

Using Optical Flow for Tracking

M. Lucena¹, J.M. Fuertes¹, and N. Perez de la Blanca²

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@ugr.es

Abstract. We present two observation models based on optical flow information to track objects using particle filter algorithms. Although, in principle, the optical flow information enables us to know the displacement of the objects present in a scene, it cannot be used directly to displace a model since flow estimation techniques lack the necessary precision. We will define instead two observation models for using into probabilistic tracking algorithms: the first uses an optical flow estimation computed previously, and the second is based directly on correlation techniques over two consecutive frames.

1 Probabilistic Tracking

The probabilistic models applied to tracking [1,2,3] 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 estimation 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 optical flow based observation models are defined. The first one uses as evidence an existing estimation of the optical flow of the sequence, and the second one is based on correlation techniques.

2 Optical Flow Estimation

The most well-known hypothesis for calculating the optical flow [4] assumes that the intensity structures found in the image, on a local level, remain approximately constant over time, at least during small intervals of time.

There is no algorithm for estimating the optical flow field which is clearly superior to the others. Each may have small advantages over the others in particular situations, but in general it may be said that from a practical point of view all are equivalent [5,6]. In this paper, we have preferred to use the algorithm in [7], for the following reasons:

- It does not impose restrictions on the sequence to be analyzed.
- It provides a dense estimation of the optical flow.
- It is designed to preserve discontinuities in the flow, which is necessary for the observation model proposed in this section to behave appropriately.

3 The Dynamical Model

Other authors have successfully used characteristics such as the gradient [8] or intensity distributions [3] for tracking tasks. The dynamical model of the object will provide an *a priori* distribution on all the possible configurations in 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 [8]

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

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, using Kalman's Filter [1]. However, in general, the distributions involved in the process are normally not Gaussian and multimodal [2]. Sampling methods for modeling this type of distribution [9] have shown themselves to be extremely useful, and *particle filter* algorithms [10, 11,3] based on sets of weighted random samples, enable their propagation to be performed effectively.

4 Observation Models

If there is a good optical flow measurement and the object is perfectly localized, it is possible to *slide* the points of the model in accordance with the flow vectors, thereby obtaining a good estimate of its position for the following frame. Unfortunately, the small errors in the flow will mount up with each frame, so that the model gradually separates from the real object, until it loses it completely.

Nevertheless, it may be supposed that the object to be tracked will move in an environment that has other displacements, and therefore it may be assumed that there will be discontinuities in the optical flow on its contour –or at least part of it. The observation model will be defined in such a way that it not only helps the flow inside the object to concur with the displacement implied by the value of \mathcal{X} , but also so that discontinuities in the optical flow appear in the contour of the model.

4.1 Observation Model Based on Optical Flow

Let us suppose that we have an estimation of the flow field \mathbf{v} for the image I in the instant t_k . The following error function may be defined, with $S \subseteq I$ being an area inside the image:

$$Z_S(\mathbf{v}; \mathbf{d}) = \sum_{(x,y) \in S} W(x,y) \|\mathbf{v}(x,y) - \mathbf{d}(x,y)\|^2 \quad (1)$$

where $W(x,y)$ is a weight function and $\mathbf{d}(x,y)$ is given by the state vector \mathcal{X} , relating the point on the model in the instant t_{k-1} with the same point in the instant t_k . This measurement will always be non-negative and will only be equal to zero when the flow vectors are perfectly adjusted to the displacement predicted by the model.

Let us now consider a point $\mathbf{x} = (x,y)$ of the image belonging to the outline of the model in the instant t_k . This point will be given by the expression

$$\mathbf{x} = f(\mathbf{X}_{t_k}; \mathbf{m})$$

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. The displacement vector can be calculated for the same point on the model between two consecutive instants of time as

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

Considering S as a 2D region centered at $f(\mathbf{X}_{t_{k-1}}; \mathbf{m})$, the measurement (1) would be:

$$Z_S(\mathcal{X}_{t_k}, \mathbf{m}) = \sum_{(x,y) \in S} W(x,y) \|\mathbf{v}(x,y) - \mathbf{d}(\mathcal{X}_{t_k}, \mathbf{m})\|^2 \quad (3)$$

The flow field is expected to present discontinuities on the boundaries of the moving objects –otherwise, it would be impossible to locate the object only from the flow vectors–, which is why if we subdivide S into two areas S_i and S_e , corresponding respectively to the parts of S interior and exterior of the object contour. If the model’s prediction is good enough, the adjustment must be much better in S_i than in S_e , so that the point in question may be considered to

be placed on the contour. In order to detect that, we compute $Z_{S_i}(\mathcal{X}_{t_k}, \mathbf{m})$, $Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m})$, and the following measurement:

$$Z(\mathcal{X}_{t_k}, \mathbf{m}) = \frac{Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m})}{Z_{S_e}(\mathcal{X}_{t_k}, \mathbf{m}) + Z_{S_i}(\mathcal{X}_{t_k}, \mathbf{m})} \quad (4)$$

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 $Z(\mathcal{X}_{t_k}, \mathbf{m}) \rightarrow 1$, which indicates that the adjustment is much better in S_i than 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 $Z(\mathcal{X}_{t_k}, \mathbf{m}) \rightarrow 0$. 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 –nor, therefore, on the contour of the object. In this case, $Z(\mathcal{X}_{t_k}, \mathbf{m}) = 1/2$.

