IDIAP RESEARCH REPORT

An implicit motion likelihood for tracking with particle filters

Jean-Marc Odobez *   Sileye Ba *
Daniel Gatica-Perez *

IDIAP–RR 03-15

Mai, 16 2003

submitted for publication

Dalle Molle Institute for Perceptual Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 – 27 – 721 77 11
fax +41 – 27 – 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

* IDIAP, Martigny, Switzerland

IDIAP Research Report 03-15

An implicit motion likelihood for tracking with particle filters

Jean-Marc Odobez   Sileye Ba   Daniel Gatica-Perez

Mai, 16 2003

submitted for publication

Abstract. Particle filters are now established as the most popular method for visual tracking. Within this framework, it is generally assumed that the data are temporally independent given the sequence of object states. In this paper, we argue that in general the data are correlated, and that modeling such dependency should improve tracking robustness. To take data correlation into account, we propose a new model which can be interpreted as introducing a likelihood on implicit motion measurements. The proposed model allows to filter out visual distractors when tracking objects with generic models based on shape or color distribution representations, as shown by the reported experiments.

1 Introduction

Visual tracking is an important problem in computer vision, with applications in teleconferencing, visual surveillance, gesture recognition, and vision based interfaces [3]. Though tracking has been intensively studied in the literature, it is still a challenging task in adverse situations, due to the presence of ambiguities (e.g. when tracking an object in a cluttered scene or when tracking multiple instances of the same object class), the noise in image measurements (e.g. lighting problems), and the variability of the object class (e.g. pose variations).

In the pursuit of robust tracking, Sequential Monte Carlo methods [1, 5, 3] have shown to be a successful approach. In this Bayesian framework, the probability of the object configuration given the observations is represented by a set of weighted random samples, called particles. This representation allows to simultaneously maintain multiple-hypotheses in the presence of ambiguities, unlike algorithms that keep only one configuration state [4], which are therefore sensitive to single failure in the presence of ambiguities or fast or erratic motion.

Visual tracking with a particle filter requires the definition of two main elements : a data likelihood term and a dynamical model. The data likelihood term evaluates the likelihood of the current observation given the current object state, and relies on the object representation we have chosen. The object representation corresponds to all the information that explicitly or implicitly characterize the object like the target position, geometry, appearance, motion etc. Parametrized shapes like splines [3] or ellipses [12] and color distributions [8, 4, 7, 12] are often used as target representation. One drawback of these generic representations is that they are quite unspecific which augment the chances of ambiguities. One way to improve the robustness of a tracker consists of combining low-level measurements such as shape and color [12]. A step further to render the target more discriminative is to use appearence-based models such as templates [9, 10], leading to very robust trackers. However, such representations do not allow for large changes of appearence, unless adaptation is performed or more complex global appearence models are used (e.g. eigen-space [2] or set of examplars [11]).

The dynamical model characterizes the prior on the state sequence. Examples of such models range from simple constant velocity models to more sophisticated oscillatory ones or even mixtures of these [6]. In the particle filter framework, the dynamics are used to predict the new state hypotheses where the data likelihood will be evaluated. Thus, the dynamical model implicitly define some search range for the new hypotheses. The difficulty of modeling the dynamics arises from the two contradictory objectives it should fulfill. On one hand, the search space should be small enough to avoid the tracker being confused by distractors in the vicinity of the true object configuration, a situation that is likely to happen for unspecific object representations such as generic shapes or color distributions. On the other hand, it should be large enough so that it can cope with abrupt motion changes.

In this paper we propose a new tracking method based on the particle filter algorithm. More precisely, we argue that a standard hypothesis of this filter, namely the independence of observations given the state sequence, is inappropriate in the case of visual tracking. In this view, we propose a model that assumes that the current observation depends on the current and previous object configuration as well as on the past observation. As we will show, the proposed model can be exploited to introduce an implicit object motion likelihood in the data term. The benefits of this new model are two-fold. First, by exploiting a kind of template correlation between successive images, it will turn generic trackers like shape or color histogram trackers into more specific ones without resorting to complex appearence based models. Second, as a consequence, it reduces the sensitivity of the algorithm to the size of the search range, since when using a large size, potential distractors should be filtered out by the introduced correlation factor.

