Interacting Multiple Model Particle Filter To Adaptive Visual Tracking

Jianyu Wang[1], DebinZhao[1,2], Wen Gao[1,2], Shiguang Shan
1Department of Computer Science, Harbin Institute of Technology, Harbin, China, 150001
2FRJDL, Institute of Computing Technology, CAS, P.O.Box 2704, Beijing, China, 100080
{jywang, dbzhao, wgao, sgshan}@jdl.ac.cn

ABSTRACT

Visual tracking could be formulated as a state estimation problem of target representation based on observations in image sequences. Approaching visual tracking problem in the Bayesian filter framework, how to sample the state evolution model to generate hypothesis of high confidence level is a critical factor. In this paper, we introduce an Interacting Multiple Model Estimation (IMME) framework for adaptive visual tracking. The essence of the IMME framework is that the state is estimated by integrating several different models in parallel and by interacting among those models' estimates probabilistically. Based on the IMME framework, we propose a new variation of particle filter named Interacting Multiple Model Particle Filter (IMMPF), in which the hypotheses can be sampled from several different state evolution models adaptively. Experiments show that, when compared with the standard particle filter, the IMMPF generates better hypotheses resulting in better tracking results, especially when the target behaves along several motion modes randomly.

1. INTRODUCTION

In general, most tracking systems or technologies consist of four parts: target representation, measurement localization, data association and filtering. Target representation and measurement localization are mostly bottom-up processes which has to cope with the appearance changes of the target, whereas data association and filtering are commonly top-down processes dealing with the dynamics of the tracked object [8].

Compared to the well-known parametric filters such as Kalman filter, the non-parametric based method, such as particle filter, becomes more popular because it does not assume functional form of the posterior. When the particles are property placed, weighted and propagated, posteriors can be estimated sequentially over time [3]. The state evolution model, which dominates the propagating of particles, plays a very important role in particle filter implementation. In the visual tracking community, following the pioneering work of CONDENSATION [1], various improvements have been proposed for visual tracking to deal with the state evolution problem [3,4,7,9,10].

Initializing to solve the irregular motion problem in visual tracking, this paper introduces an Interacting Multiple Model Estimation (IMME) framework to adaptively estimate the target's motion state from the radar tracking community. We then incorporate the IMME framework into the standard particle filter and propose a new variation of particle filter named Interaction Multiple Model Particle Filter (IMMPF). The IMMPF not only solves the irregular motion problem efficiently and accurately, but also provides a method to approach the fundamental problem of how to estimate dynamic system's state from sequential observations by integrating several different models. In IMMPF, the fittest state evolution models can be determined from several defined models to propagate the particles adaptively and the final result is obtained by probabilistically weighting among several models' results.

The rest paper is organized as follows: In Section 2, we formulate the visual tracking problem to have the structure of inference on a hidden Markov model. We introduce the Interacting Multiple Model Estimation Framework in Section 3. A new variation of particle filter, named as IMMPF is proposed in Section 4. We give the experiment results in Section 5 and conclude our work in Section 6.

2. VISUAL TRACKING PROBLEM DEFINITION

Let target's states and observations be represented by random variables X and Z, respectively. In the Bayesian approach to dynamic state estimation, one attempts to construct the posterior probability density function (pdf) of the state based on all available information, including the set of received observations. Approached the tracking problem by the temporal Bayesian filtering technique, we have a two step solution:

1) Prediction step:

$$P(x_i \mid z_0,...,z_{i-1}) = \int P(x_i \mid x_{i-1}) P(x_{i-1} \mid z_0,...,z_{i-1}) \tag{1}$$

2) Verification step:

$$P(x_i \mid z_0,...,z_i) = \frac{P(z_i \mid x_i) P(x_i \mid z_0,...,z_{i-1})}{\int P(z_i \mid x_i) P(x_i \mid z_0,...,z_{i-1}) dx_i} \tag{2}$$

The state evolution model $P(x_i \mid x_{i-1})$ and the likelihood model $P(z_i \mid x_i)$ are two key components in this recursive estimation process.

3. THE INTERACTING MULTI-MODEL ESTIMATION FRAMEWORK

For targets, it is not always reasonable to assume that the state evolution model $P(x_i \mid x_{i-1})$ will be consistent in distribution form throughout the tracking process.

