Geodesic Active Regions for Motion Estimation and Tracking*

Nikos Paragios Rachid Deriche
I.N.R.I.A
BP. 93, 2004 Route des Lucioles
06902 Sophia Antipolis Cedex, France
e-mail: {nparagio,der}@sophia.inria.fr

Abstract

This paper proposes a new front propagation method to deal accurately with the challenging problem of tracking non-rigid moving objects. This is obtained by employing a **Geodesic Active Region** *model where the designed objective function is composed of boundary and region-based terms and optimizes the curve position with respect to motion and intensity properties. The main novelty of our approach is that we deal with the motion estimation (linear models are assumed) and the tracking problem simultaneously. In other words, the optimization problem contains a coupled set of unknown variables; the curve position and the corresponding motion model. The designed objective function is minimized using a gradient descent method; the curve is propagated towards the object boundaries under the influence of boundary, intensity and motion-based forces using a PDE, while given the curve position an incremental analytical solution is obtained for the motion model. Besides, this PDE is implemented using a level set approach where topological changes are naturally handled. Very promising experimental results are provided using real video sequences.*

1 Introduction

The tracking problem has a wide variety of applications in computer vision and motion analysis and provides a good basis for high level tasks of computer vision like 3-D reconstruction and representation.

This paper addresses the problem of simultaneously tracking several non-rigid objects and estimating their motion parameters using a coupled front propagation model that integrates boundary and region-based information.

During the last years, a great variety of tracking algorithms have been proposed. These may be classified in two distinct categories:

1. **Motion-based** approaches rely on a robust method for grouping visual motion consistencies over time [14].
2. **Model-based** approaches impose high-level semantic representation and knowledge [12] but they suffer from being computationally expensive.

In both cases the tracking is performed using measurements provided by geometrical or region-based properties of the tracked object. In this direction there are two main approaches: the **boundary-based** (they are usually referred as edge-based approaches) rely on the information provided by the object boundaries (shape properties) [4, 8, 11] and are usually implemented using active contour models. On the other hand, the **region-based** approaches rely on information provided by the entire region [14, 15] (texture and motion-based properties). Finally, there are some approaches that combine intensity, boundary and motion-based information [3, 7].

The objective of this work is to deal simultaneously with the motion estimation and the tracking problem under a front propagation framework. Towards this end, we have extended our previous work on geodesic tracking [18, 20] by introducing a visual consistency module to the existing boundary, intensity and motion detection-based region modules. Thus, here we propose a coupled framework for the motion estimation (affine models are assumed) and the tracking problem. This coupling leads to a two direction front propagation tracking system, where the motion parameters and the exact object position are simultaneously recovered.

The observation set of our approach is composed of:

- A background reference frame,
- A sequence of images acquired by a static observer.

The proposed algorithm is depicted in [**fig.** 1]. Initially, the gradient values of the absolute difference frame (between the current and the background) are used to provide the boundary-based information [**fig.** 1: **Boundary Module**, **Section** 3.1].

The next step relies on analyzing the observed difference density function (histogram) as a two mixture component density, that discriminates the static from the mobile pixels in terms of conditional probabilities. These probabilities are used to define the motion detection information [**fig.** 1: **Motion Detection Module**, **Section** 3.2]. Besides, we assume

*This work was funded in part under the VIRGO research network (EC Contract No ERBFMRX-CT96-0049) of the TMR Program.

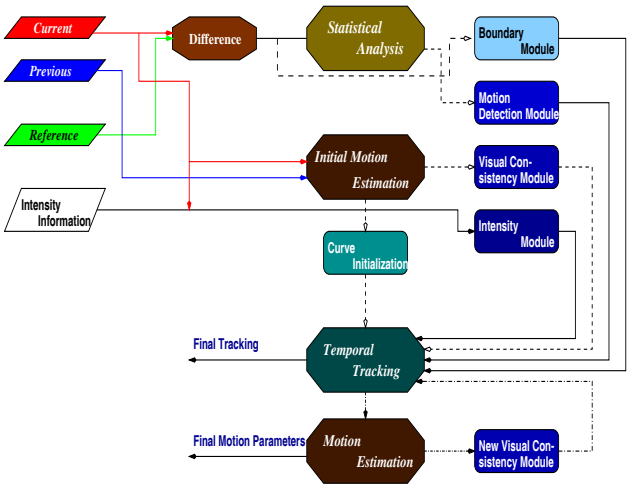


Figure 1. Flow chart for motion estimation and tracking.

that a knowledge about the background as well as the moving intensity properties is available (it is gained during the process) and is used to define the intensity-based information [**fig.** 1: **Intensity Module**, **Section** 3.3]. Finally, given the current curve position, we estimate the motion model that creates a visual consistency between the object (interior curve region) intensities on the current and on the previous frame. This model is used to determine the [**fig.** 1: **Visual Consistency Module**, **Section** 3.4], and aims at minimizing the *sum of squared differences* between the corresponding object regions over the time.