The rest of the paper is organized as follows. In the next Section, we briefly present the standard particle filter algorithm. Our model is explained in Section 3. Section 4 presents the results and some concluding remarks.

2 Particle filtering

Particle filtering is a technique for implementing a recursive Bayesian filter by Monte-Carlo simulations. The key idea is to represent the required density function by a set of random samples with associated weights. Let $c_{0:k} = \{c_l, l = 0, \ldots, k\}$ (resp. $z_{1:k} = \{z_l, l = 1, \ldots, k\}$) represents the sequence of states (resp. of observations) up to time k. Furthermore, let $\{c^i_{0:k}, w^i_k\}^{N_s}_{i=1}$ denote a set of weighted samples that characterizes the posterior probability density function (pdf) $\mathrm{p}(c_{0:k}|z_{0:k})$, where $\{c^i_{0:k}, i = 1, \ldots, N_s\}$ is a set of support points with associated weights w^i_k. The weights are normalized such that $\sum_i w^i_k = 1$. Then, a discrete approximation of the true posterior at time k is given by :

$$\mathrm{p}(c_{0:k}|z_{1:k}) \approx \sum_{i=1}^{N_s} w^i_k \delta(c_{0:k} - c^i_{0:k}) \,. \tag{1}$$

The weights are chosen using the principle of Importance Sampling (IS). More precisely, suppose that we could draw the samples $c^i_{0:k}$ from an importance (also called proposal) density $q(c_{0:k}|z_{1:k})$. Then the proper weights in (1) that lead to an approximation of the posterior are defined by :

$$w^i_k \propto \frac{\mathrm{p}(c^i_{0:k}|z_{1:k})}{q(c^i_{0:k}|z_{1:k})} \,. \tag{2}$$

The goal of the particle filtering algorithm is the recursive propagation of the samples and estimation of the associated weights as each measurement is received sequentially. To this end, the recursive equation of the posterior is employed :

$$\mathrm{p}(c_{0:k}|z_{1:k}) = \frac{\mathrm{p}(z_k|c_{0:k}, z_{1:k-1})\mathrm{p}(c_k|c_{0:k-1}, z_{1:k-1})}{\mathrm{p}(z_k|z_{1:k-1})} \times \mathrm{p}(c_{0:k-1}|z_{1:k-1}) \,, \tag{3}$$

and we assume a similar factorized form for the proposal :

$$q(c_{0:k}|z_{1:k}) = q(c_k|c_{0:k-1}, z_{1:k})q(c_{0:k-1}|z_{1:k-1}), \tag{4}$$

From there, three hypotheses are commonly made to derive the standard particle filter :

1. The state sequence $c_{0:k}$ follows a first-order Markov chain model, characterized by the definition of the dynamics $\mathrm{p}(c_k|c_{k-1})$.

2. The observations $\{z_k\}$, given the sequence of states, are independent. This leads to $\mathrm{p}(z_{1:k}|c_{0:k}) = \prod_{i=1}^{k} \mathrm{p}(z_k|c_k)$, which requires the definition of the individual data-likelihood $\mathrm{p}(z_k|c_k)$;

3. The prior distribution $\mathrm{p}(x_{0:k})$ is employed as importance function. In this case, $q(c_k|c_{0:k-1}, z_{1:k}) = \mathrm{p}(c_k|c_{k-1})$.

Replacing (4) and (3) in (2) and using the three hypotheses, we obtain the following recursive equation for the weight :

$$w^i_k \propto w^i_{k-1}\ \mathrm{p}(z_k|c^i_k). \tag{5}$$

It is known that importance sampling is usually inefficient in high-dimensionnal spaces [5], which is the case of the state space $c_{0:k}$ as k increases. To solve this problem, an additional resampling step is necessary, whose effect is to eliminate the particles with low importance weights and to multiply particles having high weights, giving rise to more variety around the modes of the posterior after the next importance sampling step.

