Particle Filtering with Factorized Likelihoods for Tracking Facial Features

I. Patras and M. Pantic
Knowledge and Data Engineering Group
Delft University of Technology
Mekelweg 4, 2628 CD Delft
The Netherlands

Abstract

In the recent years particle filtering has been the dominant paradigm for tracking facial and body features, recognizing temporal events and reasoning in uncertainty. A major problem associated with it is that its performance deteriorates drastically when the dimensionality of the state space is high. In this paper, we address this problem when the state space can be partitioned in groups of random variables whose likelihood can be independently evaluated. We introduce a novel proposal density which is the product of the marginal posteriors of the groups of random variables. The proposed method requires only that the interdependencies between the groups of random variables (i.e. the priors) can be evaluated and not that a sample can be drawn from them. We adapt our scheme to the problem of multiple template-based tracking of facial features. We propose a color-based observation model that is invariant to changes in illumination intensity. We experimentally show that our algorithm clearly outperforms multiple independent template tracking schemes and auxiliary particle filtering that utilizes priors.

1. Introduction

In the recent years, particle filtering has been the dominant paradigm [2] [3] [8] [5] [4] [7] [11] in the tracking of the state α of a temporal event given a set of noisy observations $Y = \{\ldots, y^-, y\}$ up to the current time instant. Its ability to maintain simultaneously multiple solutions, the so called particles, make it particularly attractive when the noise in the observations is not Gaussian and robust to missing or inaccurate data. However, a problem that has been reported in this framework [1] [9] is that the performance deteriorates drastically as the dimensionality of the state space A (i.e. $\alpha \in A$) increases. Indeed, as the dimensionality of the state space increases, a large number of particles that are propagated from the previous time instance are wasted in areas where the likelihood of the observations is very low. Therefore, a very large number of particles are necessary to accurately track the state.

In this paper we propose a method that deals with the above mentioned problem in the case that the state α can be partitioned in groups of random variables α_i (i.e. $\alpha = \{\alpha_i\}$), such that the likelihood $p(y|\alpha_i)$ of the observations $y$at the current time instant, given each group α_i , can be independently evaluated. We build on the particle filtering framework, which involves the following three steps: a) sample from $p(a^-|Y^-)$, where a^- is the state at the previous time instant, b) propagate the samples via the transition probability $p(a|a^-)$ and c) evaluate a new weight for the samples from the likelihood $p(y|a)$. We propose a modified scheme which can be summarized as follows. First, each partition α_i is propagated and evaluated independently. This creates a particle-based representation of $p(\alpha_i|Y)$. We subsequently use this representations to sample from a proposal function $g(a|Y) = \prod_i p(\alpha_i|Y)$. Finally, each of the particles produced in this way is reweighted by evaluating the transition probability $p(\alpha|\alpha^-)$ so that the set of particles with their new weights represents the *a posteriori* probability $p(\alpha|Y)$. In correspondence to the standard particle filtering, our approach requires only that the transition probability $p(\alpha|\alpha^-)$ can be evaluated and not that it can be sampled from. Thus, it allows easier modeling of the interdependencies between the groups of random variables α_i (for example with a Markov Random Field). Furthermore, since the particles are sampled from the proposal function $g(.)$, it is guaranteed the likelihood $p(y|a)$ is not low and, therefore, that the particles are not wasted at areas of the state space with low likelihood.

We experimentally verify our claims by applying the proposed method to the problem of multiple template-based tracking of facial features. We propose a color-based observation model that is invariant to changes in illumination intensity and utilize learned priors of the relative configurations of the facial features. We provide comparative experimental results with other particle filters on real image sequences.

The remainder of the paper is organized as follows. In Section 2 we concisely review similar works and describe the proposed particle filtering method in detail. In Section

Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)

3 we explain how it is applied to the problem of template-based tracking of multiple facial features. In Section 4 we present comparative experimental results on real data and finally in Section 5 we draw conclusions and future research directions.

