Moving Objects Segmentation Using Optical Flow Estimation

F. Ranchin, F. Dibos*
CEREMADE (URA CNRS 749), Université Paris 9 – Dauphine,
Place du Maréchal De Lattre De Tassigny,
75775 Paris cedex 16, France
4th December 2003

Abstract

In this paper, we present a new method for the segmentation of moving objects. We use one of the most powerful variational method for computing the optical flow and we exploit this information in the segmentation. This segmentation lies on well-known techniques of active contours.

Since we can distinguish moving objects from static elements of a scene by analyzing norm of the optical flow vectors, this one is incorporated in a region-based active contour model in order to attract the evolving contour to moving objects contours.

We also take gray level into account since it is known that optical flow information does not give the exact contours of the objects.

The results provided are promising and extend the numerous works on this subject.

Keywords: Segmentation, edge and feature detection.

1 Introduction

Segmentation of moving objects from a video sequence is an important stake whose applications cover domains such like video compression, video surveillance or object recognition. In video compression, the MPEG-4 video coding standard is based on the representation of the scene as different shapes-objects. This representation simplify the scene and is used for the encoding of the sequence.

There are different ways to perform moving objects segmentation, using different mathematical techniques. For Markov Random Fields based methods, we refer to the works of Bouthemy ([BL], [PHB]) and for maximum likelihood based methods, to the works of Deriche and Paragios ([DP]). For variational techniques, we refer to the works of Deriche et al. ([ADK]) and Barlaud et al. ([ABJB]). At last, mathematical morphology has been more and more used these last ten years, see the works

*{fd.ranchin}@ceremade.dauphine.fr
of Salembier, Serra and their teams ([BCJM]).

Our aim is to discriminate moving objects from a static background or from a background having a parametric motion (euclidean, affine or projective). So we need to compute the velocity of each point in the image. The most employed tool for computing the apparent motion from a video sequence is the optical flow. Computation of the optical flow can be achieved by many different methods, among a large literature, we can cite the pioneer work of Horn and Schunck [HS], the works of Aubert, Deriche and Kornprobst [ADK] and Weickert and Schnörr [WS]; there are also tensor-based methods such like in [LK] or [BGW]. In this work, we compute the optical flow by one of the most popular technique, then we use the norm of the optical flow as an input in an active contour model.

2 Weickert and Schnörr optical flow computation

Variational methods are based on the assumption of spatial smoothness of the optical flow. That means regularization terms in the energy depend on the spatial gradient of the optical flow. The idea of Weickert and Schnörr is to replace the spatial gradient \( \nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)^T \) of the optical flow by the spatio-temporal gradient \( \nabla^\theta = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial t} \right)^T \), leading to a spatio-temporal smoothness constraint. The energy to be minimized is

\[
E(\mathbf{v}) = \int_{\Omega \times [0,T]} |\nabla I \cdot \mathbf{v} + I_t|^2 \, dx \, dy \, dt \\
+ \alpha \int_{\Omega \times [0,T]} \Psi(|\nabla^\theta v_1|^2 + |\nabla^\theta v_2|^2) \, dx \, dy \, dt
\]

where \( \mathbf{v} = (v_1, v_2) \) is the optical flow and \( \Psi \) is one of the function classically used to smooth the data with keeping its discontinuities, for example \( \Psi(s^2) = \epsilon s^2 + (1 - \epsilon) \lambda^2 \sqrt{1 + s^2 \frac{\epsilon}{\lambda^2}} \). The parameter \( \epsilon \) garantees the well-posedness of the minimization problem. The Euler-Lagrange equations for the minimization problem give

\[
0 = -I_x(I_x v_1 + I_y v_2 + I_t) + \frac{1}{\alpha} \text{div} \left( \psi'(|\nabla^\theta v_1|^2 + |\nabla^\theta v_2|^2) \nabla^\theta v_1 \right) \\
0 = -I_y(I_x v_1 + I_y v_2 + I_t) + \frac{1}{\alpha} \text{div} \left( \psi'(|\nabla^\theta v_1|^2 + |\nabla^\theta v_2|^2) \nabla^\theta v_2 \right).
\]