Altogether, we obtain the standard particle filter that is displayed in Fig. 2.

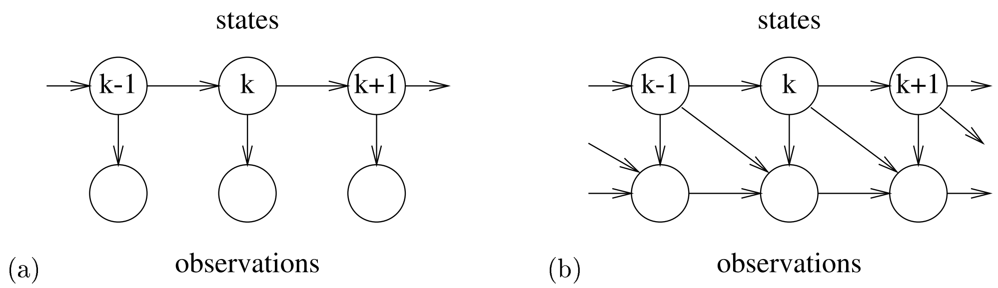

Figure 1: Graphical models for tracking. (a) standard model (b) proposed model.

1. Initialisation :

   For $i = 1, \ldots, N_s$, sample $c_0^i \sim \mathrm{p}(c_0)$ and set $k = 1$

2. Importance sampling step :

   For $i = 1, \ldots, N_s$, sample $\tilde{c}_k^i \sim \mathrm{p}(c_k | c_{k-1}^i)$ and evaluate the importance weights :

   $$\tilde{w}_k^i \propto w_{k-1}^i \ \mathrm{p}(z_k | \tilde{c}_k^i)$$

3. Selection step :

   Resample with replacement N_s particles $\{c_k^i, w_k^i = \frac{1}{N_s}\}$ from the sample set $\{\tilde{c}_k^i, \tilde{w}_k^i\}$.

   Set $k = k + 1$ and go to step 2

Figure 2: The bootstrap filter algorithm.

3 The proposed model

In this Section, we propose a new method that implicitly incorporates motion information into the measurement process. This is achieved by modifying the traditional graphical model represented in Fig. 1a, by making the current observation dependent not only on the current object configuration but also on the object configuration and observation at the previous instant (see Fig. 1b). We first provide arguments for this change and then present more precisely our approach.

3.1 Revisiting particle hypothesis

The filter described in Fig. 2a is based on the standard probabilistic model for tracking displayed in Fig. 1a and corresponding to hypotheses (1) and (2) of the previous section.

The first hypothesis is a reasonable one. Objects (and the camera) usually follows some dynamics governed by the laws of physics. The questions that arise here are more a matter of how complicated the dynamical model is, how well we can learn it, and how accurate it is to the application.

In visual tracking, the second hypothesis does not seem to be a sensible assumption.[1] In most of the tracking algorithms, the configuration state includes the parameters of a geometric transformation $\mathcal{T}$. Then, the measurements consist of implicitly or explicitly extracting some part of the image by :

$$\tilde{z}_{c_k}(\mathbf{r}) = z_k(\mathcal{T}_{c_k}\mathbf{r}) \qquad \forall \mathbf{r} \in R\,, \tag{6}$$

[1]For contour tracking, the assumption holds as the temporal auto-correlation function of contours is peaked.

Figure 3: Images at time t and $t+3$. The two local patches corresponding to the head and extracted from the two images are strongly correlated.

