Probabilistic Observation Models for Tracking Based on Optical Flow

Manuel J. Lucena[1], José M. Fuertes[1], Nicolas Perez de la Blanca[2], Antonio Garrido[2], and Nicolás Ruiz[3]

[1] Departamento de Informatica, Escuela Politecnica Superior, Universidad de Jaen
Avda de Madrid 35, 23071 Jaen, Spain
{mlucena,jmf}@ujaen.es
[2] Departamento de Ciencias de la Computacion e Inteligencia Artificial
ETSII. Universidad de Granada
C/ Periodista Daniel Saucedo Aranda s/n
18071 Granada, Spain
{nicolas,agarrido}@ugr.es
[3] Departamento de Electronica, Escuela Universitaria Politecnica
Universidad de Jaen
C/ Alfonso X el Sabio 28, 23700 Linares, Spain
nicolas@ujaen.es

Abstract. In this paper, we present two new observation models based on optical flow information to track objects using particle filter algorithms. Although optical flow information enables us to know the displacement of objects present in a scene, it cannot be used directly to displace an object model since flow calculation techniques lack the necessary precision. In view of the fact that probabilistic tracking algorithms enable imprecise or incomplete information to be handled naturally, these models have been used as a natural means of incorporating flow information into the tracking.

1 Introduction

The probabilistic models applied to tracking [10, 9, 4, 14] enable us to estimate the *a posteriori* probability distribution of the set of valid configurations for the object to be tracked, represented by a vector $\mathbf{X}$, from the set of measurements $\mathbf{Z}$ taken from the images of the sequence, $p(\mathbf{X}|\mathbf{Z})$. The likelihood in the previous instant is combined with a dynamical model giving rise to the *a priori* distribution in the current instant, $p(\mathbf{X})$. The relation between these distributions is given by Bayes' Theorem:

$$p(\mathbf{X}|\mathbf{Z}) \propto p(\mathbf{X}) \cdot p(\mathbf{Z}|\mathbf{X})$$

The distribution $p(\mathbf{Z}|\mathbf{X})$, known as the *observation model*, represents the probability of the measurements $\mathbf{Z}$ appearing in the images, assuming that a specific configuration of the model in the current instant is known.

In this paper, two observation models are defined based on the optical flow of the sequence, checking its validity within a scheme of particle filter tracking.

F.J. Perales et al. (Eds.): IbPRIA 2003, LNCS 2652, pp. 462–469, 2003.

2 Optical Flow

The most well-known hypothesis for calculating the optical flow [7] assumes that the local intensity structures found in the image remain approximately constant over time, at least during small intervals of time. This is to say,

$$I_x u + I_y v + I_t = 0 \tag{1}$$

where I_x, I_y, I_t are partial derivatives of the image, and $\mathbf{v} = (u, v)$ represents the flow vector at each point. The problem is ill-posed, since it only has one equation for the calculation of two unknowns, which makes it necessary to use various additional restrictions, in the majority of cases based on smoothness [1, 12].

3 Dynamical Model

The tracking task involves localizing, in each frame of a sequence, the object associated to a state vector that characterizes evidence of the presence of a specific configuration of the model in question. Other authors have successfully used characteristics such as the gradient [2] or intensity distributions [14]. The model which represents the dynamical model of the object will provide an *a priori* distribution on all the possible configurations at the instant t_k, $p(\mathbf{X}(t_k))$, from the estimated distributions in the previous instants of time. In this paper, a second-order dynamical model has been used in which the two previous states of the object model are considered, and this is equivalent to taking a first-order dynamical model with a state vector for the instant t_k of the form [2]

$$\mathcal{X}_{t_k} = [\mathbf{X}_{t_{k-1}},\ \mathbf{X}_{t_k}]^T$$

The integration of the *a priori* distribution $p(\mathbf{X})$ with the set $\mathbf{Z}$ of the evidences present in each image, in order to obtain the *a posteriori* distribution $p(\mathbf{X}|\mathbf{Z})$, is obtained with Bayes' Theorem. This fusion of information can be performed, if the distributions are Gaussian, by using Kalman's Filter [10]. However, in general, the distributions involved in the process are normally not Gaussian and multimodal [4]. Sampling methods for modelling this type of distribution [6] have shown themselves to be extremely useful, and *particle filter* algorithms [8, 9, 5, 14] based on sets of weighted random samples, enable their propagation to be performed effectively.