The term \( \text{div} \left( \psi'(|\nabla^\theta v_1|^2 + |\nabla^\theta v_2|^2) \nabla^\theta v_1 \right) \) is an analog of \( \text{div} \left( \psi'(|\nabla u|^2) \nabla u \right) \) which is used in scalar anisotropic diffusion PDEs. Let us recall in this case, we can develop this term to obtain

\[
\text{div} \left( \psi'(|\nabla u|^2) \nabla u \right) = \psi''(|\nabla u|) \frac{\nabla^2 u(\nabla u, \nabla u)}{|\nabla u|} + \psi'(|\nabla u|) \Delta u \\
= |\nabla u| \psi''(|\nabla u|) u_{\eta \eta} + \psi'(|\nabla u|) \Delta u
\]

where \( \eta = \frac{\nabla u}{|\nabla u|} \). When \( |\nabla u| \approx 0 \) (homogeneous zone),

\[
\text{div} \left( \psi'(|\nabla u|^2) \nabla u \right) \approx \psi'(0) \Delta u = \psi'(0)(u_{\eta \eta} + u_{\xi \xi}),
\]
the smoothing is isotropic, since it diffuses along the two directions $\eta$ and $\xi = \eta^{-1}$. When $|\nabla u| \to \infty$, under suitable conditions on function $\psi$ (see for example [ADK]), the smoothing in the direction $\eta$ (along the gradient) is almost neglectful in comparison to the one along the gradient orthogonal.

Experimentally, the assumption of spatio-temporal smoothness reduces significantly the amount of noise in the flow in comparison with spatial smoothness-based optical flow determination.

In [KWSR], Weickert and his collaborators use a tensor-based optical flow determination which was introduced by Bigun et al. in [BGW]. This method relies on the fact that the optical flow is constant on a neighborhood of each pixel. The optical flow constraint

$$I_x u + I_y v + I_t = 0$$

can be rewritten as $<\nabla^\theta I, (u \ v \ 1)^T> = 0$. Then they minimize the energy

$$\frac{1}{2} \int_{V(x_0,y_0,t_0)} |<\nabla^\theta I, (u \ v \ 1)^T>|^2 \, dx \, dy \, dt,$$

which is quite straightforward since the optical flow is constant over $V(x_0,y_0,t_0)$; the solution is given by the eigenvector corresponding to the minimum eigenvalue of the structure tensor

$$S(V(x_0,y_0,t_0)) = \int_{V(x_0,y_0,t_0)} \nabla^\theta I(\nabla^\theta I)^T \, dx \, dy \, dt$$

However this method fails if the moving objects have a too little size in comparison to the size of the neighborhood chosen. Increasing the size of the neighborhood reduces noise but it imposes a sufficiently large size of the moving objects. In the example shown below, some cars are too small for the size of neighborhood $|\mathcal{V}| = 7^3$ pixels ($7^2$ pixels on 7 successive frames) which seems to be the best choice for noise reducing, and the quality of the optical flow is not too good. This disadvantage may be avoided by performing anisotropic filtering of the tensor $\nabla^\theta I(\nabla^\theta I)^T$, but in [WB], Weickert and his collaborators have made a comparison between many optical flow algorithms, including tensor-based methods relying on a tensor anisotropic presmoothing. The conclusion is that the Weickert and Schnörr algorithm gives a better result for the average angular error than any other.

As we use the optical flow as an input of our segmentation model, the result will naturally depend on the quality of the optical flow. To overcome well-known difficulties in optical flow determination, we have embedded the Weickert and Schnörr algorithm with a multiresolution procedure. We start from a coarse (downsampled) version $I^J$ of the image $I$ at the resolution $J$ (the finest resolution is taken as 0). Then we solve

$$\inf \{ \int_{\Omega^J \times [0,T]} |\nabla I^J \cdot \mathbf{v}^J + I^J_t|^2 \, dx \, dy \, dt + \alpha \int_{\Omega^J \times [0,T]} \Psi(|\nabla^\theta v_1|^2 + |\nabla^\theta v_2|^2) \, dx \, dy \, dt \},$$