where $\mathbf{r}$ denotes a pixel position, R denotes a fixed reference region, and $\mathcal{T}_{c_k}\mathbf{r}$ represents the application of the transform $\mathcal{T}$ parameterized by c_k to the pixel $\mathbf{r}$. The data likelihood is then computed from this local : $\mathrm{p}(z_k|c_k) = \mathrm{p}(\tilde{z}_{c_k})$, with $\tilde{z}_{c_k}$ denoting the patch casted in the reference frame according to (6). However, if c_{k-1} and c_k correspond to two consecutive states of a given object, it is reasonable to assume :

$$\tilde{z}_{c_k}(\mathbf{r}) = \tilde{z}_{c_{k-1}}(\mathbf{r}) + noise \qquad \forall \mathbf{r} \in R \tag{7}$$

where *noise* usually takes some small value. This point is illustrated in Figure 3. Equation (7) is at the core of all motion estimation and compensation algorithms like MPEG and is indeed a valid hypothesis. Thus, according to this equation, the independence of the data given the sequence of states is not a valid assumption. More precisely :

$$\mathrm{p}(z_k|z_{1:k-1}, c_{1:k}) \neq \mathrm{p}(z_k|c_k) \tag{8}$$

which means that we can not reduce the left hand side to the right one as usually done. A better model for visual tracking is thus represented by the graphical model of Fig. 1b.

The new model can be incorporated in the particle framework. Starting from equation (3), all the subsequent equations can be derived exploiting this new assumption. If the proposal is equal to the dynamic model, it simply leads to the following update equation:

$$w_k^i \propto w_{k-1}^i \ \mathrm{p}(z_k|z_{k-1}, c_k^i, c_{k-1}^i) \tag{9}$$

in replacement of equation (5), which can in turn be used in the bootstrap algorithm.

3.2 Object representation, state space and dynamics definition

We follow an image-based standard approach, where the object is represented by a region R subject to some valid geometric transformation, and is characterized either by a shape or by a feature (we use color) distribution in this region. For the geometric transformations, we have chosen a subspace of the affine transformations comprising a translation $\mathbf{T}$ and a scaling factor s. Furthermore, a first-order dynamical model is defined on these parameters augmented with their respective speed. More precisely, defining the state as $c_k = (\alpha_k, \dot{\alpha}_k)$, with $\alpha = (\mathbf{T}, s)$, the dynamical model is defined by $c_k = Ac_{k-1} + B\mathbf{w}_k$ where A and B are the parameters of the model and $\mathbf{w}$ is a white noise process.

3.3 Data likelihood modeling

To implement the new particle filter, we considered the following data likelihood :

$$\mathrm{p}(z_k|z_{k-1}, c_k, c_{k-1}) = \mathrm{p}_c(z_k|z_{k-1}, c_k, c_{k-1}) \times \mathrm{p}_o(z_k|c_k) \tag{10}$$

where the first probability $\mathrm{p}_c()$ models the correlation between the two observations and $\mathrm{p}_o()$ is an object likelihood. This choice decouples the model of the correlation existing between two images, whose implicit goal is to insure that the object trajectory follows the optical flow field implied by the sequence of images, from the shape or appearence object model. We assumed that these two terms are independent. When the object is modeled by a shape, this assumption is valid since shape measurement will mainly involve measurements on the border of the object, while the correlation term will apply to the regions inside the object. When a color representation is employed, which involves measurements inside the object, the independance assumption might not hold in strict terms.

Visual object measurement

For the experiments, we considered both contour models and color models.
Contour model :
The observation model assumes that objects are embedded in clutter. Edge-based measurements are computed along L normal lines to a hypothesized contour, resulting for each line l in a vector of candidate positions $\{\nu_m^l\}$ relative to a point lying on the contour ν_0^l. With some usual assumptions [3], the likelihood can be expressed as

$$\mathrm{p}_o(z_k|c_k) \propto \prod_{l=1}^{L} \max\left(K, \exp(-\frac{\|\hat{\nu}_m^l - \nu_0^l\|^2}{2\sigma^2})\right), \tag{11}$$