Summarizing, we reformulate the tracking under the **Geodesic Active Region** framework, where boundary and region-based (motion and intensity) modules are involved. The designed objective function is minimized with respect to the curve position as well as to the motion parameters using a gradient descent method. The obtained motion equation (with respect to the curve) is implemented using a level set approach [16], which handles automatically topological change. Besides, an incremental method is used to update the estimation of the motion parameters [2]. Both problems are treated under a coupled framework; The latest curve position is used to update the motion parameters, while the latest estimation of these parameters creates forces that move the curve towards the object position. Finally, the level-set propagation is performed using a very fast front propagation algorithm [18]. Various experimental results demonstrate the performance of our approach.

The remainder of this paper is organized as follows. Section 2 briefly introduces the **Geodesic Active Region Model**, while the tracking problem is considered in Section 3. Besides, the minimization of the objective function appears in Section 4. Finally, experimental results and discussion appear in Section 5.

2 Geodesic Active Regions

The Geodesic Active Contour model has been initially proposed in [5, 9, 13] and successfully applied to a wide variety of computer vision applications. On the other hand, motivated by the work proposed in [6, 22], the Geodesic Active Region model has been initially introduced in [17] for supervised texture segmentation and has been successfully exploited in [20] to provide an early solution to the tracking problem.

We are going to introduce the generalized Geodesic Active Region model for a simple segmentation case with two possible hypothesis (h_A, h_B). In order to facilitate the notation, let us make some definitions:

- Let $\mathbf{I} : \mathcal{R} \rightarrow \mathbf{R}$ be the input frame.
- Let $\mathcal{P}(\mathcal{R}) = \{\mathcal{R}_A, \mathcal{R}_B\}$ be a partition of the frame domain into two non-overlapping regions $\{\mathcal{R}_A \cap \mathcal{R}_B = \emptyset\}$.
- And, let $\{\partial\mathcal{R}\}$ be the boundaries between $\mathcal{R}_A$ and $\mathcal{R}_B$.

Let us assume that for the given frame some information regarding the real region boundaries is available. Let $[p_C(\mathbf{I}(s))]$ be the *boundary probability density function* that measures the probability of a given point being at the boundaries between the two regions.

Besides, let us now assume that *an a priori knowledge about the desired intensity properties of the different regions is available*; the conditional probability density functions $[p_A(\mathbf{I}(s)), p_B(\mathbf{I}(s))]$ with respect to the hypothesis h_A and h_B. Then, the segmentation is equivalent to creating a consistent frame partition [determined by a curve that is attracted by the region boundaries] between *the observed data, the associated hypothesis and their expected properties*. This partition according to the Geodesic Active Region model is given by:

$$E(\partial\mathcal{R}) = (1-\alpha) \underbrace{\int_0^1 g(p_C(\mathbf{I}(\partial\mathcal{R}(p))))|\partial\dot{\mathcal{R}}(p)|dp}_{Boundary\ Term} +$$

$$\alpha \underbrace{\left\{ \iint_{\mathcal{R}_A} \xi\ [p_A(\mathbf{I}(x,y))]\, dxdy + \iint_{\mathcal{R}_B} \xi\, [p_B(\mathbf{I}(x,y))]\, dxdy \right\}}_{Region\ Term}$$

where $\partial\mathcal{R}(p) : [0,1] \rightarrow \mathbf{R}^2$ is a parameterization of the region boundaries in a planar form, $\alpha \in [0,1]$ is a positive constant balancing the contribution of the two terms, and $\{g(), \xi()\}$ are positive monotonically decreasing functions (*e.g.* Gaussians).

Let us now try to interpret the above functional. The boundary term aims at finding a geodesic curve of minimal length is attracted by the desired image properties (region boundaries) [10]. The region term uses this curve to define a partition where the interior curve region corresponds to $\mathcal{R}_A$, while the exterior to $\mathcal{R}_B$ that aims at maximizing the quality of the segmentation map, given the observed data and the associated hypothesis. Thus, if the obtained partition is the optimal one, then given the hypothesis h_A (*resp.* h_B), the conditional probabilities for the pixels of region $\mathcal{R}_A$ (*resp.* $\mathcal{R}_B$) are maximized and the corresponding energy is minimized.

The minimization of this function is performed using a gradient descent method. If $u = (x, y)$ is a point of the curve then, the computation of the Euler-Lagrange equations [22] gives:

$$\frac{du}{dt} = \underbrace{[\alpha\,(\xi(p_A(\mathbf{I}(u))) - \xi(p_B(\mathbf{I}(u))))}_{region-based\ force} + \underbrace{(1-\alpha)\,(g(u)\mathcal{K}(u) - \nabla g(u)\cdot\mathcal{N}(u))]}_{boundary-based\ force}\mathcal{N}(u)$$

where $\mathcal{K}$ is the Euclidean curvature and $\mathcal{N}$ is the unit inward normal to ∂R. The obtained PDE motion equation has two kind of *forces* acting on the curve, both in the direction of the normal. The **region force** tries to shrink or to expand the curve in the direction that maximizes the quality of the segmentation map provided by this partition. Thus, if u is a point of h_B then $[p_B(\mathbf{I}(u)) > p_A(\mathbf{I}(u))]$ and this force acts to shrink the curve $[\xi(p_A(\mathbf{I}(u))) - \xi(p_B(\mathbf{I}(u))) > 0]$ otherwise acts to expand it. Besides, the **boundary force** contains two sub-terms; one that moves the curve towards the region boundaries constrained by the curvature and one that attracts these boundaries (refinement term).

