Geodesic Active Regions for Motion Estimation and Tracking

Nikolaos Paragios, Rachid Deriche

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Geodesic Active Regions for Motion Estimation and Tracking

Nikos PARAGIOS and Rachid DERICHE

N° 3631

Mars 1999

THÈME 3

ISSN 0249-6399

Geodesic Active Regions for Motion Estimation and Tracking

Nikos PARAGIOS and Rachid DERICHE

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Robotvis

Rapport de recherche n° 3631 — Mars 1999 — 30 pages

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, while given the curve position an analytical solution is obtained for the motion model. Besides, the PDE is implemented using a level set approach where topological changes are naturally handled. Very promising experimental results are provided using real video sequences.

Key-words: *Tracking, Motion Detection, Motion Estimation, Visual Consistency, Affine Motion Models, Geodesic Active Regions, Front Propagation, Level Set.*

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

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : 04 92 38 77 77 - International : +33 4 92 38 77 77 — Fax : 04 92 38 77 65 - International : +33 4 92 38 77 65

Régions Actives Géodésiques pour l'Estimation du Mouvement et le Suivi d'Objets

Résumé : Ce rapport décrit une nouvelle méthode à base de propagation de fronts pour traiter le problème important du suivi d'objets non rigides en mouvement dans une séquence d'images. On utilise un modèle de **Régions Actives Géodésiques** qui fait usage des informations liées aux régions et aux contours des objets en mouvement pour suivre et estimer au mieux le modèle linéaire de mouvement attaché aux objets suivis, qui peuvent fusionner ou se séparer durant leur évolution. Une approche variationnelle permettant de coupler et de reformuler le problème du suivi et de l'estimation du mouvement comme un problème de propagation de fronts est à la base de cette nouvelle méthode. Une énergie à minimiser est associée au principe variationnel que doivent respecter les positions des contours des objets suivis et leur modèle de mouvement. L'équation d'Euler-Lagrange associée est alors utilisée afin de déformer les contours initiaux, considérés comme des frontières de régions d'objets en mouvement. Ces courbes vont se déplacer sous l'influence de forces associées aux contours, aux régions et aux mouvements des objets suivis. La méthode des courbes de niveaux est mise en oeuvre durant le processus d'évolution des contours, ce qui permet de gérer automatiquement d'éventuels problèmes de changement de topologie de type fusion et/ou scission d'objets au cours du processus de suivi. Plusieurs résultats expérimentaux, obtenues à partir de séquences d'images réelles, illustrent les remarquables potentialités de cette approche.

Mots-clés : *Suivi d'Objets, Estimation du Mouvement, Contours actifs Géodésiques, EDP, Courbes de niveau, Models Affine, Régions Actives Géodésiques.*

Summary

- **What is the original contribution of this work?** The originality of the proposed model relies on the coupling between the motion estimation and the tracking problem under a front propagation framework. The object trajectories are determined using affine linear models that are incorporated to the curve position-based objective function as region-based terms. This coupling leads to a two direction multi-module front propagation tracking system, where the motion parameters and the exact object position are simultaneously recovered. Additionally, the tracking is ensured by the use of several tracking modules that are expressed as boundary or region-based (intensity, motion detection, visual consistency) terms.

- **Why should this contribution be considered important?** This contribution should be considered important, because it deals simultaneously with the two most important problems of motion analysis, the motion estimation and the tracking which provide a good basis for the estimation of 3D motion and/or structure from time varying images. Besides, we have developed a very promising self-sufficient multi-module framework that can track several rigid and non-rigid moving objects, and can use different types of information (intensity, motion, etc.) and different forms (boundary, region). Additionally, thanks to the level set methodology and to the region-based terms, we propose a robust multi-module tracking tool which is free from the initialization step and deals successfully with topological changes, thereby allowing multiple moving objects to be pursued simultaneously.