Under real world conditions, one scope of problem is that the target's state variation can be modeled by the variation of parameters of some specific model. However, there exist a lot of practical situations that the target has several state evolution modes and its state variation can only be satisfying modeled by several different models. Thus, the IMME approach, which can adaptively estimate the target's state by integrating several state evolution models, is naturally the solution to such kind of problems. The approach is initially proposed by [2] in radar community and has received little attention by visual tracking community [11].

The basic idea of IMME framework is that it does state estimation in parallel by integrating several different models according to a Markov model and does interaction among models' estimates. Typically, one model can describe one state evolution mode. One step of IMME procedure can be described as followes: the previous state x(t) is evolved according to M models in parallel to generate M kinds of hypotheses. The M models have been defined in prior according to the characteristic of the application. Observation data z(i+1) is extracted to update each model's state estimate. After that, the assumed Markov transition properties between models is used and new filtered state estimates are computed for each model via mixing process. The estimated state is then transitioned according to the probability that the true target state makes a transition. Finally, the mode probabilities are updated according to the current observation and previous state. Thus, IMME approach can tell which model the target's dynamics obey. It's naturally to incorporate the IMME framework into the particle filter to get an adaptive sequential search algorithm to solve tracking problem robustly.

4. INTERACTING MULTIPLE MODEL PARTICLE FILTER

Problems like tracking that an estimate is required every time when a measurement is received, a recursive Bayesian filter is a convenient solution.

Particle Filter is a technique for implementing a recursive temporal Bayesian filter by Monte Carlo simulations. The key idea is to represent the required posterior pdf by a set of random samples with associated weights and to compute estimates based on these samples and weights.

There are mainly three key elements in particle filter implementation for sequential state estimation:

1). Sampling $P(x_{i+1} \mid x_i)$ to propagate particles to generate hypotheses.
2). Defining the likelihood function $P(z_i \mid x_i)$ to relate noisy observations to states.
3). Resampling to replace the particles with small weights.

Instead of propagating the particles by one state evolution model to generate only one kind hypothesis throughout the tracking process, we introduce the IMME framework into the particle filter to propose a new variation particle filter, which has the ability to make several kinds of hypotheses probabilistically.

To describe the approach mathematically, we define following terms:

P_{mn}^{i-} : The probability at time i that the target dynamic mode will translate from model m to model n, These probabilities are assumed to be known a priori here and satisfy: $\sum_{m=1}^{M} P_{mn}^{i-} = 1$, where M is the number of predefined models which can describing the M state evolution modes respectively. A transition matrix, which stacks the P_{mn}^{i-}, combines the M models according to a Markov model.

P_{mn}^{i+} : The conditional probability that the target made the transition from state m to state n at time i..

P_m^{i-} : The probability that the target translates its state according to model m during time interval $[i, i+1)$ and satisfy $\sum_{m=1}^{M} P_m^{i-} = 1$.

P_m^{i+} : The probability after interaction that the target will translate its state according model m and satisfy $\sum_{m=1}^{M} P_m^{i+} = 1$.

Then we have the following relations:

$$P_{mn}^{i+} = P_{mn}^{i-} P_m^{i-} / P_m^{i+} \tag{3}$$

$$P_m^{i+} = \sum_{m=1}^{M} P_{mn}^{i-} P_m^{i-} \tag{4}$$

Let $s_i = \{x_i^k, w_i^k, m_i^k \mid k = 0, ..., N\}$ denote a particle set at time i, where m_i^k means the mode according to which the particle k evolves in the state space at time i. For each particle, we define its evolution mode according to the mode probability P_m^{i-} and approximately there have the relation that the number of particles that will evolve according to mode m is proportional to P_m^{i-} . It mans that all particles are divided into M subsets probabilistically. Then, each subset of particles behaves like a standard particle filter and M filtered states are obtained according to *M* different evolutionary models.