2. Particle filtering with factorized likelihoods

The main idea of the particle filtering is to maintain a particle based representation of the *a posteriori* probability $p(\alpha|Y)$ of the state α given all the observations Y up to the current time instance. This means that the distribution $p(\alpha|Y)$ is represented by a set of pairs $\{(s_k, \pi_k)\}$ such that if s_k is chosen with probability equal to π_k, then it is as if s_k was drawn from $p(\alpha|Y)$. In the particle filtering framework our knowledge about the *a posteriori* probability is updated in a recursive way. Suppose that we have a particle based representation of the density $p(\alpha^-|Y^-)$, that is we have a collection of K particles and their corresponding weights (i.e. $\{(s_k^-, \pi_k^-)\}$). Then, the Sampling/Importance Resampling particle filtering can be summarized as follows:

1. Pick K particles s_k^- from the collection $\{(s_k^-, \pi_k^-)\}$. Each particle s_k^- is picked with probability equal to π_k^-. This is approximately equivalent to sampling from $p(\alpha^-|Y^-)$.

2. Propagate each of the chosen particles via the "transition" probability $p(\alpha|\alpha^-)$. This creates a collection of particles $\{s_k\}$ which are, approximately, sampled from the so-called, proposal density $p(\alpha|Y^-)$.

3. To each particle s_k assign a weight π_k equal to the likelihood of the observations, that is let $\pi_k = p(y|s_k)$. Normalize the weights so that they sum up to one.

This results in a collection of K particles and their corresponding weights (i.e. $\{(s_k, \pi_k)\}$) which is an approximation of the density $p(\alpha|Y)$.

There are three problems with the above scheme. The first is that a large number of the particles that result from sampling from the proposal density $p(\alpha|Y^-)$ (i.e. step 2) are wasted in areas with small likelihood. In other words, when they "go through" the "measurement" density $p(y|\alpha)$, at step 3, they are assigned very low weights. The reason is that the transition probability $p(\alpha|\alpha^-)$, naturally, cannot deterministically produce particles with high likelihood and therefore adds noise in order to "explore" the state space. This is very inefficient when the dimensionality of the state space is high. The second problem is that the above scheme ignores the fact that while a particle $s_k =< s_{k1} s_{k2} \dots s_{kN} >$ might have low likelihood, it can easily happen that parts of it (e.g. s_{k1}) might be close to the correct solution. In many practical problems, it is easy to evaluate the goodness of these subparticles since the likelihood can be factorized (i.e. $p(y|\alpha) = \prod_i p(y|\alpha_i)$). Finally, the third problem is that it might be difficult to perform the second step, that is to sample from $p(\alpha|\alpha^-)$. When there are interdependencies between the different parts of the state α (e.g. in case that α is a Markov Random Field) it is generally easier to evaluate $p(\alpha|\alpha^-)$ (up to a constant factor) than to sample from it.

There has been a number of adaptations that attempt to improve the performance of the particle filtering by giving answers to some of these problems. Icondensation by Isard and Blake [3] proposes to substitute the first two steps by sampling from an empirical "external" proposal density $g(a)$ (that is, instead of sampling from $p(\alpha|Y^-)$). While their work is very important, it addresses only the first problem and leaves the choice of $g(\alpha)$ an open issue. Auxiliary particle filtering, introduced by Pitt and Shephard [8], also proposes the use of a modified proposal density. The main idea is to sample at stage one particles s_k^- that, when propagated via the transition probability $p(\alpha|\alpha^-)$, produce particles s_k with high likelihood. Their work has the advantage that it does not require an "external" proposal functions $g(.)$ but is limited to the first problem. The work of Deutscher *et al* [1] attempt to overcome the problems associated with the high dimensionality of the state space with automatically partitioning it according to the observed variance at each of its dimensions. Finally, a number of recent works [4] [11], attempt to deal with both of the first two problems.