- **What is the most closely related work by others?**

  The most related work with our approach can be found in [5]. However they propose a sequentially three step approach that differs from the unified approach we present in this article. Following their previous work on geodesic active contours, they first start by detecting the contours of the objects to be tracked. An estimation of the velocity vector field along the detected contours is then performed using a completely separate approach, and finally another PDE is designed to move the contours to the boundary of the moving objects. These contours are then used as initial estimate of the contours in the next image, and the process is repeated. Our method deals simultaneously with the motion estimation and the tracking problem. This is a considerable extension of our previous work [35]. There, we have proposed the incorporation between boundary and region-based modules for tracking, but we didn't make any use of motion-based information apart from a statistical-based motion detection module. Thus, here we introduce a coupled framework that deals simultaneously with the motion estimation (affine models are assumed) and the tracking problem, where different boundary and

region-based tracking modules are used to provide a robust multi-propagation tracking tool.

- **How can other researchers make use of the results of this work ?**

  This work will be of great interest for researchers working on motion analysis, and tracking and curve evolution theory. The ideas presented in this paper can also be applied to develop different applications, that can be reformulated as frame partition problems. We have already applied the proposed framework to the texture segmentation problem [32], and we are currently investigating the use of the **Geodesic Active Region** framework to other domains.

1 Introduction

The tracking problem has a wide variety of applications in computer vision and motion analysis, including indexing [11], coding [41], video surveillance [20] and robotics [40]. Additionally, it provides a good basis for high level tasks of computer vision like 3-D reconstruction and representation [28, 42]. 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-based 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:

- **Motion-based** approaches rely on a robust method for grouping visual motion consistencies over time. These methods focus on recovering either the complete 3-D motion from the optical flow field [3] or 2-D motion models [2, 4, 10, 27]. Their main advantage is the fact that they are quite fast but they cannot ensure that the tracked regions have any meaning, since they are based on the a visual consistency constraint and they don't assume any object description or representation.

- **Model-based** approaches impose high-level semantic representation and knowledge and are more reliable compared with the motion-based approaches. These representations are either 3-D object models [7, 12, 19, 23, 38] or their projections, that are 2-D geometrical shapes [26, 14, 22, 20]. On the other hand they suffer from being computationally expensive due to the need to cope with scaling, translation, rotation, and deformation.

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:

- **Boundary-based** methods (they are usually referred as edge-based approaches) rely on the information provided by the object boundaries (shape properties). The idea of using boundary-based features to provide tracking has been widely adopted. The boundary-based features (edges) are well adapted to the tracking problem since they provide reliable information which is not dependent on motion type, as well as on the object shape. Usually the boundary-based tracking algorithms employ active contour models, like snakes [1, 8, 16] or geodesic active contours [6, 17, 25]. Both models are energy-minimized approaches that evolve from an initial curve under the influence of external potentials, while being constrained by internal energies. These models are quite flexible due to the fact that no restriction is imposed on the shape of the target nor on the type of motion: rigid or deformable. These methods require a good initialization step, since the initial contour converges iteratively toward the solution.

There are several approaches that have utilized a snake model to perform tracking like deformable models that have been widely adopted [15, 43]. On the other hand the use of geodesic active contours doesn't impose any shape constraint and can deal successfully with various deformations of the tracked object. The proposed approaches are limited [5, 33] and they suffer from instability, since the model is strongly depended on local-based information that can be inaccurate (**e.g.** occlusion cases).

- **Region-based** approaches rely on information provided by the entire region [29, 37, 39] (texture and motion-based properties). They use a motion estimation segmentation technique. In this case, the estimation of the velocity of the target is based on a correspondence between the associated target regions at different time instants where parametric models are usually employed to describe the target motion.These models have difficulty in tracking the target boundaries successfully, especially in the case of non-rigid targets. On the other hand these methods are robust due to the fact that they make use of information provided by the whole region and do not depend (strongly) on the initialization step.

The main 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 [33, 35] by introducing a visual consistency module. In [35], we have proposed the incorporation between boundary and region-based modules for tracking, but we didn't make any use of motion-based information apart from a statistical-based motion detection module. Thus, here we introduce a coupled framework that deals simultaneously with the motion estimation (affine models are assumed) and the tracking problem. This coupling leads to a two direction multi-module 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, an edge detector is applied to the absolute difference frame (between the current and the background) to provide the boundary-based information [**fig.** 1: **Boundary Module**].