3 Geodesic Active Region Tracking

We reformulate the tracking problem as a frame partition problem, since we would like to use the Geodesic Active Region Model. In order to facilitate the notation, let us make some definitions:

- Let $\mathbf{I}(x, y; t)$ be the current frame, $\mathbf{I}(x, y; t-1)$ the previous frame and $\mathbf{R}(x, y)$ the background reference frame,
- Let $\mathcal{P}(\mathcal{R})$ be partition of the image domain $\mathcal{R}$ into N non-overlapping regions,
- Let $\mathcal{R}_0$ be the static region, and let $\mathcal{R}_i$ be the region that corresponds to object O_i,
- Let $\partial\mathcal{R}_i$ be the boundary of region $\mathcal{R}_i$, and let $\partial\mathcal{P}(\mathcal{R}) = \{\cup_{i=1}^{N}\partial\mathcal{R}_i\}$,
- Finally, let A_i be the first-order linear 2-D motion model that creates a visual consistency between the current and the previous frame for the object O_i.

3.1 Boundary Module

Let $\mathbf{D}(x, y; t)$ the current difference frame: $[\mathbf{D} = \mathbf{I} - \mathbf{R}]$. We use the difference frame to provide some information about the object boundaries that is determined by high gradient values (the absolute values are considered $|D(x, y; t)|$). This information is captured using a Gaussian function (a more sophisticated approach can be found in [18, 20])

$$g(\mathbf{B}(x, y)|\sigma_B) = \frac{1}{\sigma_B\sqrt{2\pi}}e^{-\frac{\mathbf{B}^2(x,y)}{2\sigma_B^2}}$$

where $\mathbf{B}(x, y) \triangleq ||\{\nabla|\mathbf{D}(x, y)|\}||$. Then, we can activate the **boundary-based** tracking module;

$$\mathbf{E_B}(\partial\mathcal{P}(\mathcal{R})) = \sum_{i=1}^{N}\int_0^1 g(\mathbf{B}(\partial\mathcal{R}_i(p_i))|\sigma_B)\left|\partial\dot{\mathcal{R}}_i(p_i)\right| dp_i$$

where $\partial\mathcal{R}_i(p_i) : [0, 1] \rightarrow \mathbf{R^2}$ is a parameterization of the region boundaries $\mathcal{R}_i$ in a planar form. This term aims at finding curves that are attracted by the object boundaries.

3.2 Motion Detection Module

Besides, we assume that the observed difference frame is composed of two populations [21], the *static* that contains the background pixels and the *mobile* one contains the pixels that belong to moving objects. This assumption can be easily projected to a statistical model, where the observed histogram of the difference frame is a mixture of a single-component *static* density and a multi-component *mobile* density,

$$p_D(d) = P_{static}\ p_S(d|\Theta_S)\ +\ P_{mobile}\ p_M(d|\Theta_M)$$

where Θ_S (*resp.* Θ_M) are the unknown parameters of the static (*reps.* mobile) component, and $(P_{static},\ P_{mobile})$ are their *a priori* probabilities. We assume that the components of these probability density functions follow Gaussian or Laplacian law. The estimation of the unknown parameters of this model is done using the maximum likelihood principle (**see** [18, 20] for details).

Then, the regions that are associated to moving objects $\{\mathcal{R}_i : i \in [1, N]\}$ should be composed of mobile pixels, and should provide important conditional probabilities with respect to the *mobile* case. On the other hand, the background region $\{\mathcal{R}_0\}$ should be composed of static points. Taking this into account, the energy expression for the motion detection **region-based** module is defined as:

$$\mathbf{E_D}(\partial\mathcal{P}(\mathcal{R})) = \iint_{\mathcal{R}_0} g\,(p_S(\mathbf{D}(x, y; t))|\sigma_D)\,dxdy\ + \sum_{i=1}^{N}\iint_{\mathcal{R}_i} g\,(p_M(\mathbf{D}(x, y; t))|\sigma_D)\,dxdy$$

For stability reasons, and to preserve the regularity between the boundary and the region-based terms, we use also a Gaussian function $[g(.|\sigma_D)]$ to capture the motion detection properties.

3.3 Intensity Module

Since we assume a background reference frame, we can extract some intensity-based information related to it. This can be done by applying some invariant operators. In our case these operators are selected to be invariant to translation and rotation $[F = \{I, I_{\eta\eta}, I_{\xi\xi}\}]$, where $\{I_{\eta\eta}, I_{\xi\xi}\}$ are the second directional derivatives $\left[\vec{\eta} = \frac{\vec{\nabla}I}{|\nabla I|}, \xi \perp \eta\right]$.