For the resolution $j < J$, we use the optical flow determined at the resolution $j + 1$ by searching the residue $w^j = v^j - T v^{j+1}$ ($T$ is an oversampling operator which
constructs the image of $v^{j+1}$ on the grid at the resolution $j$)

\[
\inf_w \left\{ \int_{\Omega \times [0,T]} |\nabla \tilde{I}^j : w^j + \tilde{I}_t^j|^2 \, dx \, dy \, dt + \alpha \int_{\Omega \times [0,T]} \Psi(|\nabla^\theta (Tv_1^{j+1} + w_1^j)|^2 + |\nabla^\theta (Tv_2^{j+1} + w_2^j)|^2) \, dx \, dy \, dt \right\}
\]

where $\tilde{I}^j(x, t) = I^j(x - Tv^{j+1}, t)$ is a shifted version of the image $I^j$ and $\tilde{I}_t^j(x, t) = I^j(x, t + 1) - I^j(x - Tv^{j+1}, t)$. Then we iterate this procedure from $j = J$ to $j = 0$. Introduced by Mémé and Pérez in [MP], this procedure allows to overcome the locality problem of the optical flow, that is to say one cannot determine it correctly when displacements are larger than the size of the mask used for the approximation of the gradient. In our case, it is also a good way to avoid too weak optical flow values in very homogeneous zones where $|\nabla I| \approx 0$ in choosing a suitable value for the regularization parameter $\alpha$. Thus the moving objects are quite well filled even they contain homogeneous zones.

3 Shape sensitivity analysis

For the derivation of energies depending on a shape, we need the notion of shape derivatives. This was introduced by Sokolowski and Zolésio in [SZ]. Following [SZ], we consider a transformation $T_t(V) : x \to X(t)$, $X(t)$ being given as the solution of the ordinary differential equation ($V$ is a smooth vector field with compact support)

\[
\begin{cases}
  X'(t) &= V(X(t)) \\
  X(0) &= x.
\end{cases}
\]
Then, the images of the shape $\Omega$ and its boundary $\Gamma$ by the transformation $T_t(V)$ are simply defined as

$$
\Gamma_t = \{ T_t(V)(x), x \in \Gamma \} = T_t(V)(\Gamma) \\
\Omega_t = \{ T_t(V)(x), x \in \Omega \} = T_t(V)(\Omega).
$$

From these preliminaries, the Eulerian shape derivative of a function $J(\Omega)$ (resp. $J(\Gamma)$) is defined as

$$
dJ(\Omega; V) = \lim_{t \to 0} \frac{1}{t} \left( J(\Omega_t) - J(\Omega) \right) \quad (\text{resp. } dJ(\Gamma; V) = \lim_{t \to 0} \frac{1}{t} \left( J(\Gamma_t) - J(\Gamma) \right)) \quad (2)
$$

We need also to introduce two other notions, the material derivative and the shape derivative

**Definition 1** Let $\psi$ be a function, we denote

$$
\psi_t = \psi(\Omega_t, X(t)) \quad (\text{resp. } \psi_t = \psi(\Gamma_t, X(t)))
$$

and $\psi^0 = \psi(\Gamma, x)$, where $\Gamma_t = T_t(V)(\Gamma)$, $\Omega_t = T_t(V)(\Omega)$, $T_t(V) : x \to X(t)$ and $X(t)$ is a solution of (1). The material derivative is defined as the limit

$$
\dot{\psi}(\Gamma; V) = \lim_{t \to 0} \frac{1}{t} \left( \psi_t - \psi^0 \right) \quad (3)
$$

Since $\psi_t$ depends on the flows of the point and the shape, its derivative can be seen as a total derivative. The derivative of a function $\psi$ which does not depend on a shape can be simply expressed by $\langle \nabla \psi, V \rangle$. As a consequence, it seems natural to define the shape derivative as the difference of the material derivative and the derivative with respect to the point

**Definition 2** Given the material derivative $\dot{\phi}(\Omega; V)$ of the functional $\phi(\Omega)$ (resp. $\dot{\phi}(\Gamma; V)$ for $\phi(\Gamma)$), the shape derivative $\phi'(\Omega; V)$ (resp. $\phi'(\Gamma; V)$) is defined as

$$
\phi'(\Omega; V) = \dot{\phi}(\Omega; V) - \langle \nabla \phi, V \rangle \quad (4)
$$

From Sokolowski and Zolésio works, we will use the following theorem

**Theorem 1** Let $\phi$ given such that $\dot{\phi}(\Omega; V)$ and the shape derivative $\phi'(\Omega; V)$ exist. Then, the cost functional

$$
J(\Omega) = \int_{\Omega} \phi(x, \Omega) \, dx
$$

is shape differentiable and we have

$$
dJ(\Omega; V) = \int_{\Omega} \phi'(\Omega; V) \, dx + \int_{\Gamma} \phi < V, n > \, d\Gamma. \quad (6)
$$