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**]. Besides, we assume that a knowledge about the background as well as the moving intensity properties is available

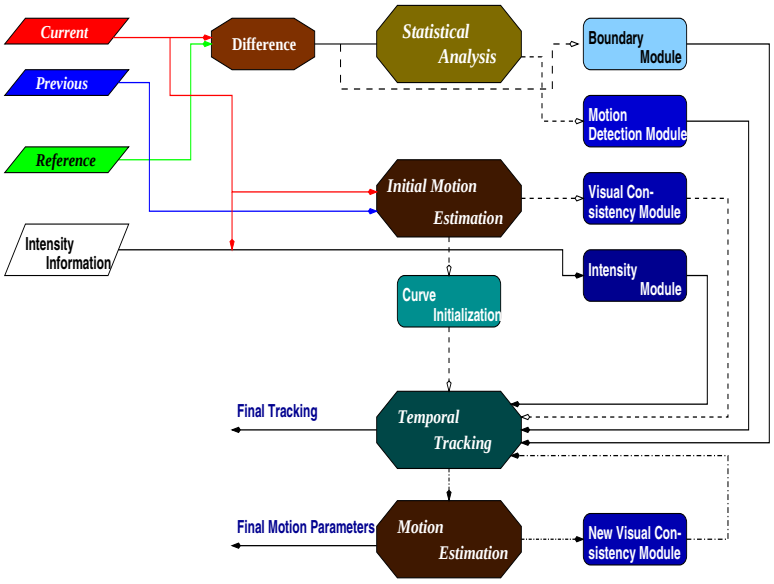


Figure 1: Geodesic Active Regions for tracking.

(it is gained during the process) and is used to define the intensity-based information [**fig.** 1: **Intensity Module**]. 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**].

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 that deforms the initial curve, consists of *boundary, intensity* and *motion-based* forces and is constrained by a *regularity* force. This PDE is implemented using a Level-Set approach where topological changes are naturally handled [30], thereby allowing multiple moving objects to be pursued simultaneously. Besides, an incremental method is used to update the estimation of the motion parameters. Both problems are treated under a coupled framework, where the latest curve position is used to update the estimation of the motion parameters, while the latest estimation of these parameters creates forces that move the curve towards the real object position. Finally, the level-set propagation is performed using a very fast front propagation algorithm [33]. 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 Region

The Geodesic Active Contour model and the Level Set methods have been initially proposed in [6, 17, 25], and successfully applied to a wide variety of computer vision applications [5, 9, 18, 21, 24, 34]. These models are based on boundary-based information, and aims at finding the best minimal-length smooth curve that takes into account the desired image properties.

Additionally, motivated by the excellent work proposed in [44], the Geodesic Active Region model has been initially introduced in [32] to deal with the problem of supervised texture segmentation and has been successfully exploited in [35] to deal with the tracking problem. This model is a considerable extension to the classic geodesic active contour model since it incorporates region-based information and aims at finding a partition [determined by the best minimal-length smooth curve that takes into account the desired image properties] where the interior as well as the exterior region preserves the desired image properties.

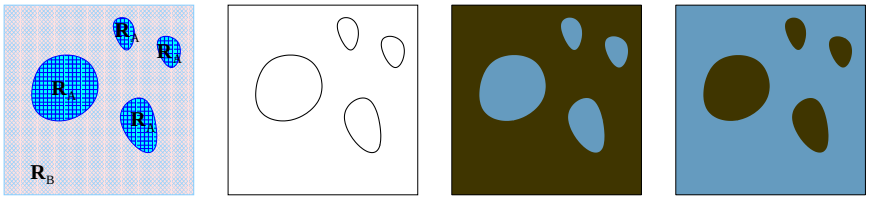
We are going introduce this model for a simple segmentation case with two possible decisions. In order to facilitate the notation, let us make some definitions:

- 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, where $\mathcal{R}_A$ is the region of interest (hypothesis h_A).
- And, let $\{\partial\mathcal{R}_A\}$ be the $\mathcal{R}_A$ region boundaries.