Then, we can model the operator responses using low level statistics, where the conditional probability density functions are expressed directly from their observed histograms. The output of this operation is a vector of conditional probability density functions

- $\mathbf{p_0(x)} = \left(p_{0|x}(x), p_{0|\eta\eta}(x_{\eta\eta}), p_{0|\xi\xi}(x_{\xi\xi})\right)$

Additionally, if we assume that there are N moving objects in our scene and these objects have been tracked, we can produce the same statistical modules with respect to these objects:

- $\mathbf{p}_i(.) = \left(p_{i|x}(.), p_{i|\eta\eta}(.), p_{i|\xi\xi}(.)\right),\ i \in [1, N]$

A moving object is well tracked if the corresponding curve at different time instants refers to a region with constant intensity properties, the object properties. Besides, the exterior curve region corresponds to the static hypothesis and should preserves the background intensity properties. This assumption can be easily projected to the Geodesic Active Region model, by creating a **region-based** intensity module given by,

$$\mathbf{E_I}(\partial\mathcal{P}(\mathcal{R})) = \sum_{i=0}^{N} \sum_{o\in F} w_o \iint_{\mathcal{R}_i} g\left(p_{i|o}(\mathbf{I}(x,y;t))|\sigma_I\right) dxdy$$

where w_o are positive normalized weights that take into account the operator uncertainties and $[g(.|\sigma_I)]$ is a Gaussian function. For our case, since the intensity information is not of good quality (see **section 4.4**), we use equal weights for the different filter operators.

3.4 Visual Consistency Module

The last module, consists of defining a consistency term for the object intensities with respect to their position over the time and is based on the optical flow constraint. If we assume that the 2-D motion of the object is known, then we can predict the position of the object O at the current frame, given its position at the previous frame (or the opposite). In other words, if A is the motion model of object O between the frames t and $t-1$, then for a given point s of O at the frame t, we can find the corresponding point at frame $t-1$ $(s + A(s))$. The inverse motion model is used.

The most common way to validate the motion model is the optical flow constraint. This means that the observed intensities at the position s in the frame t, and at $s + A(s)$ in the frame $t-1$ should be the same if there aren't any global illumination changes. The optical flow constraint relies on minimizing the *sum of squared differences* (SSD)

$$E(A) = \iint_{O} \left[I(x,y;t) - I(x + A_x(x,y), y + A_y(x,y); t-1)\right]^2 dxdy$$

where $A(x,y) = [A_x(x,y), A_y(x,y)]$. Given the motion models for the moving objects, we can build a visual consistency term that demands a point to point correspondence between the object positions over the time. Additionally, we build a similar visual consistency for the background points, where there isn't any motion. This consistency can be easily added as a region-based term to the designed objective function as,

$$E(\partial\mathcal{P}(\mathcal{R}), A) = \iint_{\mathcal{R}_0} \left[\mathbf{I}(x,y;t) - \mathbf{R}(x,y)\right]^2 dxdy + \sum_{i=1}^{N} \iint_{\mathcal{R}_i} \left[\mathbf{I}(x,y;t) - \mathbf{I}(x + A_{ix}(x,y), y + A_{ix}(x,y); t-1)\right]^2 dxdy$$

In order to preserve the regularity with respect to the other tracking modules, we express the visual consistency term using a function $h(.)$ $[h(x|\sigma_C) = 1 - g(x|\sigma_C)]$ that gives the following **region-based** term,

$$\mathbf{E_C}(\partial\mathcal{P}(\mathcal{R}), A) = \iint_{\mathcal{R}_0} h\left(\mathbf{I}(x,y;t) - \mathbf{R}(x,y)|\sigma_C\right) dxdy + \sum_{i=1}^{N} \iint_{\mathcal{R}_i} h\left(\mathbf{I}(x,y;t) - \mathbf{I}(x + A_{ix}(x,y), y + A_{ix}(x,y); t-1)|\sigma_C\right) dxdy$$

The interpretation of this term is following: the first integral is applied to the background region $\{\mathcal{R}_0\}$. If the optimal segmentation map is derived, then the sum of squared differences for the points of region $\{\mathcal{R}_0\}$ between the current frame and the background reference frame is minimum since the intensity values are the same. On the other hand, the background pixels charge the objective function if they are labeled as object pixels. Additionally, if the curve position as well as the motion models are optimized, then, the *sums of squared differences* (SSD), over the objects are minimum.

It is known that for a sufficiently small field of view and independently moving objects, the image velocity field inside patch can be well approximated by a linear transformation. We assume that the object motion can be described using a global affine motion model which is valid for the majority of the object pixels. This model is given by

$$A(x,y) = \begin{bmatrix} A_x(x,y) \\ A_y(x,y) \end{bmatrix} = \begin{bmatrix} \mathbf{a}_0 & \mathbf{a}_1 & \mathbf{a}_2 \\ \mathbf{b}_0 & \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ y \end{bmatrix}$$

3.5 The Complete Model

We fuse the different functionals, by defining a Geodesic Active Region Tracking model:

$$E(\partial\mathcal{P}(\mathcal{R}), A) = \alpha\, \mathbf{E_B}(\partial\mathcal{P}(\mathcal{R})) + \beta\, \mathbf{E_D}(\partial\mathcal{P}(\mathcal{R})) + \gamma\, \mathbf{E_I}(\partial\mathcal{P}(\mathcal{R})) + \delta\, \mathbf{E_C}(\partial\mathcal{P}(\mathcal{R}), A)$$

where $\mathbf{E_B}, \mathbf{E_D}, \mathbf{E_I}, \mathbf{E_C}$ are the different tracking modules, and $\alpha, \beta, \gamma, \delta$ are positive normalized constants that balance their contribution.