4 Observation Models

4.1 Observation Model Based on Intensity Restrictions

In order to build this model, we will use a technique derived from the Lucas-Kanade algorithm[11] taking advantage of knowledge of the position of the flow discontinuities predicted by the object model. Our hypothesis is based on the fact that the point $\mathbf{x}$ of the model outline is situated on the real outline of the

object, and therefore we assume that the flow in a neighborhood of $\mathbf{x}$ shall only take two values: one on the inner part of the model, and the other on the outer part.

Let $\mathbf{x} = f(\mathbf{X}_{t_k}; \mathbf{m})$ (where $\mathbf{X}_{t_k}$ defines the specific configuration of the object model, and $\mathbf{m}$ is the parameter vector which associates each point within the model with a point on the image plane), a point belonging to the model outline at the instant t_k. Let S be a neighborhood of $\mathbf{x}$ subdivided in S_i and S_e (corresponding respectively to the parts of the neighborhood which remain towards the interior and exterior of the outline of the object), and $\mathbf{d}(\mathcal{X}_{t_k}, \mathbf{m})$ be calculated using the expression:

$$\mathbf{d}(\mathcal{X}_{t_k}, \mathbf{m}) = f(\mathbf{X}_{t_k}; \mathbf{m}) - f(\mathbf{X}_{t_{k-1}}; \mathbf{m}) \tag{2}$$

The system of equations [11] is therefore solved

$$\begin{bmatrix} \sum_{S_x} (I_x^{(k-1)})^2 & \sum_{S_x} I_x^{(k-1)} I_y^{(k-1)} \\ \sum_{S_x} I_x^{(k-1)} I_y^{(k-1)} & \sum_{S_x} (I_y^{(k-1)})^2 \end{bmatrix} \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} -\sum_{S_x} I_x^{(k-1)} I_t \\ -\sum_{S_x} I_y^{(k-1)} I_t \end{bmatrix} \tag{3}$$

with $I^{(k-1)}$ and $I^{(k)}$ being the images corresponding to the instants of time t_{k-1} and t_k. In order to obtain the flow vector $\mathbf{f}_{S_x} = (f_x, f_y)$, where S_x shall be S_i or S_e, respectively, I_x and I_y are the spatial derivatives of the image and

$$I_t(\mathbf{x}) = I^{(k)}(\mathbf{x} + \mathbf{d}(\mathcal{X}_{t_k}, \mathbf{m})) - I^{(k-1)}(\mathbf{x})$$

In this way, two different flow estimations are obtained, $\mathbf{f}_{S_i}(\mathcal{X}_{t_k}, \mathbf{m})$ and $\mathbf{f}_{S_e}(\mathcal{X}_{t_k}, \mathbf{m})$, corresponding to the inner and outer area of the neighborhood of $\mathbf{x}$, respectively.

The squared norm of the estimated flow vectors are then calculated, which is equivalent to obtaining the quadratic differences with the expected flow, which in this case equals zero:

$$Z_{S_i}(\mathcal{X}_{t_k}, \mathbf{m}) = \|\mathbf{f}_{S_i}(\mathcal{X}_{t_k}, \mathbf{m})\|^2, \qquad Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m}) = \|\mathbf{f}_{S_e}(\mathcal{X}_{t_k}, \mathbf{m})\|^2 \tag{4}$$

It should be noted that if the point $\mathbf{x}$ is really situated on a flow discontinuity, and the flow in S_i coincides with $\mathbf{d}(\mathcal{X}_{t_k}, \mathbf{m})$, the value of Z_{S_i} must be close to zero and the value of Z_{S_e} must be considerably greater. Using the following expression, these values may be combined and a value of $Z(\mathcal{X}_{t_k}, \mathbf{m})$ may therefore be obtained:

$$Z(\mathcal{X}_{t_k}, \mathbf{m}) = \begin{cases} \dfrac{Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m})}{Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m}) + Z_{S_i}(\mathcal{X}_{t_k}, \mathbf{m})} & \text{if } Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m}) \neq Z_{S_i}(\mathcal{X}_{t_k}, \mathbf{m}) \\ 1/2 & \text{if } Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m}) = Z_{S_i}(\mathcal{X}_{t_k}, \mathbf{m}) \end{cases} \tag{5}$$

The value of $Z(\mathcal{X}_{t_k}, \mathbf{m})$ satisfies the following properties:

- $0 \leq Z(\mathcal{X}_{t_k}, \mathbf{m}) \leq 1$
- If $Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m}) \gg Z_{S_i}(\mathcal{X}_{t_k}, \mathbf{m})$, then $Z(\mathcal{X}_{t_k}, \mathbf{m}) \to 1$, which indicates that the adjustment is much better in S_i than it is in S_e, and therefore the point must be situated exactly in a flow discontinuity, in which the inner area coincides with the displacement predicted by the model.
- If $Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m}) \ll Z_{S_i}(\mathcal{X}_{t_k}, \mathbf{m})$, then $Z(\mathcal{X}_{t_k}, \mathbf{m}) \to 0$. The adjustment is worse in the inner area than it is in the outer area, and therefore the estimated flow does not match the model's prediction.
- If $Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m}) = Z_{S_i}(\mathcal{X}_{t_k}, \mathbf{m})$, then the adjustment is the same in the inner area as it is in the outer area, and therefore the flow adequately matches the displacement predicted by the model, but it is impossible to guarantee that it is situated on a flow discontinuity. In this case, $Z(\mathcal{X}_{t_k}, \mathbf{m}) = 1/2$.

It is possible that some of the areas S_i or S_e lack enough *structure* to give a good flow estimate. In this paper, we have used the inverse *condition number* [13] of the coefficient matrix in the expression (3), $R = \lambda_{min}/\lambda_{max}$, in order to check the stability of the equation system, so that if it is too small ($< 10^{-10}$), it is necessary to discard the flow values obtained, and therefore $Z(\mathcal{X}_{t_k}, \mathbf{m})$=1/2.

We shall consider that the presence probability of the measurements obtained for the image, since they have been caused by the point of the outline corresponding to the vector $\mathbf{m}$ of the sample in question, defined by the vector $\mathcal{X}_{t_k}$, must be proportional to the function $Z(\mathcal{X}_{t_k}, \mathbf{m})$ obtained previously,

$$p(\mathbf{Z}|\mathcal{X}_{t_k}, \mathbf{m}_i) \propto Z(\mathcal{X}_{t_k}, \mathbf{m}_i) \tag{6}$$

and that, given the independence between the different points of the outline,

$$p(\mathbf{Z}|\mathcal{X}_{t_k}) \propto \prod_i Z(\mathcal{X}_{t_k}, \mathbf{m}_i) \tag{7}$$

4.2 Observation Model Based on Similarity Measures

If the prediction which the model makes is good and the intensity maps corresponding to the neighborhood of each point are superimposed, the inner part of the model must fit better than the outer part. In the model defined in this section, in order to estimate the observation probability of each point of the outline, similarity measurements shall be used to quantify the degree to which the inner part fits better than the outer part.

Let $\mathbf{x} = f(\mathbf{X}_{t_k}; \mathbf{m})$ be a point belonging to the model outline at the instant t_k, let S be a neighborhood of $\mathbf{x}$ subdivided in turn into S_i and S_e, let $\mathbf{d}(\mathcal{X}_{t_k}, \mathbf{m})$ be calculated from expression (2), and let $I^{(k-1)}$ and $I^{(k)}$ be images corresponding to the instants of time t_{k-1} and t_k. The quadratic errors are therefore calculated in the following way:

$$\begin{aligned} Z_{S_i}(\mathbf{X}_{t_k}, \mathbf{m}) &= \textstyle\sum_{S_i} W(\mathbf{x}) \Big(I^{(k-1)}(\mathbf{x}) - I^{(k)}(\mathbf{x} - \mathbf{d}(\mathcal{X}_{t_k}, \mathbf{m})) \Big)^2 \\ Z_{S_e}(\mathbf{X}_{t_k}, \mathbf{m}) &= \textstyle\sum_{S_e} W(\mathbf{x}) \Big(I^{(k-1)}(\mathbf{x}) - I^{(k)}(\mathbf{x} - \mathbf{d}(\mathcal{X}_{t_k}, \mathbf{m})) \Big)^2 \end{aligned} \tag{8}$$

where $W(\mathbf{x})$ is a weighting function. Two non negative magnitudes are obtained, that may be combined using expression (5), in order to obtain a value of $Z(\mathcal{X}_{t_k}, \mathbf{m})$. Since the magnitudes Z_{S_i} and Z_{S_e} are restricted, $Z(\mathcal{X}_{t_k}, \mathbf{m})$ may be considered to be proportional to the observation density $p(\mathbf{Z}|\mathcal{X})$, and therefore we again have:

$$p(\mathbf{Z}|\mathcal{X}_{t_k}, \mathbf{m}_i) \propto Z(\mathcal{X}_{t_k}, \mathbf{m}_i) \tag{9}$$

Supposing the measurements on each point are statistically independent, the following observation model is finally arrived at:

$$p(\mathbf{Z}|\mathcal{X}_{t_k}) \propto \prod_i Z(\mathcal{X}_{t_k}, \mathbf{m}_i) \tag{10}$$

5 Experiments

In order to check the validity of the observation models proposed, they were incorporated into the CONDENSATION algorithm [9], and their performance was compared with that of the observation model based on normals as proposed in [2].