If we assume that for the given frame [**fig.** (2.a)] some information regarding the real region boundaries is available [**fig.** (2.b)], then the extraction of the region of interest can be viewed as the problem of accurately extracting its boundaries.

Let $[\mathbf{p}(\mathbf{I}(\mathbf{s})|\mathbf{B})]$ be the **conditional boundary density function** that measures the probability of a given point being at the real boundaries of $\mathcal{R}_A$. Then, the region of interest can be obtained using the geodesic active contour framework, thus minimizing

$$E(\partial\mathcal{R}_A) = \alpha \underbrace{\int_0^1 \left|\partial\dot{\mathcal{R}}_A(p)\right|^2 dp}_{Internal\ Term} + (1-\alpha) \underbrace{\int_0^1 g^2(p(\mathbf{I}(\partial\mathcal{R}_A(p)))|B)dp}_{External\ Term} \quad (1)$$



(a) (b) (c) (d)

Figure 2: Geodesic Active Region Model: (a) the input, (b) the boundary-based information, (c) the region-based information correspond to hypothesis h_A, [the information is proportional to the frame intensities] (d) the region-based information correspond to hypothesis h_B [the information is proportional to the frame intensities].

where $\partial\mathcal{R}_A(p) : [0,1] \rightarrow \mathbf{R}^2$ is a parameterization of the region boundaries in a planar form, the dot operator $\{\partial\dot{\mathcal{R}}_A(p)\}$ denotes the partial derivative with respect to time, $\alpha \in [0,1]$ is a positive constant balancing the contribution of the two terms, and g is a monotonically decreasing function. Finally, the internal term accounts for the expected spatial properties (*i.e. regularity*) of the curve, while the external term stands for the *attraction* of the curve towards the real region boundaries.

The minimization of the designed functional leads to a geodesic curve with a new metric [6],

$$E(\partial\mathcal{R}_A) = \int_0^1 g(p(\mathbf{I}(\partial\mathcal{R}_A(p)))|B) \left|\partial\dot{\mathcal{R}}_A(p)\right| dp \tag{2}$$

Let 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 [**fig.** (2.c,2.d)].

Then, the extraction of the region of interest is equivalent with creating a consistent frame partition between *the observed data, the associated hypothesis and their expected properties.* This partition can be viewed as an optimization problem with respect to the *a posteriori* segmentation probability, given the observation set.

Let $[p(\mathcal{P}(\mathcal{R})|\mathbf{I})]$ be the *a posteriori* segmentation density function with respect to the different partitions $\mathcal{P}(\mathcal{R})$ given the input data $\mathbf{I}$. This density function is given by the Bayes rule as:

$$p(\mathcal{P}(\mathcal{R})|\mathbf{I}) = \frac{p(\mathbf{I}|\mathcal{P}(\mathcal{R}))}{p(\mathbf{I})} p(\mathcal{P}(\mathcal{R})) \tag{3}$$

If we assume that all the partitions are *a priori* equally possible then we can ignore the constant terms $p(\mathbf{I})$, $p(\mathcal{P}(\mathcal{R}))$ and we can rewrite the density function as:

$$p(\mathcal{P}(\mathcal{R})|\mathbf{I}) = \; p(\mathbf{I}|\{\mathcal{R}_A, \mathcal{R}_B\}) = p([\mathbf{I}_A \mid \mathcal{R}_A] \cap [\mathbf{I}_B \mid \mathcal{R}_B]) = \; p(\mathbf{I}_A \mid \mathcal{R}_A)\; p(\mathbf{I}_B \mid \mathcal{R}_B) \quad (4)$$

Besides, if we assume that the points within each region are independent, we can replace the region probability with:

$$p(\mathbf{I}_X \mid \mathcal{R}_X) = \prod_{s \in \mathcal{R}_X} p_X(\mathbf{I}(s)) \quad (5)$$