4 Minimizing the Energy

The objective function is minimized using a gradient descend method.

4.1 With Respect to Motion Model

In order to derive the defined objective function with respect to the motion parameters, we expand the related function using fist order Taylor series. In that case we can define the intensity error at each pixel as

$$\mathbf{e}(u) = I(u;t) - I(x + A_x(u), y + A_y(u); t-1) = I(u;t) - A_x(u)I_x(u;t-1) - A_y(u)I_y(u;t-1) - I(u;t-1)$$

where $\{u = (x,y)\}$. The minimization of the objective function with respect to the motion parameters A_i depends only from the visual consistency term and is equivalent with minimizing of the following functional

$$E(A_i) = \iint_{\mathcal{R}_i} \left[\mathbf{I}(x,y;t) - \mathbf{I}(x + A_{ix}(x,y), y + A_{ix}(x,y); t-1)\right]^2 dxdy$$

We adopt an incremental way to update the motion parameters estimation [1, 2] that relies on estimating the improvement to the existing affine estimate A_i, $\Delta A_i = \begin{bmatrix} \hat{\mathbf{a}}_{i0} & \hat{\mathbf{a}}_{i1} & \hat{\mathbf{a}}_{i2} \\ \hat{\mathbf{b}}_{i0} & \hat{\mathbf{b}}_{i1} & \hat{\mathbf{b}}_{i2} \end{bmatrix}$ that minimizes a modified objective functional given by,

$$E(A_i, \Delta A_i) = \iint_{\mathcal{R}_i} [\mathbf{I}(x, y; t) - \mathbf{I}(x + [A_{ix} + \Delta A_{ix}](x, y) , y + [A_{ix} + \Delta A_{iy}](x, y); t - 1)]^2 \, dxdy$$

where the unknown parameters consist of the matrix ΔA. Using the first order Taylor expansion we obtain the following functional

$$E(A_i, \Delta A_i) = \iint_{\mathcal{R}_i} \left[\mathbf{I}(x, y; t) - \nabla \mathbf{I}\left([\mathbf{1} + A_i](x, y); t - 1\right) \begin{bmatrix} \Delta A_{ix}(x, y) \\ \Delta A_{iy}(x, y) \end{bmatrix} - \mathbf{I}([\mathbf{1} + A_i](x, y); t - 1) \right]^2 dxdy$$

where $\mathbf{1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is a special identity matrix. Then minimizing the designed function using the Euler-Lagrange equations with respect to the motion parameters ΔA_i, we obtain a 6×6 linear system, with six unknown parameters $\left\{ \hat{\mathbf{a}}_{i0}, \hat{\mathbf{a}}_{i1}, \hat{\mathbf{a}}_{i2}, \hat{\mathbf{b}}_{i0}, \hat{\mathbf{b}}_{i1}, \hat{\mathbf{b}}_{i1} \right\}$ which has a solution in a closed form. We use this solution to update the motion parameters $\left[A_i^{\tau+1} = A_i^{\tau} + \Delta A_i \right]$. We perform this motion estimation step until the motion model is optimized. Besides, for each iteration we use the object points for which the optical flow constraint is valid, (*e.g.* points with intensity error below a threshold).

4.2 With Respect to the Curve Position

Let $u = (x, y)$ be a point of the initial curve that is located between the background region and the object O_r $\{\partial \mathcal{R}_r\}$. We compute the Euler-Lagrange equations (Section 2) and we deform the initial curve towards the minima of the objective function using the following equation:

$$\frac{d}{dt}u = \Bigg(\alpha \underbrace{\left[g(\mathbf{B}(u)|\sigma_B)\, \mathcal{K}(u) - \nabla g(\mathbf{B}(u)|\sigma_B) \cdot \mathcal{N}(u) \right]}_{boundary\ force \triangleq \mathbf{f}_{B1} + \mathbf{f}_{B2}} + \beta \underbrace{\left[g\left(p_M(\mathbf{D}(u))|\sigma_D\right) - g\left(p_S(\mathbf{D}(u))|\sigma_D\right) \right]}_{motion\ detection\ force \triangleq \mathbf{f}_D} + \gamma \underbrace{\sum_{o \in F} w_o \left[g\left(p_{\mathbf{r}o}(\mathbf{I}(u))|\sigma_I\right) - g\left(p_{0o}(\mathbf{I}(u))|\sigma_I\right) \right]}_{intensity\ segmentation\ force \triangleq \mathbf{f}_I} + \delta \underbrace{\left[h\left(\mathbf{I}(u; t) - \mathbf{R}(u)|\sigma_C\right) - h\left(\mathbf{I}(u; t) - \mathbf{I}(u + A_r(u); t - 1)|\sigma_C\right) \right]}_{visual\ consistency\ force \triangleq \mathbf{f}_C} \Bigg) \mathcal{N}(u)$$

where $\mathcal{K}$ is the Euclidean curvature and $\mathcal{N}$ is the unit inward normal to $\partial \mathcal{R}_r$. We will try to interpret the above PDE motion equation that is composed of several "forces" acting on the contour, all in the direction of the normal.