For the experiments, two image sequences were used, lasting 10 seconds, with 25 frames per second, 320×240 pixels, 8 bits per band and pixel, corresponding to the movement of a hand over a background with and without noise. Results can be downloaded from `http://wwwdi.ujaen.es/~mlucena/invest.html`

5.1 Tracking an Object over a Background without Noise

In order to model the hand, an outline model based on a closed spline with 10 control points and a Euclidean similarity deformation space were used. A second-order dynamical model was used in which the object tended to maintain velocity, and a preliminary tracking was carried out of the hand by using the gradient observation model along the contour normals. With the data obtained, the multidimensional learning method proposed in [3, 2] was used to determine the dynamic parameters.

For the observation model based on contour normals, 20 normals were sketched for each sample. The observation model was applied with parameters $\alpha = 0.025$ and $\sigma = 3$, incorporated into the CONDENSATION algorithm with 200 samples. The initialization was carried out manually, indicating the position of the object in the first frame. Figure 1.a shows the weighted average of the distribution obtained.

The observation model based on intensity restrictions was used on the same 20 points along the outline, defining a neighborhood for each point of 7×7 pixels. In order to calculate the spatial derivatives, each frame was convolved with two Gaussian derivative masks, with vertical and horizontal orientations respectively, with $\sigma = 1.0$. The number of samples was also 200, and the results obtained are shown in Figure 1.b.

In order to apply the observation model based on similarity measures, the same conditions were used as in previous experiments (200 samples and 20 points along the outline, considering a neighborhood of 5×5 pixels for each point). The result obtained is illustrated in Figure 1.c.

5.2 Tracking an Object over a Background with Noise

In this case, the parameters of the dynamic model were adjusted manually for the first 50 frames, and these were used to learn dynamics and to perform an initial tracking of the sequence, using CONDENSATION with the observation model for the contour normals. From the results obtained, and using the same learning method as in the previous experiment, the dynamic parameters were determined.

In order to use the observation model based on contour normals, 18 normals were sketched to each outline. The number of samples was still 200, and the parameters for the observation model in this case were $\sigma = 3$ and $\alpha = 0.055$. The results are shown in Figure 2.a.

The same parameters were used in the observation model based on intensity restrictions as in the previous sequence, that is to say, spatial derivatives from Gaussian derivative masks with $\sigma = 1.0$, and neighborhoods of 7×7 pixels for each point. The results obtained with 200 samples are shown in Figure 2.b.

For the observation model based on similarity measures, neighborhoods of 5×5 pixels and 200 samples for the CONDENSATION algorithm were also used. The results obtained are shown in Figure 2.c.

6 Discussion and Conclusions

The observation model based on contour normals behaves appropriately in the two sequences (Figures 1.a and 2.a). At no time does the tracker lose the object, although it does have problems with noise in the sequence due to the presence of clutter.

For the first sequence, the model based on intensity restrictions, although never completely losing the object, does have problems focusing exactly on its outline. This is due to the absence of texture on the outer part of the object. However, for the second sequence, the presence of a background with a lot of noise is not only irrelevant but also favors the good behavior operation of the observation model.

Due to the fact that there is hardly any texture on the outer part of the object in the first sequence, the observation model based on similarity measures tends to minimize the inner measurement. Nevertheless, a slight deviation occurs at times towards the hand's shadow, since there is actually an ambiguity in the sequence, as the shadow moves jointly with the hand, and by only considering the optical flow, it is impossible to separate them. In the second sequence, there are no significant deviations from the real outline of the object.

The results obtained suggest that the observation models based on optical flow are, in a way, complementary to those based on gradient along normals. The