Under similar hypothesis, for a cost functional

$$
J(\Gamma) = \int_{\Gamma} \psi(x, \Gamma) \, d\Gamma,
$$

$$
J(\Gamma) = \int_{\Gamma} \psi(x, \Gamma) \, d\Gamma,
$$
the shape derivative exists and is given by

$$dJ(\Gamma; V) = \int_{\Gamma} \psi'(\Gamma; V) d\Gamma + \int_{\Gamma} \kappa \psi < V, n > d\mathbf{x}. \quad (8)$$

If \(\psi\) is the restriction on \(\Gamma\) of a function defined on \(\Omega\), the shape derivative can be rewritten as

$$dJ(\Gamma; V) = \int_{\Gamma} \phi'(\Gamma; V)|_{\Gamma} d\Gamma + \int_{\Gamma} \left( \frac{\partial \phi}{\partial n} + \kappa \psi \right) < V, n > d\mathbf{x} \quad (9)$$

where \(\kappa\) denotes the curvature of \(\Gamma\) and \(n\) its outward normal.

4 A region-based active contour model

Active contour models have been introduced first by Kass, Witkin and Terzopoulos in [KWT]. A parametrized curve is searched as the solution of the minimization problem

$$J_{KWT}(\Gamma) = \int_{0}^{L(\Gamma)} -|\nabla I(\Gamma(s))|^2 ds + C \int_{0}^{L(\Gamma)} (a + |\kappa(s)|^2) ds.$$

The first term helps to attract the curve towards high gradient values of the image, i.e. the edges, while the last integral guarantees the smoothness of the curve (it has a finite perimeter and a finite total curvature). Unfortunately, this model does not allow the curve break since the curve is parametrized and the smoothness is controlled by the last integral.

In [CCCD], Caselles, Catté, Coll and Dibos have introduced a \textit{snake} model that allows topology changes. The curve is represented as the zero level set of a function \(u\) that is evolved by the PDE

$$\frac{\partial u}{\partial t} = g(|\nabla G_{\sigma} I|) |\nabla u| \left( \text{div} \left( \frac{\nabla u}{|\nabla u|} \right) + \nu \right) \quad (10)$$

where \(g(r) = \frac{1}{1+r^2}\), \(\nu\) is a parameter chosen such that non convex contours may be obtained. This is a level set model inspired from the level sets methods introduced by Osher and Sethian [OS] and so allows topology changes, that is to say that the curve can split into several Jordan curves during its evolution.

Contour-based segmentation is a difficult task if we use the optical flow as an input of our video segmentation. In [KWSR], Weickert \textit{et al.} use geodesic active contours, but the function \(g\) used is the characteristic function based on the coherence function \(c_t = \exp(-\frac{1}{|\lambda_1 - \lambda_3|})\) (\(\lambda_1\) and \(\lambda_3\) are the extremal eigenvalues of the structure tensor \(\int \nabla^2 I(\nabla^2 I)^T d\mathbf{x}\)) and the optical flow norm \(|\mathbf{v}|\)

$$g = 1(c_t(x, y) > 1 - \epsilon \text{ and } |\mathbf{v}| > \alpha)$$

where \(\epsilon\) and \(\alpha\) are two given parameters. Such a function is more related to a region information, as a consequence, the idea to incorporate region information from the inside and outside regions seems more natural than using this function as a boundary potential. What we want to do is to have the optical flow norm
upper a certain threshold on the inside region and lower on the outside region. In addition, we would like the boundary be an edge of the image $I$. The most natural way to realize this is to minimize the function $g(|\nabla G_{\alpha}I|)$ along the boundary with $g(r) = \frac{1}{1+r^2}$. For notation simplicity, we will denote $g(|\nabla G_{\alpha}I(x)|)$ by $g(x)$. Finally, we minimize the functional

$$E(\Omega) = \int_{\Omega} \alpha \, dx + \int_{D\setminus \Omega} |v| \, dx + \int_{\partial \Omega} g(x(s)) \, ds.$$ 

Following the theorems on the derivation of region functionals and boundary functionals, the shape derivative of this energy is given by

$$dE(\Omega; V) = \int_{\Gamma} (\alpha - |v| + g \kappa + <\nabla g, n>) <V, n> \, ds.$$ 

The minimization of this energy can be achieved by a gradient descent method on the contour by choosing the direction $V$ as $- (g(x(s))\kappa + <\nabla g, n>) + \alpha - |v|) \, n$

$$\frac{\partial \Gamma}{\partial t} = - (g(x(s))\kappa + <\nabla g, n>) + \alpha - |v|) \, n.$$