- The first force $\{\mathbf{f}_{B1} + \mathbf{f}_{B2}\}$ is a boundary-based and is composed of two sub-terms;one that shrinks or expands the curve constrained by the curvature effect towards the object boundaries $\{\mathbf{f}_{B1}\}$ and one that attracts the curve to the objects boundaries (refinement term) $\{\mathbf{f}_{B2}\}$.
- The second term is a motion detection force $\{\mathbf{f}_D\}$ that aims at shrink the curve when it is located at the background and aims at expand the curve when it is located inside a moving object.
- The third term is an intensity-based force $\{\mathbf{f}_I\}$ that moves the curve at the direction that creates interior regions with the desirable intensity properties. In other words, the curve is expanded if the it is located inside an object and is shrunk if it is located at the background.
- Finally, the last term is a visual consistency force $\{\mathbf{f}_C\}$ that deforms the curve in the direction that minimizes the intensity error between the interior curve region and the object position at the previous frame.

4.3 Level Set Formulation

The obtained PDE can be implemented using a Lagrangian approach, that is limited since it cannot deal with topological changes of the moving front and suffers from instability in the domain of numerical approximations.

This can be avoided by introducing the work of Osher and Sethian [16] in our scheme. The central idea is to represent the moving front $\partial R(t)$ as the zero-level set $\{\Phi = 0\}$ of a function Φ. This representation of $\partial R(t)$ is implicit, parameter-free and intrinsic. Additionally, it is topology-free.It is easy to show, that if the embedding function Φ deforms according to

$$\frac{d}{dt}\Phi(p, t) = \mathbf{f}(p) \left| \nabla \Phi(p, t) \right|$$

then the corresponding moving front evolves according to:

$$\frac{d}{dt}C(p, t) = \mathbf{f}(p)\mathcal{N}$$

Thus, the minimization of the proposed geodesic active region objective function is equivalent to searching for a steady-state solution of the following equation:

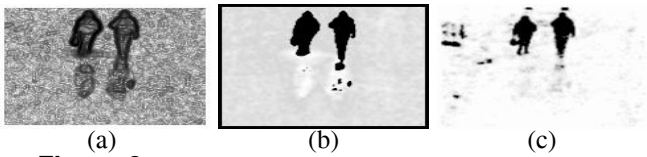
$$\frac{d\Phi}{dt}(u) = \Big(\alpha(\mathbf{f}_{B1} + \hat{\mathbf{f}}_{B2}) + \beta\, \mathbf{f}_D + \gamma\, \mathbf{f}_I + \delta\, \mathbf{f}_C \Big) \left| \nabla \Phi(u) \right|$$

where $\left[\hat{\mathbf{f}}_{B2} = \nabla g(\mathbf{B}(u)|\sigma_B) \cdot \frac{\nabla \Phi(u)}{|\nabla \Phi(u)|} \right]$.

The Level Set Equation is implemented using the **Hermes** algorithm [18] that proposes a fast way to deform locally the initial curve towards the minimum of the objective function.

4.4 Implementation Issues

The proposed algorithm is self-sufficient and works as follows: Given the *first*, the *reference* frame and an initial curve, the Geodesic Active Region model is activated, and detects the moving objects using the boundary [**Section** 3.1] and the motion detection module [**Section** 3.2]. Then, each object is associated to a motion model, and to an intensity-based descriptor (it is estimated directly from the observed values). Then, for each object we estimate the



(a) (b) (c)

Figure 2. Tracking Modules for the Walking in Sweden Sequence [**fig.** (4.3), second frame]. (a) Boundary, (b) Motion Detection, (c) Visual Consistency.

motion model between the *current* and the *next* frame. This model is used to initialize the curve at the next frame.

Then, we proceed to the next frame. We use the inverse motion model to create correspondence between the current curve region and the corresponding object at the *previous* frame, and we perform temporal tracking using the complete set of tracking modules, thus the curve position is updated [**Section** 4.2]. Then, given the latest curve position, we update the motion estimates [**Section** 4.2]. Then, these estimates are used to update the visual consistency module. These two complementary steps are performed until the optimization of the motion model, which also gives the best tracking result. Then, we update the reference frame and the intensity descriptors, and we proceed to the *next* frame.

Finally, we have to define the weights $\{\alpha, \beta, \gamma, \delta\}$ of the different tracking modules. Experimentally, it has been found that the visual consistency module provides the most reliable information [**fig.** (2.c); the white corresponds to positive, while the black to negative propagation velocities]. Besides, the motion detection module is quite reliable [**fig.** (2.b); similar interpretation with the visual consistency module]. On the other hand, the intensity module fails constantly, since it is based on global statistics and the quality of our sequences is pure. Finally, the boundary module has an unaccountable behavior [**fig.** (2.a)]. This module presents negative values only due to the curvature effect, thus the curve is propagated towards one direction under a regularity constraint. Thus, if the initial curve has a part inside the object and a part outside the object, then this term has a beneficial contribution for the exterior part, while it discourages the interior part to evolves towards the correct direction (outwards). On the other hand, this term is very important since it ensures the regularity of the curve. We take these remarks into consideration and we determine the modules contributions as follows $\{\alpha = 0.30, \beta = 0.25, \gamma = 0.05, \delta = 0.40\}$. Besides, when the final result is obtained, we can refine the final curve to the real object boundaries by applying a correction step using only the boundary-based module.