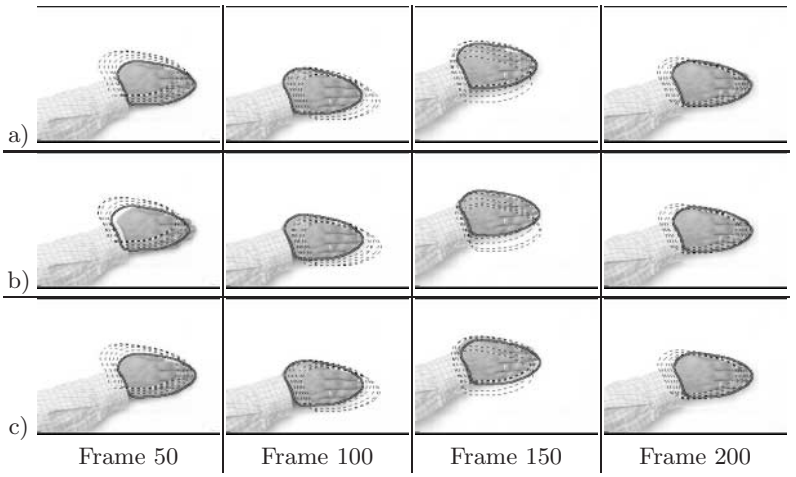

Fig. 1. a) Results obtained with the observation model for the contour normals. b) Results obtained with the observation model based on intensity restrictions. c) Results obtained with the observation model based on similarity measures. The distribution average appears in continuous line in the current frame, and the averages in some previous frames appear in dotted line

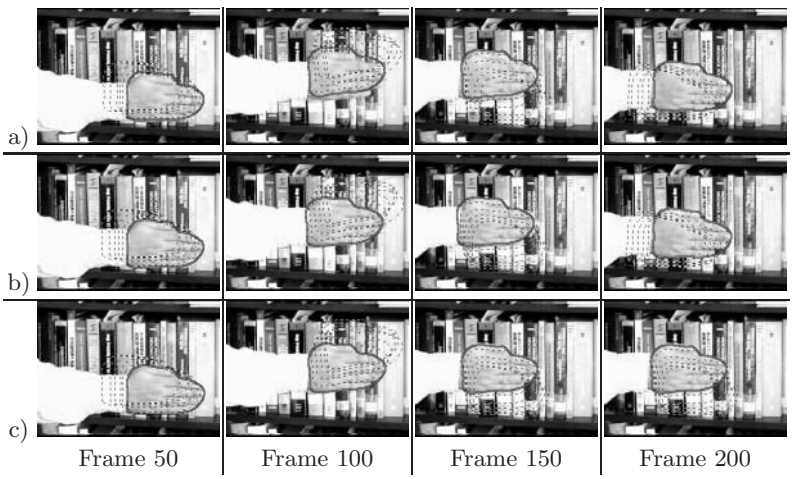

Fig. 2. a) Results obtained with the observation model for the contour normals. b) Results obtained with the observation model based on intensity restrictions. c) Results obtained with the observation model based on similarity measures. The distribution average appears in continuous line in the current frame, and the averages in some previous frames appear in dotted line

presence of clutter constitutes a source of noise for the first models, while favoring the good behavior operation of models based on flow. In addition, the model based on similarity is more stable numerically than the model based on intensity restrictions, because the former doesn't need to compute image derivatives.

Acknowledgment

This work has been financed by grant TIC-2001-3316 from the Spanish Ministry of Science and Technology.

References

[1] E. H. Adelson and J. R. Bergen. The extraction of spatiotemporal energy in human and machine vision. In *Proceedings of IEEE Workshop on Visual Motion*, pages 151–156, Los Alamitos, CA, 1986. IEEE Computer Society Press. 463

[2] A. Blake and M. Isard. *Active Contours*. Springer, 1998. 463, 466

[3] A. Blake, M. Isard, and D. Reynard. Learning to track the visual motion of contours. *Journal of Artificial Intelligence*, 78:101–134, 1995. 466

[4] J. Deutscher, A. Blake, B. North, and B. Bascle. Tracking through singularities and discontinuities by random sampling. In *Proceedings of International Conference on Computer Vision*, volume 2, pages 1144–1149, 1999. 462, 463

[5] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle filtering. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000. 463

[6] A. Gelfand and A. Smith. Sampling-based approaches to computing marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990. 463

[7] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, 1981. 463

[8] M. Isard and A. Blake. Contour tracking by stochastic propagation of conditional density. In *Proceedings of European Conference on Computer Vision*, pages 343–356, Cambridge, UK, 1996. 463

[9] M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *International Journal on Computer Vision*, 1998. 462, 463, 466

[10] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960. 462, 463

[11] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of DARPA IU Workshop*, pages 121–130, 1981. 463, 464

[12] B. McCane, K. Novins, D. Crannitch, and B. Galvin. On benchmarking optical flow. *Computer Vision and Image Understanding*, 84:126–143, 2001. 463

[13] W. H. Press, S. A. Teulosky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C, Second Edition*. Cambridge University Press, 1992. 465

[14] J. Sullivan, A. Blake, M. Isard, and J. MacCormick. Bayesian object localisation in images. *International Journal of Computer Vision*, 44(2):111–135, 2001. 462, 463