Our method attempts to answer all of the three problems in the case that the likelihood can be factorized, that is in the case that $p(y|\alpha) = \prod_i p(y|\alpha_i)$ (see fig. 1). Similarly to [8] and [3] we propose the use of an adapted proposal function. In contrast to them, we utilize the fact that the likelihood $p(y|\alpha)$ can be factorized and propose to use as the proposal distribution $g(\alpha)$ the product of the *posteriors* of each α_i given the observations, that is $g(\alpha) = \prod_i p(\alpha_i|y)$. Our work is closer to [4] and [11] but we do not introduce an artificial hierarchy to the propagation of each partition α_i as in [4] and do not restrict to Markov Random Fields models for $p(\alpha)$ as in [11]. Furthermore we do not require iterative sampling from a complicated $p(a)$. For the remainder of this section we will show how we draw samples s_k from the proposal distribution and, subsequently, how we assign them weights π_k, such that the collection $\{(s_k, \pi_k)\}$ is a representation of the *a posteriori* probability $p(\alpha|Y)$.

2.1 Updating the posterior

For the proposal distribution $g(a)$ we ignore the interdependencies between the different α_i. This fact, which is witnessed by the factorized form of $g(\alpha)$, allows us to construct a sample $s_k =< s_{k1} \dots s_{ki} \dots s_{kN} >$ by independently sampling from $p(\alpha_i|Y)$. In, a slightly modified, particle fil-

Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)
0-7695-2122-3/04 $ 20.00 © 2004 IEEE

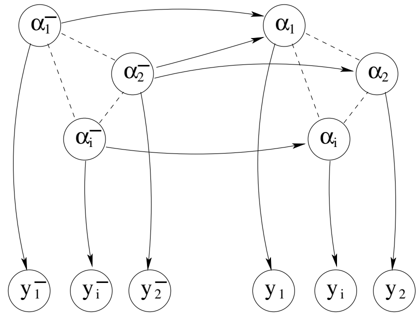

Figure 1: The assumed model. The broken lines represent interdependencies, that might or might not exist, between the partitions of the state variables.

tering framework we express the $p(\alpha_i|Y)$ as follows:

$$p(\alpha_i|Y) = p(y|\alpha_i) \sum_{\alpha^-} p(\alpha_i|\alpha^-) p(\alpha^-|Y^-) \qquad (1)$$

We use a modified version of the auxiliary particle filtering algorithm to end up with a particle based representation of $p(\alpha_i|Y)$, for each i. This procedure, which is repeated for each partition i, can be summarized as follows:

1. Propagate all particles s_k^- via the transition probability $p(\alpha_i|\alpha^-)$ in order to arrive at a collection of K subparticles μ_{ik}. Note, that while s_k^- has the dimensionality of the state space, the μ_{ik} has the dimensionality of the partition i.

2. Evaluate the likelihood associated with each subparticle μ_{ik}, that is let $\lambda_{ik} = p(y|\mu_{ik})$.

3. Draw K particles s_k^- from the probability density that is represented by the collection $\{(s_k^-, \lambda_{ik}\pi_k^-)\}$. This is the essence of the auxiliary particle filtering; in this way it favors particles with high λ_{ik}, that is particles which, when propagated with the the transition density, end up at areas with high likelihood.

4. Propagate each particle s_k^- with the transition probability $p(\alpha_i|\alpha^-)$ in order to arrive at a collection of K sub-particles s_{ik}. Note, that s_{ik} has the dimensionality of the partition i.

5. Assign a weight π_{ik} to each subparticle as follows, $w_{ik} = \frac{p(y|s_{ik})}{\lambda_{ik}}$, $\pi_{ik} = \frac{w_{ik}}{\sum_j wij}$

With this procedure, we have a particle-based representation for each of the N *posteriors* $p(\alpha_i|Y)$. That is, we have N collections (s_{ik}, π_{ik}), one for each i. Then, sampling K particles from our proposal function $g(\alpha) = \prod_i p(\alpha_i|Y)$ is, approximately, equivalent to constructing each particle $s_k =< s_{k1} \dots s_{ki} \dots s_{kN} >$ by sampling independently each s_{ik} from $p(\alpha_i|Y)$.