5 Conclusions, Results

Very promising experimental results have been obtained using the proposed framework for various real outdoor video sequences. Thanks to the incremental way of motion estimation, we can deal with important motion displacements (highway, Soccer sequence) [**fig.** (3)][1]. Besides, the tracking results are very promising and the object deformations are successfully handled [**fig.** (4)][2].

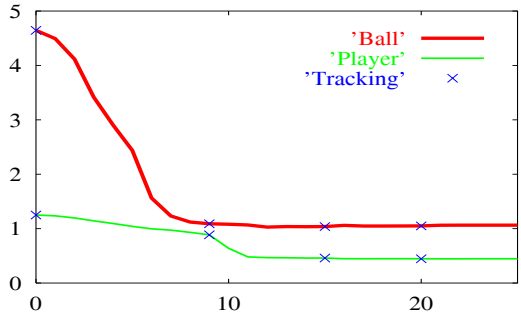


Figure 3. Motion Estimation: Mean Square Error.

Summarizing, in this paper we presented a new front propagation approach to deal with the motion estimation and the tracking problem. The main contribution of this work consists of **creating a coupled front propagation model that deals simultaneously with the motion estimation and the tracking problem, and combines different types of information** (*e.g. boundary, region*) **and different sources** (*e.g. boundary, intensity, motion*). This leads to a multi-propagation system where several modules operate simultaneously. The contour propagation is guided by *smoothing*, *boundary-based*, and *region-based* forces. The proposed model preserves robustness thanks to the coupling between the motion estimation and the tracking problem (visual consistency module), and is *independent* from the initialization step thanks to the level set implementation and to the region-based terms which create data-dependent positive and negative propagation forces.

As it concerns the future directions of this work, we would like to incorporate a multi-phase level set propagation which will permit us to deal with the occlusion cases. Besides, for the time being we try to extend our method for cases with a mobile observer (work in progress).

An extended version of this paper can be found in [19]. Various experimental results (in MPEG format), including the ones shown in this article, can be found at:
http://www.inria.fr/robotvis/personnel/nparagio/demos
http://www.inria.fr/robotvis/personnel/der/der-eng.html

[1] The graph corresponds to the motion estimates between the first two frames of the Soccer sequence [**fig.** (4.(1,c))]. The X-axis of the graph corresponds to the total iteration number, while the Y-axis, to the mean square error. The presence of crosses denotes that before this step the curve position is updated (temporal tracking). Besides, thanks to the incremental way of motion estimation, we can deal with important motion displacements (highway, Soccer sequence).

[2] Three real outdoor sequences have been used to validate our approach. The images on each row are intendical. The first column corresponds to the initial curve, and the second to the final curve using the proposed approach. Besides, the last column shows the motion field between the current and the next frame, which is used to initialize the curve position.

Finally, as it concerns the computational cost of our approach, it is related with the number of times that we couple the two unknown variables (curve position, motion parameters), as well as the number of iterations which are performed to the motion estimation step. Approximately, we can say that for a 160x128 sequence (Walking in Sweden), where we couple two times the unknown variables and we perform 20 iterations to each motion estimation step, we need about 2.25 seconds per frame using an ULTRA 10, 299 MHz.

References

[1] H. Adelson and Y. Weiss. Unified Mixture Framework for Motion Segmentation: Incorporating Spatial Coherence and Estimating the number of Models. In *IEEE CVPR*, San Fransisco, USA, 1996.

[2] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical Model-Based Motion Estimation. In *ECCV*, pages 237–252, Santa Margarita, Italy, 1992.

[3] M. Black. Combining Intensity and Motion for Incremental Segmentation and Tracking over long Image Sequences. In *ECCV*, pages 485–493, Santa Margarita, Italy, 1992.

[4] V. Caselles and B. Coll. Snakes in Movement. *SIAM Journal on Numerical Analysis*, 33:2445–2456, 1996.

[5] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *IJCV*, 22:61–79, 1997.

[6] A. Chakraborty, H. Staib, and J. Duncan. Deformable Boundary Finding in Medical Images by Integrating Gradient and Region Information. *IEEE Transactions on Medical Imaging*, 15(6):859–870, 1996.

[7] F. Heitz and P. Bouthemy. Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields. *IEEE PAMI*, 15:1217–1232, 1993.

[8] M. Isard and A. Blake. Contour Tracking by Stochastic Propagation of Conditional Density. In *ECCV*, volume I, pages 343–356, Cambridge, UK, 1996.

[9] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *IEEE ICCV*, pages 810–815, Boston, USA, 1995.

[10] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Conformal curvature flows: from phase transitions to active vision. *Archive of Rational Mechanics and Analysis*, 134:275–301, 1996.

[11] F. Leymarie and M. Levine. Tracking deformable objects in the plane using an active contour model. *IEEE PAMI*, 33:617–634, 1993.

[12] D. Lowe. Robust model based motion tracking through the integration of search and estimation. *International Journal of Computer Vision*, 8:113–122, 1992.

[13] R. Malladi, J. Sethian, and B. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE PAMI*, 17:158–175, 1995.