where $\hat{\nu}_m^l$ is the nearest edge on l, and K is a constant used when no edges are detected.
Color model :
As color models we used color distributions represented by normalized histograms in the HSV space and gathered inside the candidate region $R(c_k)$ associated with the state c_k. To be robust to illumination effects, we only considered the HS values. Besides, to add some spatial layout information, the candidate region was split into N_r sub-regions $R_r(c_k)$. Then, a multidimensional histogram was computed (and normalized), resulting in a vector $\mathbf{b}(c_k) = (\mathrm{b}^j(c_k))_{j=1..N}$, where $N = N_h \times N_s \times N_r$ with N_h and N_s representing the number of bins along the hue and saturation dimensions respectively, and where an index j correspond to a triple (h, s, r), with h and s denoting hue and saturation bin numbers, and r the region number.
At time k, the candidate color model $\mathbf{b}(c_k)$ is compared to a reference color model $\mathbf{b}_{ref}$. Currently, we use the histogram computed in the first frame as reference model. As a distance measure we employed a Bhattacharyya distance measure [4, 7]:

$$D_{bhat}(\mathbf{b}(c_k), \mathbf{b}_{ref}) = \left(1 - \sum_{j=1}^{N} \sqrt{\mathrm{b}^j(c_k)\mathrm{b}^j_{ref}}\right)^{1/2}$$

and assumed that the probability distribution of the square of this distance for a given object was following an exponential law,

$$\mathrm{p}_o(z_k|c_k) \propto \exp\{-\lambda D^2_{bhat}(\mathbf{b}_k(c_k), \mathbf{b}_{ref})\}.$$

Image correlation measurement

We model this term in the following way :

$$\mathrm{p}_c(z_k|z_{k-1}, c_k, c_{k-1}) \propto \exp^{-\lambda_c \mathrm{d}_c(\tilde{z}_{c_k}, \tilde{z}_{c_{k-1}})} \tag{12}$$

where d_c denotes a distance between two image patches. Many such distances have been defined and used in the literature [9, 11]. The choice of the distance should take into account the followings considerations, which are illustrated in Fig. 4 :

Figure 4: Tracking the white box on the left image : green boxes on the right should have a higher probability than the red one. Green boxes, generated from the dynamical model, will almost never fall on the optimal match (white box on the right). If the head and the background undergo different motion, the predicted boxes associated with the blue box (on the left) should in general get a lower probability than those associated with the white box.

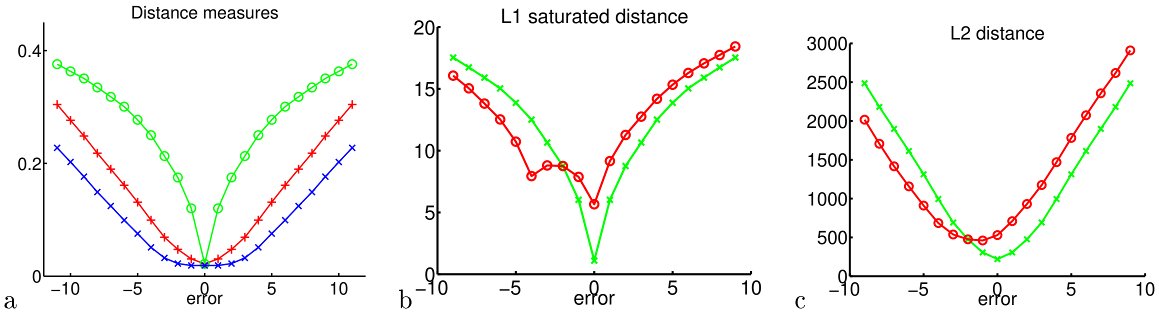

Figure 5: a) Distance profiles (scaled), in function of the error (in pixel units) with respect to the perfect match (error=0). The size of the considered patches is 40x40, and the values have been averaged over 20 different patches from an image. Distances : (red,+) L2 (green,o) L1 saturated (blue,x) Haussdorf. b) and c) Distance profiles for a predicted particle that originaly covers only the object (green,x) or part of the object (60%) and of the background (red,o). b) L1 saturated distance c) L2 distance. The difference between the background motion and object motion is 4, and the error is set to 0 when the predicted particle has the same motion than the object.

1. The distance should model the underlying motion content, i.e. the distance should increase as the error in the predicted configuration grows. For example, in Fig. 4, the predicted red box should receive a larger distance than the green boxes if they originate from the same particle in the previous frame (the white box).