With the level sets formalism and with the convention $u > 0$ inside the curve $\Gamma$ and $u < 0$ outside, we have $n = -\frac{\nabla u}{|\nabla u|}$ and

$$\frac{\partial u}{\partial t} = |\nabla u| \left( \text{div} \left( g(x) \frac{\nabla u}{|\nabla u|} \right) + |v| - \alpha \right). \quad (11)$$

5 Experimental results

We show the results obtained for this method on figures 1, 3, 4 and 7. The difference between figures 3 and 4 is only the using of a multiresolution procedure for the Weickert and Schnörr optical flow computation, we can see that it gives slightly better results on homogeneous zones not very well segmented, for example the dark car on the left side at the bottom of the images. The results presented on figure 7 are obtained from the Weickert and Schnörr algorithm without multiresolution procedure, since there are less problems for this sequence.

For the initialization, we have only used it on frames on which moving objects appear, since it requires more iterations to the algorithm to converge toward the objects; on the other frames, we use the final state of the snake on the previous frame as the initialization to be close to the solution and to reduce the iteration number. On a computer with a 1.8 GHz Pentium processor and 1 Go of RAM, the computing time of our snake algorithm is approximately of 4 minutes for 1000 iterations and a time step of 5 on frames where moving objects appear and 20 seconds on the other frames with 100 iterations.

We show on figure 5 some results of our implementation of the optical flow tensor-based active contour model of [KWSR]. We notice that some cars are missed by the algorithm, we think it is due to the size of the objects which are not big enough in comparison to the size of the neighborhood, as we have explained it in section 2.

---

1 All the images are freely provided on the web site http://i21www.ira.uka.de/image_sequences/
One probably can obtain very better results with a presmoothing of the structure tensor like in [WB].

Another direction we have involved is a non variational active contour model inspired from the Caselles, Catté, Coll and Dibos snake model

\[
\frac{\partial u}{\partial t} = |\nabla u| \left( g(x) \text{div} \left( \frac{\nabla u}{|\nabla u|} \right) + |v| - \alpha \right).
\]  

(12)

The results obtained are not very different, we simply notice that contours are slightly regularized in comparison those obtained from the upper variational method. The results are presented on figure 6. The computing time is almost the same since we employ Romeny-Viergerer-Weickert AOS schemes ([RVW]) for the two models.

In addition to this algorithm, we add a post-processing consisting in pushing the curves toward the true location of moving objects edges, this can be simply done by applying the geodesic criterion

\[
\inf \left\{ \int g(x(s)) \, ds \right\}
\]

with the same function \( g \) as before. Actually, we think this step could be avoided by choosing a more precise value for the parameter \( \alpha \) in our algorithm, but it would also be more sensible to noise and to the problem of optical flow determination in gray level homogeneous zones. The experiments have shown that choosing a value for \( \alpha \) such that the curves roughly surround the moving objects and then apply the post-processing is a better way to segment correctly the moving objects.

At last, we must precise this method applies well on rigid or approximately rigid body movements since we search areas of sufficiently large optical flow norm. For objects which have a very inhomogeneous and non rigid movements, this is a hard task to apply our algorithm and one should probably build a new model or incorporate an \textit{a priori} information on the shape to converge to the border of the whole object and not only to the parts of sufficiently high displacements. Another limitation is the problem of too gray level homogeneous zones where the optical flow can be too weak, despite the use of multiresolution procedures in optical flow it remains a difficult problem.

References


Figure 2: Initialization on the first image.


[CCCD] V. Caselles, F. Catté, B. Coll and F. Dibos, A geometric model for active contours in image processing, Numerische Mathematik 66 (1993), 1-31;


[DP] R. Deriche and N. Paragios, Geodesic Active Regions for Motion Estimation and Tracking, INRIA Research Report n° 3631 (1999);


Figure 3: Segmented objects for our method on frames 5, 10, 15, 20, 25, 30.
Figure 4: Multiresolution version of the Weickert and Schnörr algorithm: segmented objects for our method on frames 5, 10, 15, 20, 25, 30.
Figure 5: Segmented objects for the tensor-based method on frames 1, 5, 10.
**Figure 6:** Segmented objects for our non variational method on frames 5, 10, 15, 20, 25, 30.
Figure 7: Segmented objects for our variational method on frames 450, 455, 460, 465 and 470 (the initialization is chosen close to the border as before).