At the end of this phase we have a collection of particles $s_k =< s_{k1} \dots s_{ki} \dots s_{kN} >$ each of which is drawn from $g(\alpha)$. In order that this collection represents the *posterior* $p(\alpha|Y)$ we need to assign a weight to each particle equal to [3][8]:

$$\pi_k = \frac{p(y|s_k)p(s_k|Y^-)}{g(s_k)} \qquad (2)$$

$$= \frac{p(y|s_k)p(s_k|Y^-)}{\prod_i p(s_{ik}|Y)} \qquad (3)$$

$$= \frac{p(y|s_k)p(s_k|Y^-)}{\prod_i p(y|s_{ik})p(s_{ik}|Y^-)} \qquad (4)$$

$$= \frac{p(s_k|Y^-)}{\prod_i p(s_{ik}|Y^-)} \qquad (5)$$

In words, the numerator of the final weight of a particle is the probability of the particle s_k given the observations at the previous time instant. That is, at the numerator the interdependencies between the different s_{ik} are taken into consideration. On the contrary, at the denominator, the different s_{ik} are considered independent. In other words, the reweighting process favors particles for which the joint is higher than the product of the marginals. Note, that in the degenerate case that the α_i are indeed independent and, therefore, the weights are one, our scheme is equivalent to N independent particle filters, one for each of the α_i. Indeed, each s_{ik} is drawn independently from $p(\alpha_i|Y)$.

In the general case that eq. 5 cannot be evaluated by an appropriate model, the weights need to be estimated. We do so by utilizing the particle based representation of $p(\alpha^-|Y^-)$ as follows:

$$\pi_k = \frac{p(s_k|Y^-)}{\prod_i p(s_{ik}|Y^-)} \qquad (6)$$

$$= \frac{\sum_l p(s_k|s_l^-)p(s_l^-|Y^-)}{\prod_i \sum_l p(s_{ik}|s_l^-)p(s_l^-|Y^-)} \qquad (7)$$

Finally, the weights are normalized to sum up to one. With this, we end up to a collection $\{(s_k, \pi_k)\}$ that is a particle-based representation of $p(\alpha|Y)$.

3. Tracking multiple facial micro-features

In what follows we adapt the particle filtering scheme that has been proposed in Section 2 to the problem of multiple template-based tracking. We then apply it to the problem of simultaneously tracking multiple facial micro-features such as the lip corners, the middle of the mouth and a number of points on the eyebrows. We propose a novel observation

Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)
0-7695-2122-3/04 $ 20.00 © 2004 IEEE

likelihood for each template, that is based on robust color consistency. The proposed likelihood attempts to deal with changes in the appearance of the templates due to changing illumination conditions, facial expressions and occlusions. Furthermore, we propose tracking the facial micro-features by utilizing prior knowledge on the configurations of the facial micro-features, that is knowledge of $p(\alpha)$. We show how prior knowledge can be easily incorporated in our factorized particle filtering scheme.

In what follows we will first formally pose the problem, second describe the observation model, third the transition probability models and the use of priors in the reweighting phase. We will subsequently present comparative experimental results and close with some conclusions and directions for future research.

3.1 Problem formulation

Formally, we aim at tracking in an image sequence the 2-D positions α_i of N facial micro-features. At the first frame of the sequence the position of the facial micro-features are initialized by the user (e.g. fig. 2). In the notation of Section 2, $\alpha =< \alpha_1 \ldots \alpha_i \ldots \alpha_N >$ is the $2N$ dimensional random variable that represents the unknown state at a given time instant and the observations y at the current time instant is the current image frame.

3.2 Robust color-based observation model

Various observation models have been proposed for template-based tracking with special attention being given to robustness in clutter and occlusions and in the adaptation of the observation model (e.g. [6] [10]). Recently, attention has been drawn to color-based tracking [12] [7].

Our observation model is initialized at the first frame of the sequence when a set of N windows are centered by the user around the facial micro-features that will be tracked. Let us denote with o_i the template feature vector, which contains the RGB color information at window i. Obviously, o_i has dimensionality equal to three times the number of pixels in window i.

