MAP PARTICLE SELECTION IN SHAPE-BASED OBJECT TRACKING

A. Dore and C. S. Regazzoni

University of Genova
Department of Biophysical and Electronic Engineering
16145 - Via Opera Pia 11A, Genova - Italy

M. Musso

TechnoAware S.r.l.
Via Greto di Cornigliano,
Genova - Italy

ABSTRACT

The Bayesian filtering for recursive state estimation and the shape-based matching methods are two of the most commonly used approaches for target tracking. The Multiple Hypothesis Shape-based Tracking (MHST) algorithm, proposed by the authors in a previous work, combines these two techniques using the Particle Filter algorithm. The state of the object is represented by a vector of the target corners (i.e. points in the image with high curvature) and the multiple state configurations (particles) are propagated in time with a weight associated to their probability. In this paper we demonstrate that, in the MHST, the likelihood probability used to update the weights is equivalent to the voting mechanism for Generalized Hough Transform (GHT)-based tracking. This statement gives an evident explanation about the suitability of a MAP (Maximum a Posteriori) estimate from the posterior probability obtained using MHST. The validity of the assertion is verified on real sequences showing the differences between the MAP and the MMSE estimate.

Index Terms**—** Particle Filter, Shape Tracking, MAP estimate

1. INTRODUCTION

Tracking objects is one of the most challenging and not yet satisfactorily solved problem within the discipline of Computer Vision. Major problems occur in presence of non-rigid, fast moving objects and in complex and crowded scenes, where frequent occlusions cause partial or total lack of target's observations.

The representation of the shape of an object by means of different types of features has been demonstrated to be a powerful approach to accomplish this task. The basic idea is to match a target model, if necessary continuously updated, with the observations of the objects in the scene. Hariharakrishnan and Schonfeld [1] proposed a shape-based target tracking technique that copes with frequent occlusions and non-rigid objects by handling contour with a region-based approach. Object boundary is predicted computing the block motion estimation and performing an occlusion/disocclusion detection for the sub-parts of the target. This procedure renders more reliable the tracker relying on regions association. Gabriel *et al.* [2] make use of Interest Points (i.e. points in the image where significant changes occur, as corners, junctions, etc.) to characterize the shape of an object. To track a target, points are extracted with Harris detector and, frame by frame, they are matched by means of the Mahalanobis distance.

Tracking can be also considered as a state estimation given the available observations. According to this consideration the Bayesian filtering represents an efficient solution to accomplish tracking. Sequential Monte Carlo techniques (e.g. *Particle Filter*) are widely employed methods to approximate the Bayesian filtering in non-linear, non-Gaussian situations. In [3, 4] tracking is accomplished by means of Particle Filter, describing the shape of the target with contours. More in details, the CONDENSATION [3] algorithm, proposed by Isard and Blake, tracks objects, modelling their shapes with B-splines. Lanz, in [5], faces the problem of tracking defining a Bayesian model of multiple parts of the body and a related particle filtering based method to handle partial and complete occlusions.

This paper aims at providing a motive for MAP (Maximum A Posteriori) state estimate for a particle filter based tracking algorithm, the *Multiple Hypothesis Shape-based Tracking* (MHST) [6], by demonstrating that this choice corresponds to a shape matching technique.

The remainder of the paper is organized as follows: in Sect. 2 an overview of the Particle Filter is provided. In Section 3 the *Multiple Hypothesis Shape-based Tracking* (MHST) algorithm is described and the suitability of the *Maximum a Posteriori* (MAP) estimate is stated. In Sect. 4 results on real-world sequence demonstrate the advantages of MAP estimate with respect to MMSE (Minimum Mean Square Error) and, finally in Sect. 5 we conclude.

2. THE PARTICLE FILTER ALGORITHM

The recursive Bayesian state estimation (Bayesian filtering) is one of the mathematical tools most commonly employed in tracking to evaluate step-by-step the target state, usually defined by its kinematics features. In Bayesian filtering two steps can be identified: the prediction and the update. The

system transition model
$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ and the set of available observations $\boldsymbol{z}_{1:k-1} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{k-1}\}$ provides the posterior prediction as:

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1}|\boldsymbol{z}_{1:k-1}) d\boldsymbol{x}_{k-1} \quad (1)$$

New observations $\boldsymbol{z}_k$ at time k and the observation model supply the likelihood probability $p(\boldsymbol{z}_k|\boldsymbol{x}_k)$, that is used to correct the prediction by means of the update process:

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) = \frac{p(\boldsymbol{z}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1})}{p(\boldsymbol{z}_k|\boldsymbol{z}_{1:k-1})} \quad (2)$$

The Particle Filter [7] approximates Bayesian filtering when a closed-form solution cannot be computed (i.e. when transition and observation models are non-linear and non-Gaussian) by representing the posterior as a finite set of weighted samples $\boldsymbol{\chi}_k = \{\boldsymbol{x}_k^{(m)}, w_k^{(m)}\}_{m=1}^{N_s}$. The set of N_s candidate samples (i.e. *particles*) $\{\tilde{\boldsymbol{x}}_k^{(m)}\}_{m=1}^{N_s}$ representing the prediction are drawn from the so called proposal distribution (or importance distribution) $q_k = (\boldsymbol{x}_k|\boldsymbol{x}_{1:k-1}, \boldsymbol{z}_{1:k})$. In many applications the proposal distribution can be reasonably obtained by the transition model so that particles are drawn from $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(m)})$. Values of the associated weights are obtained by means of the equation:

$$w_k^{(m)} = \frac{p(\boldsymbol{z}_k|\boldsymbol{x}_k^{(m)})p(\boldsymbol{x}_k^{(m)}|\boldsymbol{x}_{k-1}^{(m)})}{q(\boldsymbol{x}_k^{(m)}|\boldsymbol{x}_{0:k-1}^{(m)}, \boldsymbol{z}_{0:k})} w_{k-1}^{(m)} \quad (3)$$

When the proposal distribution is given by the transition model, the weight computation is simplified so that it can be derived by $w_k^{(m)} = p(\boldsymbol{z}_k|\boldsymbol{x}_k^{(m)})w_{k-1}^{(m)}$. However, this choice, whereas computationally efficient, provokes a degeneration of performances, under the form of the propagation of several particles with low weight, not representative of the state. To overcome this issue the Sequential Importance Resampling (SIR) algorithm procedure redistributes (*resample*) the particles to have a more accurate approximation of the posterior.

From the approximated posterior probability two possible state estimate can be easily obtained: the *MMSE* is deduced as the mean of the weighted particles; instead, the *MAP* estimate coincides with the sample associated to the highest weight.

3. MULTIPLE HYPOTHESIS SHAPE-BASED TRACKING (MHST)

3.1. The MHST Algorithm

In [6], the authors describe a method, based on the Particle Filter algorithm, that tracks objects basing conjunctively on shape and position. To accomplish that, they proposed an update procedure inspired by the Generalized Hough Transform (GHT). According to the GHT the shape of an object can be described with a list of corners (i.e. points in an image characterized by high curvature) positions.

The state of the object is represented by a vector, $\boldsymbol{x} = [(x_{(1)}, y_{(1)}), (x_{(2)}, y_{(2)}), \ldots, (x_{(N)}, y_{(N)})]^T$, constituted by the image plane position of N corners. Then for each object a $2 \times N$ vector is initialized with a set of corners equally distributed in the bounding box. This procedure is necessary to increase the quantity of information carried by each of the two dimensional subspaces of the state vector (i.e. the corners position). Prediction step is fulfilled using a second-order autoregressive model:

$$\boldsymbol{x}_k = A\boldsymbol{x}_{k-1} + B\boldsymbol{u}_{k-1} + \boldsymbol{\nu}_{k-1} \quad (4)$$

where A and B are identity matrices of dimension $2N \times 2N$, $\boldsymbol{\nu}$ is a Gaussian noise $N(0, \sigma^2)$, and $\boldsymbol{u}_{k-1}$ is the input vector defined as $\boldsymbol{u}_{k-1} = \hat{\boldsymbol{x}}_{k-1} - \hat{\boldsymbol{x}}_{k-2}$, with $\hat{\boldsymbol{x}}$ indicating the estimate. The prediction process provides multiple hypothesis regarding the position and the shape of the object in the following instant according to the transition model specified in Eq. 4. In fact the motion model describes the movement of each corner defined in the state vector.
In the update step, particle weights are recomputed according to the likelihood probability. To accomplish that the observed corners are compared to the predicted state by matching between the hypothesized configurations of the state vector and the detected corners. A function $s(\boldsymbol{x}_{k(i)}^{(m)})$ can be defined to determine if a corner of the model is close to an observed corner $\boldsymbol{z}_{k(j)}$, that is:

$$s(\boldsymbol{x}_{k(i)}^{(m)}) = \sum_{j=1}^{M} \exp\left(-\left(d_{k(i,j)}^{(m)}\right)^2\right) \quad (5)$$

where M is the total number of extracted corners and $\left(d_{k(i,j)}^{(m)}\right)^2 = \left\|\boldsymbol{z}_{k(j)} - \boldsymbol{x}_{k(i)}^{(m)}\right\|^2$ is the Euclidean distance between the i-th predicted corner of the m-th particle $\boldsymbol{x}_{k(i)}^{(m)} = (x_i, y_i)$ and the j-th extracted corner.
Then, in order to provide a one-by-one association between predicted and observed corners, $s(\boldsymbol{x}_{k(i)}^{(m)})$ is compared to a unitary distribution, i.e.

$$V_k = \sum_{i=1}^{N} \left(s(\boldsymbol{x}_{k(i)}^{(m)}) - 1\right)^2 \quad (6)$$

The likelihood probability between the detected corners and the m-th particle representing a possible position and configuration of the object is:

$$p(\boldsymbol{z}_k|\boldsymbol{x}_k^{(m)}) \propto exp(-V_k) \quad (7)$$

The set of weighted samples $\boldsymbol{\chi}_k = \{\boldsymbol{x}_k^{(m)}, w_k^{(m)}\}_{m=1}^{N_s}$ with $w_k^{(m)} = w_{k-1}^{(m)} p(\boldsymbol{z}_k|\boldsymbol{x}_k^{(m)})$ then gives an approximation of the posterior distribution. A resampling step follows to remove the not probable particles.

3.2. Analogies between GHT-Based Tracking and MHST

Oberti *et al.* in [8] use a variation of the GHT to track objects by comparing the observed corners with the model of the object (i.e. the GHT). The GHT is composed by a list of corners characterized by the their relative position (δx, δy) with respect to a reference point $\boldsymbol{x}_{REF}$ (e.g. the centroid, the upper left corner of a bounding box, etc.) by its gradient and by the persistence of the corner over successive frames. When a corner belonging to the model matches (with respect to gradient and relative position) a detected one, a vote is given to the reference point. The reference point that obtains the highest number of votes identifies the displacement of the object.

The approach described in Sect. 3 is also based on a shape matching between a model (a particle) and the observed corners by means of Eq. 6. In this case corners are localized and compared in the image plane coordinates without taking into account a reference point, as it was in [8]. However, if we consider the model corner position for the particle m as $\boldsymbol{x}_{k(i)}^{(m)} = \delta\boldsymbol{x}_{k(i)}^{(m)} + \boldsymbol{x}_{REF_k}^{(m)}$ it is possible to demonstrate that Eq. 6 corresponds to a voting mechanism on the reference point (e.g. the centroid with respect to the corners of $\boldsymbol{x}_k$).

The prediction step applied to a particle, which represents a target position and configuration, establishes a shift of corners based on the transition model. Coherently the reference point is subjected to a shift which, on the average, corresponds to the corners displacements. Then, the relative position of the corner $\delta\boldsymbol{x}_{k(i)}^{(m)}$ with respect to $\boldsymbol{x}_{REF_k}^{(m)}$ is also predicted.

In the update step the function $s(\boldsymbol{x}_{k(i)}^{(m)})$ (see Eq. 5) is computed. When this function is close to the unity it means that the predicted corner match an observed one. If we write the j-th extracted corner and the i-th corner of the model positions with respect to the reference point we have:

$$\tilde{s}(\boldsymbol{x}_{k(i)}^{(m)}) = \sum_{j=1}^{M} \exp\left(-\left(\tilde{d}_{k(i,j)}^{(m)}\right)^2\right) \tag{8}$$

where $\left(\tilde{d}_{k(i,j)}^{(m)}\right)^2 = \left\|(\boldsymbol{z}_{k(j)} - \boldsymbol{x}_{REF_k}^{(m)}) - \delta\boldsymbol{x}_{k(i)}^{(m)}\right\|^2$. Then, as this distance is close to one for a corner i that matches an observation $\boldsymbol{z}_{k(j)}$, the value of $V_k(i)$ is near zero entailing an high likelihood for corner i. According to this procedure a predicted reference point $\boldsymbol{x}_{REF_k}^{(m)}$ is supported by observations when for the majority of the the i-th corners of the model there is one (or more) j detected corner such that $\left(\tilde{d}_{k(i,j)}^{(m)}\right)^2 \sim 0$. Then, as in [8] the quantity $\left(\tilde{d}_{k(i,j)}^{(m)}\right)^2$ is used to find the $\boldsymbol{x}_{REF_k}^{(m)}$ displacement, in this case it is employed to confirm the predicted shift.