2. The random nature of the prediction process in the SMC filtering will rarely produce configurations corresponding to exact matches, as shown in Fig. 4.

3. The distance between a particle and its descendents should in general be higher for particles that cover parts of an object and the background than for particles covering only an object, when object and background undergo different motion.

Fig. 5 displays some distances measures in function of the predicted error.[2] The values have been averaged by performing the same experiment on random patches taken in an image. This Figure shows that a robust distance (here a L1 saturated distance) leads to a peaked distance profile that

[2]i.e. the error (in pixels) between the predicted position and the one corresponding to the optimal match.

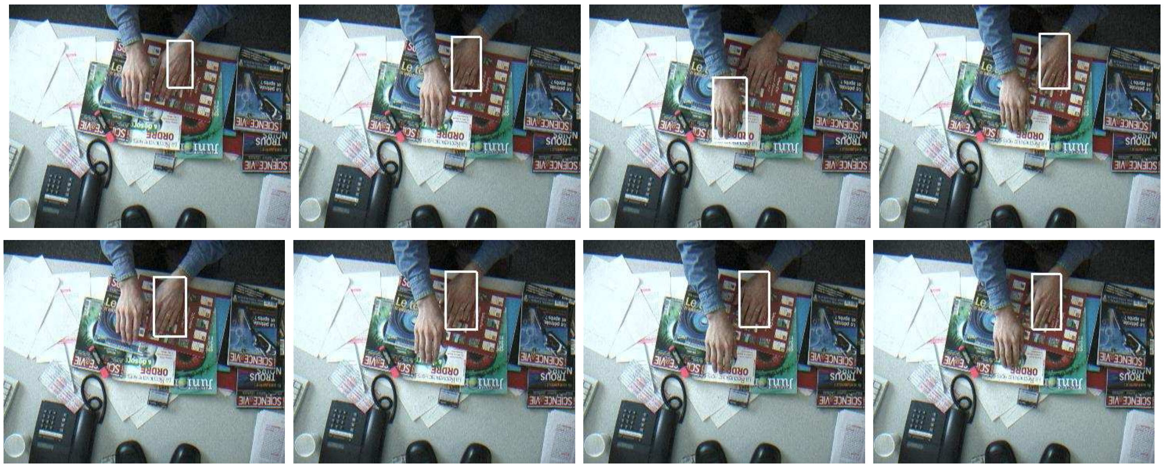

Figure 6: Tracking the left hand at time t_{840}, t_{845}, t_{850} and t_{855}. Top: histogram tracker. Bottom: histogram plus correlation. White box: most likely particle. Green boxes: other likely particles.

would not be so appropriate to account for point 2 above. Moreover, the robust distance doesn't satisfy so well point 3, since particles fully covering the object but with small predicted errors (e.g. 1 pixel) receive distance measures similar to those of particles covering object and background with a predicted configuration that matches very well the object motion, as illustrated in Figure 5c. On the other hand, the L2 norm is more appropriate[3]. Thus, we selected the L2 distance, which indeed corresponds to an additive Gaussian noise model in Eq.(7),

$$\mathrm{d}_c(\tilde{z}_{c_k}, \tilde{z}_{c_{k-1}}) = \sum_{\mathbf{r}\in R} \rho(\tilde{z}_{c_k}(\mathbf{r}) - \tilde{z}_{c_{k-1}}(\mathbf{r})) \text{ with } \rho(x) = x^2 \tag{13}$$

and λ_c is set to $\frac{1}{2\sigma_c^2}$ where σ_c denotes the noise standard deviation.

Regarding the above equation, it is important to emphasize that the method is not performing template matching, as in [9]. No object template is learned off-line or defined at the begining of the sequence, and the tracker does not maintain a single template object representation at each instant of the sequence. Thus, the correlation term is not object specific (except through the definition of the reference region R). A particle "lying" on the background would thus receive a high weight if the predicted motion is in adequation with background motion. Nevertheless, the methodology could be the extended to be more object dependent, by allowing the region R to vary over time (using examplars for instance), or by introducing spatially variant noise in the definition of the correlation term.