Let x_i denote the estimated state at time i and $\hat{x}_{i+1}^{m}$ denote the estimated state at time i+1 by particles in subset m. First, a mixed state estimation is computed for each subset m:

$$\tilde{x}_{i+1}^{n} = \sum_{m=1}^{M} P_{mn}^{(i+1)+} \hat{x}_{i+1}^{m} \qquad (5)$$

Then the final estimated state x_{i+1} is computed as:

$$x_{i+1} = \sum_{n=1}^{M} P_{n}^{(i+1)+} \tilde{x}_{i+1}^{n} \qquad (6)$$

There will be a difference residual distance

$$(d_{i+1}^{m})^2 = (\frac{N_m}{\sum_{k=1}^{N_m} w^k})^2 \qquad (7)$$

with the subset m at the time i+1, which d_{i+1}^{m} is some distance defined between extracted observation and expected observation computed from $P(z_{i+1} \mid x_{i+1})$ .Then, assuming measurement dimension R and Gaussian statistics, the likelihood function for the observation given model m is

$$V_{i+1}^{m} = \frac{\exp(-(d_{i+1}^{m})^2 / 2)}{\sqrt{(2\pi)^R \sigma_{i+1}^{m}}} \qquad (8)$$

where σ_{i+1}^{m} is the corresponding covariance matrix of the distance d_{i+1}^{m} . Finally, using Bayes's rule, the updated model probability is $P_{m}^{(i+1)-} = V_{i+1}^{m} P_{m}^{i+} / C$, where $C = \sum_{m=1}^{M} V_{i+1}^{m} P_{m}^{i+}$ is the normalizing constant.

The state estimation is done in such a recursive process by propagating several kinds of hypotheses in parallel.

5. EXPERIMENTS

The tracking performance of the IMMPF is examined in this section. We record a 69-secondes sequence at resolution 320X240 pixels and 3 frames/second showing a toy car running on the ground. The car runs at the speed approximately 1m/s. At the first period, the car runs along a straight line. Then it turns around along a circle and finally, the car switches back to the motion mode along a straight line. That is to say the car has changed its motion mode two times in the whole process. The IMMPF is tested on this sequence in order to show its effectiveness on adaptive visual tracking by combining several models in parallel and its ability to decide the target's motion mode automatically. Figure 1 shows some tracking results on the test sequence.

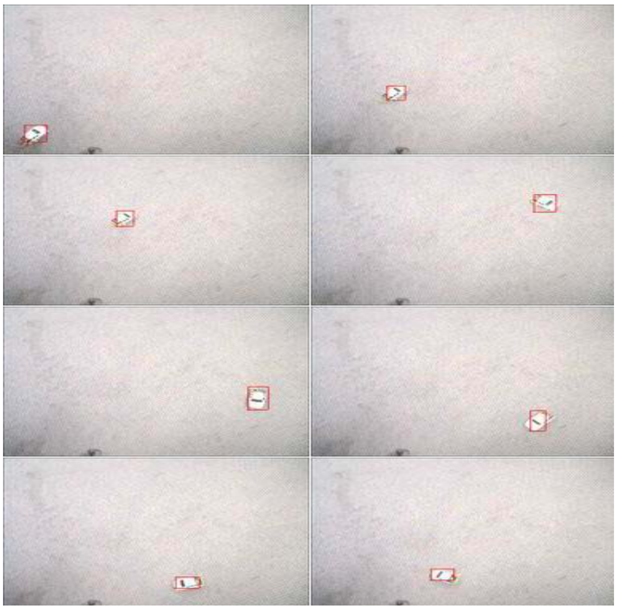

Figure 1. Some tracking results on the test sequence

5.1. Target State Representation and Motion Models

We simply concatenate positions and velocities of the car into a four dimension state vector:

$$x = (p_x, p_y, v_x, v_y)^T \qquad (9)$$

Two motion models are implemented in this experiment. One is the Nearly Constant Velocity Motion Model (NCVMM), and the other is the Nearly Constant Horizontal Turn Motion Model (NHTMM). The detailed implementation of these two models can be found in [11].

Additionally, the Markov transition matrix between two motion models is assumed constant in the experiment and is empirically set to:

$$M_T = \begin{Bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{Bmatrix}. \qquad (10)$$

In our experiments, the tracking performance is insensitive to the little changes of this predefined probability matrix M_T.

5.2. Measurement Localization and the Likelihood Function

In the experiment, a non-parametric object representation by 2D color distribution similar to [6] is employed. For each hypothesis generated by state evolution model in previous subsection, the corresponding color distribution is computed and compared with the target reference color distribution. The Bhattacharyya distance between them can be regarded as the output of the likelihood model $P(x_i \mid z_i)$.