We need to define $p(y|\alpha_i)$. Let us denote with q_i the template feature vector that contains the RGB color information at the window around α_i. We propose a color-based difference between the vectors o_i and q_i that is invariant to global changes in the intensity as follows:

$$c(o_i, q_i) = \left(\frac{o_i}{E\{o_i\}} - \frac{q_i}{E\{q_i\}} \right) \tag{8}$$

where $E\{o_i\}$ is the mean of the vector o_i, that is the average intensity of the color template o_i. It is easy to show that the proposed color difference vector $c(o_i, q_i)$ is invariant to global changes in the light intensity.

Finally, we define the scalar color distance using a robust function ρ. Let us denote here with j the pixel index and with $c_j(o_i, q_i)$ the color difference at pixel j. The scalar color distance is then defined as:

$$d(o_i, q_i) = E_j \left\{ \rho(c_j(o_i, q_i)) \right\} \tag{9}$$

where the robust function that has been used in our experiments is the L_1 norm. Then, we define the observation likelihood as:

$$p(y|\alpha_i) = e^{-\frac{d(o_i, q_i)}{\sigma_i}} \tag{10}$$

where σ_i is a scaling parameter. Note, that the color-based distance as defined in eq. 8 becomes unreliable as the templates become darker. In order to compensate for this fact we adapt σ_i to the average intensity of each template by letting $\sigma_i = 1.5/E\{o_i\}$.

3.3 Transition models and priors

Once the observation model is defined we need to model the transition density $p(\alpha_i|\alpha^-)$ and to specify the scheme for reweighting the particles, either by eq. 5, or by eq. 7. To avoid sampling from a complicated distribution we consider that $p(\alpha_i|\alpha^-) = p(\alpha_i|\alpha_i^-)$, that is each feature can be independently propagated. We use a very simple zero order model with Gaussian noise, that is $p(\alpha_i|\alpha_i^-) = \alpha_i^- + \mathcal{N}(0, \sigma)$.

In order to define the reweighting scheme we utilize a more elaborate scheme that uses training data to learn the interdependencies between the positions of the facial micro-features. In our modeling, the ratio of eq. 5 is approximated with a *prior* on the relative positions of the facial features. In the graphical model of fig. 1 that would correspond to the evaluation related to the dashed lines. This is evaluated with Parzen density estimation, that is,

$$\pi_k = \sum_j \phi(d(q(s_k)), h_j), \sigma_p) \tag{11}$$

where the collection $\{h_j\}$ is the collection of training data and $q(.)$ is a transformation function that registers the data from the current face (i.e. the particle s_k) to the training data (i.e. to the collection $\{h_j\}$). For our experiments we removed the translational component by compensating for the position of the facial features at the first frame and for the position of a stable facial feature (such as the nose) at the current frame. Scaling is done by person-related scaling factors (such as the dimensions of the mouth of the subject) that are estimated from the positions of the facial features at the first frame. With $\phi(.)$ we denote the Parzen kernel, which in our case is a Gaussian kernel with standard deviation σ_p. Finally, with $d(q(s_k)), h_j)$ we denote the distance function between the registered particle s_k and a training

Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)
0-7695-2122-3/04 $ 20.00 © 2004 IEEE

datum h_j. In our experiments we have used as $d(.,.)$ the Euclidean distance.

4. Experimental Results

We have applied the proposed method to a number of image sequences and here we present results for tracking five facial features, that is the position of the nose, the mouth corners and the upper and lower lip. We define small rectangular templates around each mouth feature and a larger template around the nose.

We present comparative results with a standard auxiliary particle filtering that attempts to track independently each of the facial micro-features. The results are summarized in fig. 2. As expected, the nose is tracked reliably for the whole sequence, since its template is rather large and the appearance does not change due to facial expressions. However, the tracking of mouth corners fails at the first facial expressions that affects their appearance for long time. Similarly, at a later time instant, occlusion of the upper lip causes a failure of the tracking of the corresponding template.