4 Results

We present results on three sequences. In all cases, the tracker is initialized by hand.

Hand tracking

In this sequence (see Fig. 6), the two hands are active, grasping or displacing objects, or waving from left to right simultaneously. The hands move over a cluttered background, both in shape and color, and the camera is moving as well. As a hand model, we used a box split in 9 regions[4] of equal size

[3]The need for point 3 is arguable. If partial occlusion is expected, a robust norm might be more appropriate.

[4]When using less regions, the histogram tracker was frequently jumping from one hand to the other. This was almost never the case of the histogram-correlation tracker, but because of the crude spatial representation, the tracker would often lock on a small part of the hand only.

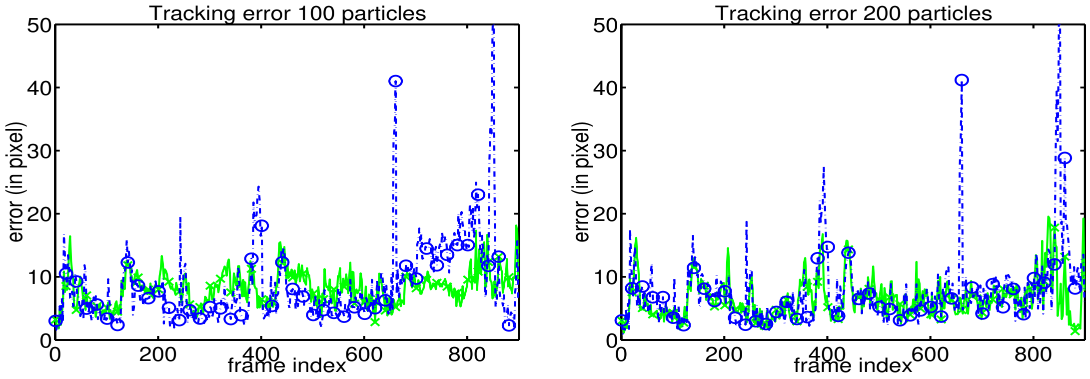

Figure 7: Error between the position of the hand provided by the tracker (box center) and a hand-labeled hand position. The plotted curves correspond to an average over 10 runs of each tracker. (+,green) HC and (o,blue) H trackers. Left, 100 particles. Right, 200.

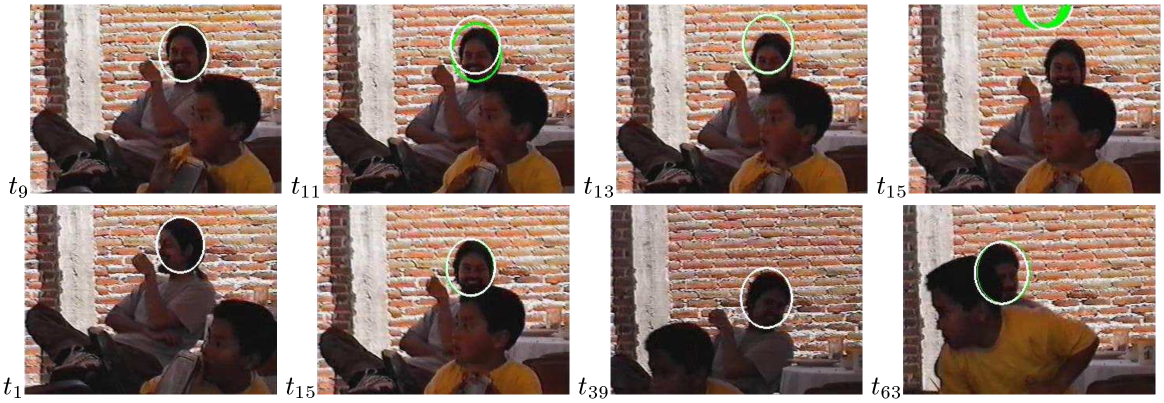

Figure 8: Head tracking 1 : top, shape tracker only. Bottom, shape+correlation tracker.

with associated color histograms. Tracking the right hand with the histogram only (H) or histogram plus correlation (HC) trackers produced similar results. However, when tracking the left hand, the H tracker was several times confused by the presence of the right hand, as illustrated in Fig. 6, top row. This confusion usually lasted for several frames, but most of the time, the H tracker was able to resume. The HC tracker did never undergo this ambiguity problem. This is demonstrated in Fig. 7, which displays the error in the hand position averaged over 10 runs.