5.3. Tracking Results

1000 particles are used to implement both the Standard Particle Filter (SPF) and the IMMPF described in Section 4. NCVMM is adopted as the state evolution model for the SPF as usual. Some tracking results are compared in Figure 2.

From Figure 2, we can see that when the car switches its motion mode at first time, the SPF loses the target while our proposed IMMPF works well.

Our system can work only at a rate about 3 frames/second on a Pentium *III*, 667MHz PC currently. The implementation of our proposed IMMPF is not optimized yet. High processing rate can be anticipated by further optimization.

Other conducted experiments, which prove the efficiency of the IMMPF, are not described here due to the short of page.

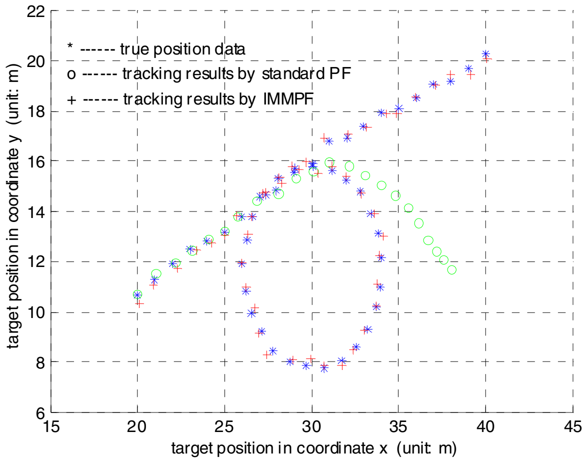

Figure 2. Comparison of tracking results of SPF, IMMPF and true position data

6. CONCLUSIONS

In this paper, we introduce an IMME framework which can do state estimation adaptively by combining several predefined models according to a Markov model. Based on the IMME framework, we propose a new variation of particle filter, named IMMPF, to solve the irregular motion problem in visual tracking. The IMMPF not only has the ability to integrate several models like the IMME method, but also inherits the power of particle filter to maintain a pool of hypotheses.

7. ACKNOWLEDGEMENT

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, National Hi-Tech Program of China (No. 2001AA114190 and 2002AA118010), and ISVISION Technologies Co., Ltd.

8. REFERENCES

[1] M. Isard and A. Blake, CONDENSATION -- conditional density propagation for visual tracking, In *International Journal of Computer Vision*, 29, 1, 5--28, 1998.

[2] H.A.P Blom and Y. Bar-Shalom, The interacting multiple model algorithm for systems with Markovian switching coefficients, *IEEE Transaction on Automatic Control*, vol. 33, no. 8, pp. 780-783, Aug. 1988.

[3] Y. Rui and Y. Chen, Better Proposal Distributions: Object Tracking Using Unscented Particle Filter, In *IEEE Proceedings of Computer Vision and Pattern Recognition*, pp. II-786 to 793, Kauai, Hawaii, December 11-13, 2001.

[4] Y. Wu and T. S. Huang, A co-inference approach to robust visual tracking, In *IEEE International Conference on Computer Vision*, Vol. II, pp.26-33, Vancouver, Canada, July, 2001.

[5] C. Rasmussen and G. Hager, Probabilistic Data Association Methods for Tracking Complex Visual Objects, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, June 2001.

[6] D. Comaniciu, V. Ramesh and P. Meer, Real-time tracking of non-rigid objects using Mean Shift, In *IEEE Proceedings of Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, Vol. 2, 142-149, 2000.

[7] Merwe, R., Doucet, A., Freitas, N., and Wan, E., The unscented particle filter, *Technical Report CUED/F-INFENG/TR 380*, Cambridge University Engineering Department, August 2000.

[8] D. Comaniciu, V. Ramesh and P. Meer, Kernel-Based Object Tracking, *IEEE Transaction on Pattern Analysis and. Machine Intelligence,* October 2002.

[9] J. MacCormick and M. Isard, Partitioned sampling, articulated objects, and interface-quality hand tracker, In *Europe Conference on Computer Vision*, Vol.2, pp.3-19, 2000.

[10] M. Isard and A. Blake, A mixed-state Condensation tracker with automatic model-switching, In *IEEE International Conference on Computer Vision*, 107-112, 1998.

[11] Bar-Shalom, Y., and X. R. Li, Estimation and tracking: principles, techniques and software, published by Artech House,1993.