3.3. Estimate Selection

Given the particles approximating the posterior distribution, an estimate of the state has to be defined at each step. As previously stated (see Sect. 2) MMSE and MAP estimates can be derived from $\boldsymbol{\chi}_k = \{\boldsymbol{x}_k^{(m)}, w_k^{(m)}\}_{m=1}^{N_s}$. The most appropriate estimate for the MHST is the MAP, that is the particle with the highest weight. This statement comes from the fact that, as we have just demonstrated, (see Sect. 3.2) the likelihood of the (m)-th particle is equivalent to the votes received by the related reference point. Then choosing the particle with the highest weight coincides to select the $\boldsymbol{x}_{REF_k}^{(m)}$ receiving the highest number of votes from the observations and then the most probable displacement of the object. The MMSE estimate, on the contrary, is not suitable since it would provide a weighted mean of the predicted shift which, therefore, is not representative of the object motion given by the model-observation matching.

4. EXPERIMENTAL RESULTS

We demonstrate the suitability of the choice of the MAP estimate for the MHST algorithm in real-world test sequences of rigid and non-rigid object. As outlined in [6] this tracking method can work for static cameras without the necessity of a change detection algorithm to localize moving targets. To verify this fact, observed corners are extracted using a SUSAN corner detector [9] in a searching area surrounding the target.

Targets are manually selected at the first frame and the prior distribution is initialized with a Gaussian having a mean equivalent to the corners position. To choose the N corners constituting the mean an automatic method aiming at having them equally distributed on the object is employed. In our experiments the state vector is 16 dimension, i.e. 8 corners are used to represent the target. This choice is a trade-off between the necessity of an accurate model and the computational complexity. Moreover the increase of the state dimension leads to sample degeneracy and impoverishment. Since no change detection module has been employed the observed corners $\boldsymbol{z}_k$ used to compute the likelihood are extracted from a searching area of variable dimension according to the kinematics characteristic of the target.

Two situations that can be commonly encountered in the tracking domain of application were chosen to show the better results of the MAP estimate with respect to MMSE: 1) occlusion of two non rigid targets (humans); 2) a rigid object modifying its shape due to the motion, in this case, a curving car. The posterior is approximated with 200 particles and the transition model variance, that takes into account the accelerations and deformations of the corners, is $\sigma^2 = 0.5$ for 1) and is $\sigma^2 = 1$ in 2), the curving vehicle test sequence. In these conditions, with a non optimized code, on a Pentium 4 3.0 GHz machine with 1 GB RAM, the tracker works at 9.1 frames/sec with one target and 4.6 in the test 1), where two targets are present. In Figures 1(a)-1(c) it is shown that, in the case of two persons walking and coming to an occlusion, the MHST is able to maintain the track of the two targets after the overlap when the MAP estimate is used. When MMSE estimate is employed (see Figures 1(d)-1(f)), on the contrary, the tracking procedure fails. In this case, in fact, when the occlusion occurs the MMSE is not capable to cope with the non-linearity of the situation and this leads the Particle Filter to propagate wrong states hypothesis. Using the MAP estimate, on the other hand, only the particle, which best matches the observations, is preserved allowing to recover the object after the occlusion. The second test (see Fig. 2), where a curving car is tracked, shows again the failures of the MHST with MMSE estimates for non-linear shifts. In particular, worth of noticing is that the MMSE is not able to handle the changes in car dimension and pose and the track tends to propagate, as expected, a linear behaviour. Table 1 shows the Root Mean Square Error (RMSE) of the MHST with the two estimate computed with respect to the Ground Truth showing the significant better results if MAP estimate is used.