Assuming that the values of Z_{S_e} and Z_{S_i} are bounded, it may be assumed that the probability of a point on the image corresponding to the point on the outline of the model given by the vectors \mathcal{X}_{t_k} and \mathbf{m} is proportional to $Z(\mathcal{X}_{t_k}, \mathbf{m})$:

$$p(\mathbf{Z}|\mathcal{X}_{t_k}, \mathbf{m}_i) \propto Z(\mathcal{X}_{t_k}, \mathbf{m}_i) \quad (5)$$

Finally, assuming statistical independence, we may obtain the expression for the observation model based on optical flow vectors, as the product of the values obtained for each individual point on the contour:

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

with \mathbf{m}_i being the vector which identifies the i -nth point on the contour of the model.

When it comes to partitioning the neighborhood S corresponding to a point \mathbf{x} of the contour of the model into two halves, one (S_i) inside and the other (S_e) outside the model, respectively, a good approximation consists in using the tangent to the contour in \mathbf{x} as the dividing line between S_i and S_e .

The difficulty in determining a dense flow [12] has led us to establish that those points with a more reliable flow measurement are of more use when calculating the internal and external values for the measurement Z –expression (1). This is easily achieved by calculating the value of $W(x, y)$ in this expression from any of these *reliability* measurements of the calculated flow. In the experiments carried out for this paper, the magnitude of the intensity gradient has been used,

$$W(x, y) = \|\nabla I(x, y)\|$$

4.2 Observation Model Based on Similarity Measures

In the model defined in this section, in order to estimate the observation probability of each point of the contour, similarity measurements shall be used. 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.

Let $\mathbf{x} = f(\mathbf{X}_{t_k}; \mathbf{m})$ be a point belonging to the model contour 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:

$$\begin{aligned} Z_{S_i}(\mathbf{X}_{t_k}, \mathbf{m}) &= \sum_{S_i} W(\mathbf{x}) \left(I^{(k-1)}(\mathbf{x}) - I^{(k)}(\mathbf{x} - \mathbf{d}(\mathcal{X}_{t_k}, \mathbf{m})) \right)^2 \\ Z_{S_e}(\mathbf{X}_{t_k}, \mathbf{m}) &= \sum_{S_e} W(\mathbf{x}) \left(I^{(k-1)}(\mathbf{x}) - I^{(k)}(\mathbf{x} - \mathbf{d}(\mathcal{X}_{t_k}, \mathbf{m})) \right)^2 \end{aligned} \quad (7)$$

where $W(\mathbf{x})$ is a weighting function. Two non negative magnitudes are obtained, that may be combined using expression (4), 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:

$$p(\mathbf{Z}|\mathcal{X}_{t_k}, \mathbf{m}_i) \propto Z(\mathcal{X}_{t_k}, \mathbf{m}_i) \quad (8)$$

Supposing the measurements on each point are statistically independent, we can use the expression (6) to compute the final observation probability.

5 Experiments

The observation models proposed have been incorporated into the CONDENSATION algorithm [8], and its performance has been compared with that of the observation model based on normals as proposed in [8]. Two image sequences are 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 an uniform and non uniform background.

5.1 Tracking an Object over an Uniform Background

For modelling the hand, a contour model based on a closed spline with 10 control points and a Euclidean similarity deformation space were used.

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

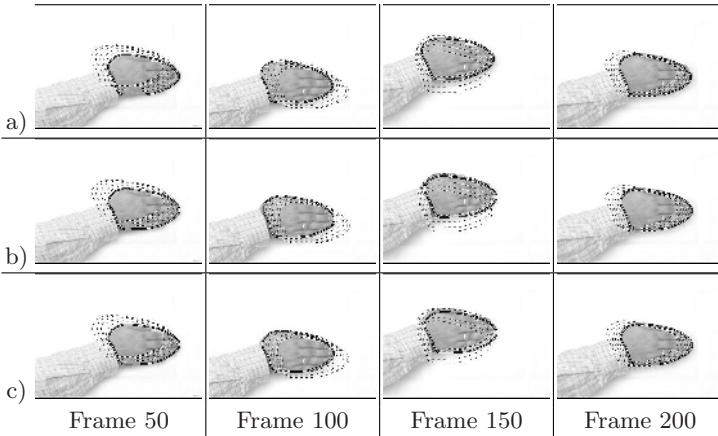


Fig. 1. a) Results obtained with the observation model for the contour normals. b) Results obtained with the observation model based on optical flow. c) Results obtained with the observation model based on similarity measures. The distribution average appears in solid line in the current frame, and the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

of the object in the first frame. Figure 1.a shows the weighted average of the distribution obtained.

