DTW, FSM, HMM, TDNN, natural language description of behaviors, etc [11]. If this problem is breached, the simulation of interpersonal behavior in autonomous pedestrians will gain a great improvement.

References

[1] Lee, K. H., Choi, M. G., Hong, Q. and Lee, J.. Group Behavior from Video: A Data-Driven Approach to Crowd Simulation. In. Proc. ACM SIGGRAPH / Eurographics Symposium on Computer Animation,2007, p. 109 – 118.

[2] Z. Khan, T. Balch, and F. Dellaert. An MCMC-based particle filter for tracking multiple interacting targets. In Proc. ECCV, Prague, May 2004.

[3] Deriche, R., Faugeras, O.: Tracking line segments. Image and Vision Computing 8,1990, p. 261–270

[4] Cox, I., Leonard, J.: Modeling a dynamic environment using a Bayesian multiple hypothesis approach. Artificial Intelligence 66,1994, p. 311–344

[5] Rasmussen, C., Hager, G.: Probabilistic data association methods for tracking complex visual objects. PAMI 23,2001, p. 560–576

[6] Schulz, D., Burgard, W., Fox, D., Cremers., A.B.: Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. In: IEEE Intl. Conf. on Robotics and Automation (ICRA),2001.

[7] Smith, K., Gatica-Perez, D., Odobez, J.M.. Using Particles to Track Varying Numbers of Interacting People. Computer Society Conference on Computer Vision and Pattern Recognition, 2005

[8] J. Mac Cormick and A. Blake. A probabilistic exclusion principle for tracking multiple objects. In Proc. IEEE ICCV.

[9] HU Hai-dong, HUANG Xian-lin, LI Ming-ming, SONG Zhuo-yue. Federated unscented particle filtering algorithm for SINS/CNS/GPS system. J.Cent.South Univ.Technol,2010, p. 778-785

[10] CARLSON N A.Federated square root filter for decentralized parallel processes[J].IEEE Transactions on Aerospace and Electronic Systems,1990, p. 517-525.

[11] Wei-ming Hu, Tie-niu Tan. A Survey on Visual Surveillance of Object Motion and Behaviors. IEEE TRANSACTIONS ON SYSTEMS, 2004, p. 334-352