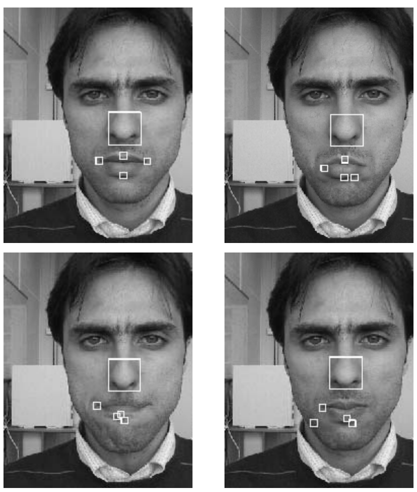

Figure 2: Tracking independently facial micro-features with 1000 particles. Results, from left to right and top to bottom, are presented for frames 1 (user initialization), 35, 51 and 101.

The results with the proposed method are summarized in fig. 3. The use of the priors provide constraints that are sufficient for the reliable tracking of the templates at the presence of appearance changes due to facial expressions. Furthermore, the tracker can successfully recover from long term occlusions, some of them occurring for more than 40 frames. This is achieved with only 100 samples in total, which is an order of magnitude less than the number of particles that have been used for tracking each template independently. Note however, that during occlusions the estimated positions of the occluded parts do not very often coincide with their true positions (e.g. frame 51 at fig. 3).

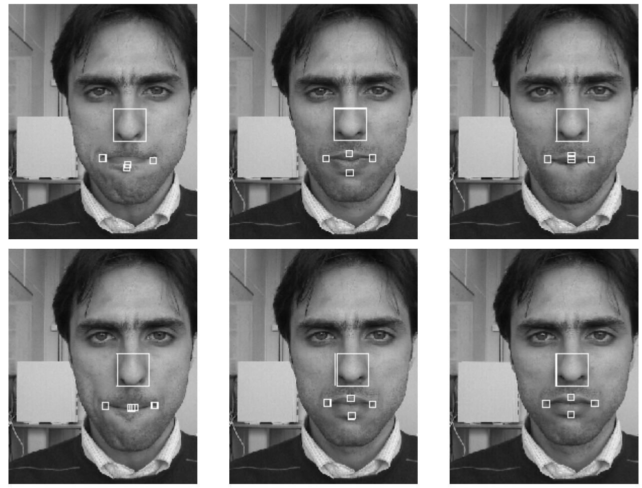

Figure 3: Proposed method with 100 particles. Results are presented for frames 51, 101, 151, 201, 251 and 301.

Finally, we present results for another image sequence in which the subject exhibits another set of facial expressions and talks naturally. As expected, independent tracking of the facial features, even with a large number of samples, fails soon after the first facial expression and, for some of the facial features, it never recovers. To compare our method with a particle filtering that utilizes the priors, we have devised a modified version of the standard auxiliary particle filtering. This scheme utilizes the priors $p(\alpha)$ by using the function $p(y|s_k)p(s_k)$ as the observation model. The results improve considerably in comparison to tracking each feature independently, but still (fig. 4), the tracking is lost even with a relatively large number of particles.

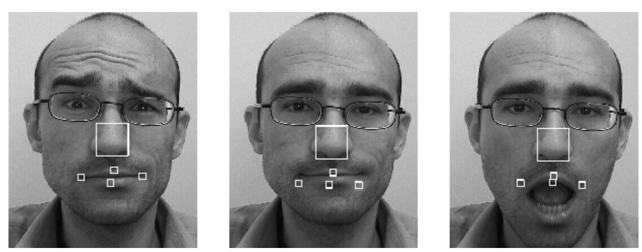

Figure 4: Modified auxiliary particle filtering with priors and 1000 particles. Results are presented for frames 51, 101, 151.

In comparison, the proposed method clearly outperforms the independent feature tracker and the modified auxiliary particle filtering scheme. We have used the same training

Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)
0-7695-2122-3/04 $ 20.00 © 2004 IEEE

data to learn the priors and the same parameters in our transition and observation models. With only 100 samples we were able to reliably track the mouth features in the presence of different facial expressions. The proposed method has always been able to recover from errors and track reliably the facial features for the rest of the sequence.