For our first observation model, the algorithm of [7] was used on the images to obtain an optical flow map between each two consecutive frames. The size of the area centered on each point was 5×5 pixels. As a reliability measure $W(x, y)$ when it comes to weighting the quadratic differences in expression (3), the magnitude of the intensity gradient ∇I was used on each point.

The CONDENSATION algorithm was applied in exactly the same conditions as for the previous model, obtaining the results showed 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 contour, 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 Non Uniform Background

In order to use the observation model based on contour normals, 18 normals were sketched to each contour, and the same technique was used to detect the boundaries as the one used in the previous series of experiments, with a slightly lower threshold (0.04). The number of samples is still 200, and the parameters for the observation model in this case were $\sigma = 3$ and $\alpha = 0.055$. The results can be seen in Figure 2.a.

For the observation models based on optical flow, the algorithm in [7] was used once again, areas of 5×5 pixels, and $W(x, y) = \|\nabla I\|$. With the same

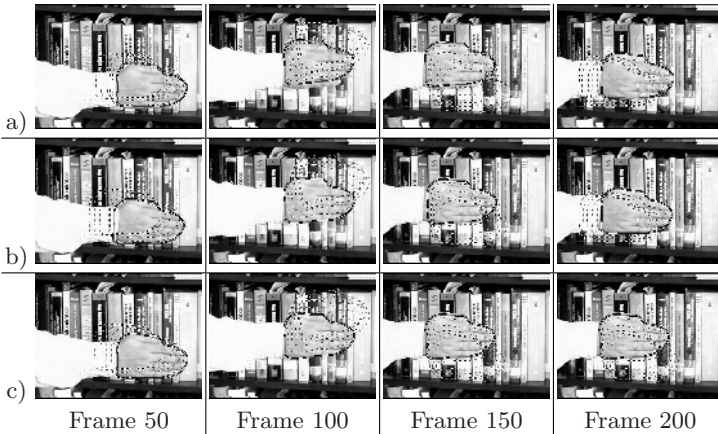


Fig. 2. a) Results obtained with the observation model for the contour normals. b) Results obtained with the observation model based on optical flow. c) Results obtained with the observation model based on similarity measures. The distribution average appears in solid line in the current frame, and the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

number of samples as in the previous experiment (200), and the same 18 points on the contour, the results (Figure 2.b) are clearly better for this observation model.

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 experimental results obtained by the two proposed observation models on the sequence with an uniform background are satisfactory, although at one moment the distribution average strays slightly below and to the right of the hand, covering its shadow. This is due to the fact that, since there is no texture on the background, the shadow appears as a small grey patch which moves around with the hand, which is why the flow boundary can be placed on the contour of the hand-shadow set.

In the second sequence, there were significant differences in the results obtained in the tracking according to which observation model was used. With the observation model for the contour normals, as there are many edges on the background, samples emerge with a significant likelihood value, although they are not placed on the object. Consequently, the distribution average strays from the real position of the object in some frames, although at no time does it lose it completely.

With the two new observation models, it can be seen how the model never loses the object, and that it is not affected by the presence of clutter outside the real object, since the only discontinuities in the flow map will be given by the contour of the hand.

As can be seen, in a non uniform background, the observation models proposed here perform better than the model based on contour normals. In a uniform background, the absence of texture means that the model based on normals behaves better. This suggests that the proposed model and the contour normals model can be considered, in some way, complementary.

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

References

1. R.E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
2. 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*, 1999, vol. 2, pp. 1144–1149.
3. J. Sullivan, A. Blake, M. Isard, and J. MacCormick, "Bayesian object localisation in images," *International Journal of Computer Vision*, vol. 44, no. 2, pp. 111–135, 2001.
4. Berthold K. P. Horn and Brian G. Schunck, "Determining optical flow.," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.
5. 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*, Los Alamitos, CA, 1986, pp. 151–156, IEEE Computer Society Press.
6. B. McCane, K. Novins, D. Crannitch, and B. Galvin, "On benchmarking optical flow," *Computer Vision and Image Understanding*, vol. 84, pp. 126–143, 2001.
7. M. Proesmans, L. Van Gool, E. Pauwels, and A. Oosterlinck, "Determination of optical flow and its discontinuities using non-linear diffusion," in *Proceedings of 3rd European Conference on Computer Vision*, 1994, vol. 2, pp. 295–304.
8. A. Blake and M. Isard, *Active Contours*, Springer, 1998.
9. A. Gelfand and A. Smith, "Sampling-based approaches to computing marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.
10. M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *Proceedings of European Conference on Computer Vision*, Cambridge, UK, 1996, pp. 343–356.
11. 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.
12. J. L. Barron, David J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques.," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.