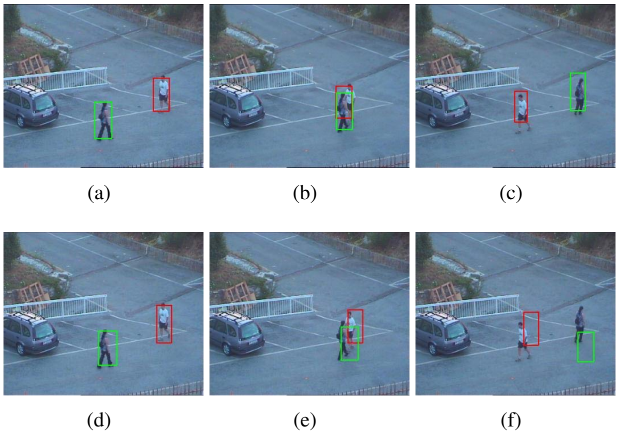

Fig. 1. a), b), and c) MHST results with MAP estimate; d), e), and f) MHST results with MMSE estimate

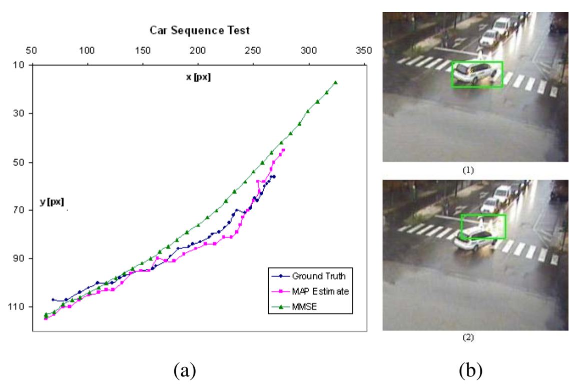

Fig. 2. MHST results for a curving car: a) the real and estimated trajectories in the image plane; b1) MAP estimate; b2) MMSE estimate

Table 1. RMSE of the MHST using MAP and MMSE estimates

	Sequence 1 (occlusion) 60 frames		Sequence 2 (car)
	Target 1	Target 2	30 frames
MAP	6.1	9.1	13.1
MMSE	9.1	20	27.2

5. CONCLUSIONS AND FUTURE WORKS

In this paper we presented a demonstration of the equivalence between the likelihood probability used to update the weights in a Particle Filter based tracking algorithm, the Multiple Hypothesis Shape-Based Tracking (MHST) algorithm, and the voting mechanism for Generalized Hough Transform (GHT)- based tracking. This consideration provides a motive regarding the suitability of a MAP state estimate instead of the MMSE estimate and the significant benefits of this choice have been shown over two real-world sequences.

Ongoing researches are focused on selecting the most appropriate corners of the object to initialize the state vector and to define a transition model more robust to sudden direction and speed changes of the target.

6. REFERENCES

[1] K. Hariharakrishnan and D. Schonfeld, "Fast object tracking using adaptive block matching," *IEEE Transactions on Multimedia*, vol. 7, no. 5, pp. 853–859, 2005.

[2] P. Gabriel, J. B. Hayet, J. Piater, and J. Verly, "Object tracking using color interest points," in *IEEE Proc. of Int. Conf. on Advanced Video and Signal based Surveillance, AVSS '05*, Como, Italy, September 2005, pp. 159–164.

[3] M. Isard and A. Blake, "CONDENSATION – conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[4] P. Li, T. Zhang, and A. E. C. Pece, "Visual contour tracking based on Particle Filters," *Image and Vision Computing*, vol. 21, no. 1, pp. 111–123, 2003.

[5] O. Lanz, "Approximate Bayesian multibody tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1436–1449, 2006.

[6] A. Dore, M. Asadi, and C. S. Regazzoni, "Multiple hypothesis shape tracking using Particle filtering and Hough-based observation models," in *SPIE Electronic Imaging - Visual Communications and Image Processing, VCIP 2007*, San Jose, CA, USA, January 2007.

[7] B. Ristic, S. Arulapalam, and N. Gordon, *Beyond the Kalman Filter*, Artech House Publishers, 2004.

[8] F. Oberti, S. Calcagno, M. Zara, and C. S. Regazzoni, "Robust tracking of humans and vehicles in cluttered scenes with occlusions," in *IEEE International Conference on Image Processing, ICIP2002*, Rochester, NY, USA, 2002, pp. 629 – 632.

[9] S. M. Smith and J. M. Brady, "SUSAN – a new approach to low level image processing," *Int. J. Comput. Vision*, vol. 23, no. 1, pp. 45–78, 1997.