[14] F. Meyer and P. Bouthemy. Region-based tracking using affine motion models in long image sequences. *CVGIP: Image Understanding*, 60:119–140, 1994.

[15] J-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6:348–365, 1995.

[16] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed : algorithms based on the hamilton-jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[17] N. Paragios and R. Deriche. Geodesic Active Regions for Texture Segmentation. Research Report 3440, INRIA, France, 1998. http://www.inria.fr/rapports/sophia/RR-3440.html.

[18] N. Paragios and R. Deriche. A PDE-based Level Set approach for Detection and Tracking of moving objects. In *IEEE ICCV*, pages 1139–1145, Bombay, India, 1998.

[19] N. Paragios and R. Deriche. Geodesic Active Regions for Motion Estimation and Tracking. Research Report 3631, INRIA, France, 3631 1999. http://www.inria.fr/rapports/sophia/RR-3631.html.

[20] N. Paragios and R. Deriche. Unifying Boundary and Region-based Information for Geodesic Active Tracking. In *IEEE CVPR*, Colorado, USA, 1999.

[21] N. Paragios and G. Tziritas. Adaptive Detection and Localization of Moving Objects in Image Sequences. *Signal Processing: Image Communication*, 14:277–296, 1999.

[22] S. Zhu and A. Yuille. Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE PAMI*, 18:884–900, 1996.

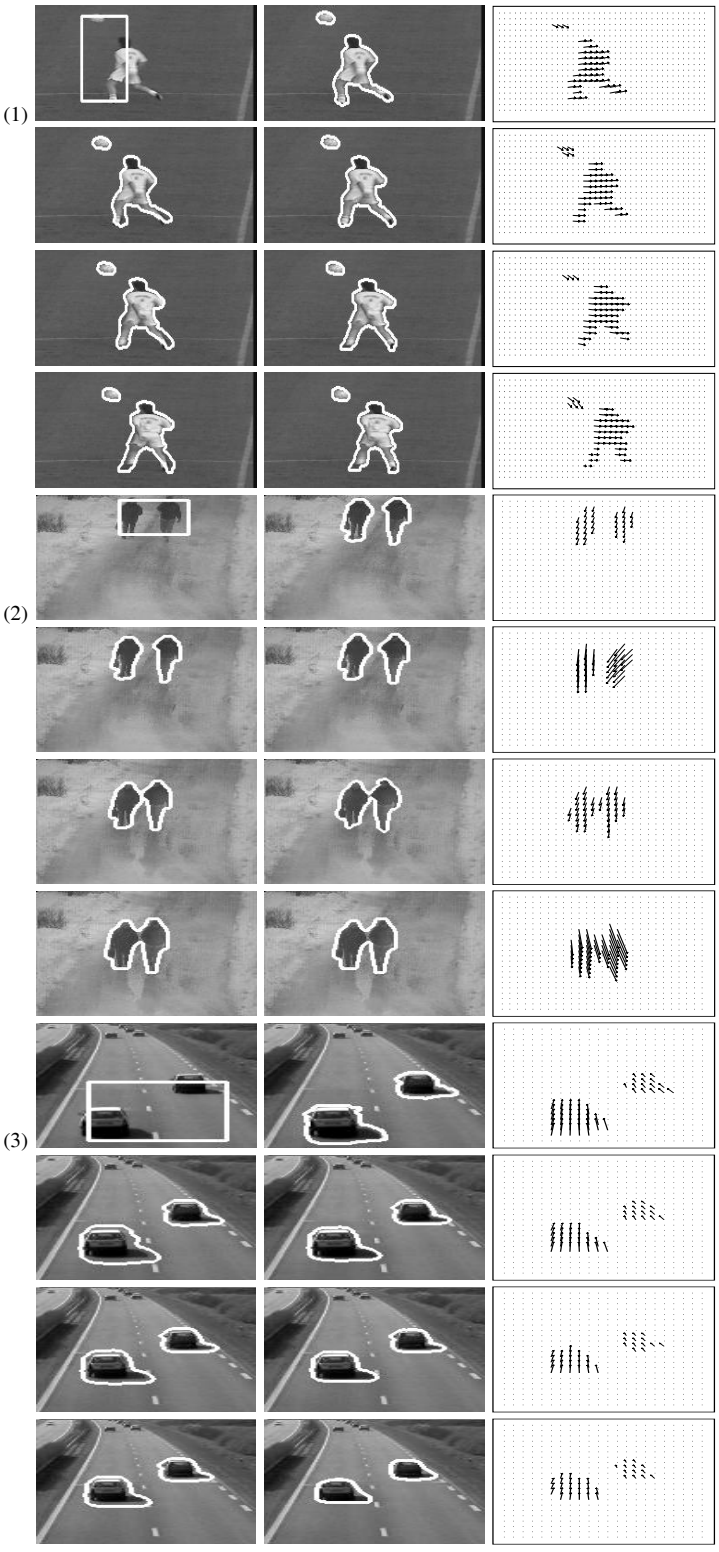


Figure 4. (a) Initial Curve , (b) Final Curve, (c) Motion Estimation. (1) Soccer Sequence, (2) Walking in Sweden Sequence, (3) Highway Sequence.