Head tracking

The first sequence (Fig. 8) illustrates the benefit of the method in the presence of ambiguities. Despite the presence of a highly textured background producing very noisy shape measurements, the camera and head motion, the change of appearence of the head, and partial occlusion, the head is correcly tracked using our method. Whatever the number of particles or the noise variance in the dynamical model, the shape tracker alone is unable to perform a correct tracking after the t_{12} instant. Note that for this sequence, the histogram only tracker fails as well but a joint histogram and shape tracker was successful.

In the second sequence (Fig. 9), histogram models are inappropriate, as the tracked person is undergoing several 360^o head turns. In this sequence, the shape tracker is perturbed around frame 60 by different factors : abrupt vertical camera motion and the absence of head contours as the head moves in front of the bookshelves. The tracker is then unable to recover. On the other hand, the shape and correlation tracker successfuly tracks the head turn [5], remains less influenced by the lack of contour measurements and abrupt motion changes, and successfully tracks the head over the rest of the sequence.

5 Conclusion

We proposed a method for visual tracking with particle filters, which takes into account the temporal correlation that exists between successive images of the same object by introducing a new data likelihood modeling. The presented model allows for the integration of motion measurements in an implicit way, which is helpful to remove tracking ambiguities that occurs when using generic shape-based or color-based object models.

6 Acknowledgments

This work was carried out in the framework of the Swiss National Center of Competence in Research (NCCR) on Interactive Multimodal Information Management (IM)2, and in the more specific context of the MUCATAR project (Multiple Camera Tracking and Activity Recognition).

References

[1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian. *IEEE Trans. Signal Processing*, pages 100–107, 2001.

[2] M. J. Black and A. D. Jepson. A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions. In *European Conf. on Computer Vision, ECCV-98*, pages 909–924, Freiburg, Germany, 1998. Springer-Verlag.

[3] Andrew Blake and Michael Isard. *Active Contours.* Springer, 1998.

[4] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *CVPR*, pages 142–151, 2000.

[5] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice.* Springer-Verlag, 2001.

[6] Michael Isard and Andrew Blake. A mixed-state CONDENSATION tracker with automatic model-switching. In *ICCV*, pages 107–112, 1998.

[7] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *Eur. Conf. on Computer Vision, ECCV'2002, LNCS 2350*, pages 661–675, June 2002.

[8] Y. Raja, S. McKenna, and S. Gong. Colour model selection and adaptation in dynamic scenes. In *5th European Conference on Computer Vision*, pages 460–474, 1998.

[9] J. Sullivan and Rittscher J. Guiding random particles by deterministic search. In *Int. Conf. on Computer Vision*, pages 323–330, 2001.

[10] Hai Tao, Harpreet S. Sawhney, and Rakesh Kumar. Object tracking with bayesian estimation of dynamic layer representations. *IEEE PAMI*, 24(1):75–89, 2001.

[5] The head turn is indeed a difficult case for the new method, as in the extreme case, the motion inside the head region indicates a right (or left) movement while the head outline remains static.

Figure 9: Top, shape tracker. Bottom, shape+correlation tracker. 800 particles.

[11] K. Toyama and A. Blake. Probabilistic tracking in a metric space. In *Proc.* 8^{th} *IEEE Int. Conf. Computer Vision*, Vancouver, July 2001.

[12] Y. Wu and T. Huang. A co-inference approach for robust visual tracking. In *Proc.* 8^{th} *IEEE Int. Conf. Computer Vision*, Vancouver, July 2001.