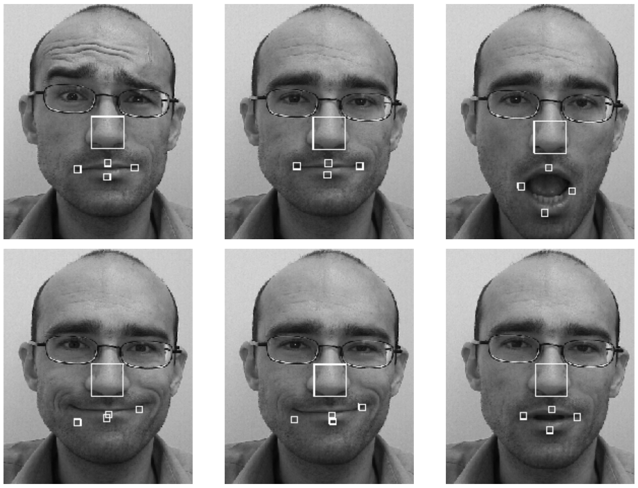

Figure 5: Proposed method with 100 particles. Results are presented for frames 51, 101, 151, 201, 251 and 301.

5. Conclusions

We have presented a method for particle filtering that overcomes some of the limitations of the classical particle filters in the case that the likelihood can be factorized. We devise a proposal density that is the product of the marginal posteriors of each of the state partitions. With show that with the proposed method we end up in a scheme that requires sampling from simple transition probabilities and only evaluation of more complicated interdependencies between the partitions of the state. We utilize the developed framework for the problem of template-based tracking of multiple facial features. We have devised a color-based scheme and we have incorporated prior knowledge of the relative positions of the facial features that is learned from training data. We have experimentally shown that the proposed method clearly outperforms the auxiliary particle filter both for multiple independent template tracking and when prior knowledge is utilized. For future work we consider the use of more sophisticated methods for evaluating the prior (or the transition) density than the simple Parzen window approach that we have used and the adaptation of the appearance model in the particle filtering framework (e.g. [5]).

Acknowledgements

This work is supported by the Innovational Research Grant of the Dutch Scientific Organisation (EW-639.021.202).

References

[1] J. Deutscher, A. Davison, and I. Reid. Automatic partitioning of high dimensional search spaces associated with articulated body motion capture. In *Int' Conf. Computer Vision and Pattern Recognition*, Dec. 2001. Kauai, Hawaii.

[2] M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *Int'l Journal of Computer Vision*, 29(1):5–28, 1998.

[3] M. Isard and A. Blake. Icondensation: Unifying low-level and high-level tracking in a stochastic framework. In *European Conf. on Computer Vision*, pages 893–908, Jun. 1998. Freiburg, Germany.

[4] J. MacCormick and A. Blake. Probabilistic exclusion and partitioned sampling for multiple object tracking. *Intl J. Computer Vision*, 39(1):57–71, 2000.

[5] K. Murphy and S. Russell. *Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks*. Springer-Verlag, 2001.

[6] H.T. Nguyen, M. Worring, and R.vd. Boomgaard. Occlusion robust adaptive template tracking. In *Int'l Conf. Computer Vision*, 2001. Vancouver, Canada.

[7] P. Perez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *European Conf. on Computer Vision*, volume 1, pages 661–675, May 2002. Copenhagen, Denmark.

[8] M.K. Pitt and N. Shephard. Filtering via simulation: auxiliary particle filtering. *J. American Statistical Association*, 94:590–, 1999.

[9] C. Sminchisescu and B. Triggs. Covariance scaled sampling for monocular 3d body tracking. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pages 447–454, 2001. Maui, Hawaii.

[10] J. Vermaak, P. Perez, M. Gangnet, and A. Blake. Towards improved observation models for visual tracking: Selective adaptation. In *European Conf. on Computer Vision*, pages 645–660, 2002. Copenhagen, Denmark.

[11] Y. Wu, G. Hua, and T. Yu. Tracking articulated body by dynamic markov network. In *Int'l Conf. on Computer Vision*, Oct. 2003. Nice, France.

[12] Y. Wu and T. Huang. A co-inference approach to robust tracking. In *Intl. Conf. on Computer Vision*, volume 2, pages 26–33, 2001. Vancouver, Canada.

Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)