The maximization of *a posteriori* probability is equivalent with the minimization of the [**-log**()] function (which transforms the probability product to a sum) of this probability,

$$E(\partial\mathcal{P}(\mathcal{R})) = -\iint_{\mathcal{R}_A} \mathbf{log}\left[p_A(\mathbf{I}(x,y))\right] dxdy - \iint_{\mathcal{R}_B} \mathbf{log}\left[p_B(\mathbf{I}(x,y))\right] dxdy \quad (6)$$

We fuse the two different segmentation models by defining the Geodesic Active Region objective function as

$$\begin{aligned} E(\partial\mathcal{R}_A) = \alpha \int_0^1 g(p(\mathbf{I}(\partial\mathcal{R}_A(p))))|\partial\dot{\mathcal{R}}_A(p)|dp - \\ (1-\alpha)\left\{\iint_{\mathcal{R}_A} \mathbf{log}\;\left[p_A(\mathbf{I}(x,y))\right] dxdy + \iint_{\mathcal{R}_B} \mathbf{log}\left[p_B(\mathbf{I}(x,y))\right] dxdy\right\} \end{aligned} \quad (7)$$

The minimization of this function is performed using a gradient descent method. If $u = (x, y)$ is a point of the initial curve and we compute the Euler-Lagrange equations [6, 44], then we should deform the curve to this point using the following equation:

$$\begin{aligned} \frac{du}{dt} = -\alpha\left[\mathbf{log}(p_A(\mathbf{I}(u)))\vec{\mathcal{N}}_{\partial R_A}(u) + \mathbf{log}(p_B(\mathbf{I}(u)))\vec{\mathcal{N}}_{\partial R_B}(u)\right] \\ +\; (1-\alpha)\left[g(u)\mathcal{K}_{\partial R_A}(u) - \nabla g(u)\cdot\vec{\mathcal{N}}_{\partial R_A}(u)\right]\vec{\mathcal{N}}_{\partial R_A}(u) \end{aligned} \quad (8)$$

where $\mathcal{K}_{\partial R_A}$ is the Euclidean curvature of ∂R_A and $\vec{\mathcal{N}}_{\partial R_A}$ (*resp.* $\vec{\mathcal{N}}_{\partial R_B}$) is the unit inward normal to ∂R_A (*resp.* ∂R_B). Since the curves ∂R_A and ∂R_B have inverse inwards vectors at each common point u, we have $\vec{\mathcal{N}} = \vec{\mathcal{N}}_{\partial R_A} = -\vec{\mathcal{N}}_{\partial R_B}$. Taking this into account and replacing $\mathcal{K}_{\partial R_A}$ by $\mathcal{K}$, the motion equation for u can be rewritten as:

$$\begin{aligned} \frac{du}{dt} = \underbrace{\left[\alpha\mathbf{log}\left(\frac{p_B(\mathbf{I}(u))}{p_A(\mathbf{I}(u))}\right)\right.}_{region\quad term} + \\ \underbrace{\left.(1-\alpha)\left(g(u)\mathcal{K}(u) - \nabla g(u)\cdot\vec{\mathcal{N}}(u)\right)\right]}_{boundary\quad term}\vec{\mathcal{N}}(u) \end{aligned} \quad (9)$$

The obtained PDE motion equation has two kind of *forces* acting on the curve, both in the direction of the normal.

- The **region force** aims to shrink or to expand the curve to the direction that maximizes the *a posteriori* segmentation probability. 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 $\left[\alpha \mathbf{log}\left(\frac{p_B(\mathbf{I}(u))}{p_A(\mathbf{I}(u))}\right) > 0\right]$ otherwise acts to expand it.
- The **boundary force** contains information regarding the region boundaries and is composed of two sub-terms; one that shrinks or expands the curve constrained by the curvature effect towards the object boundaries and one that attracts the curve to the objects boundaries (refinement term).

2.1 Generalizing the model

The proposed model can be generalized quite easily [32, 35] as follows:

$$E(\partial\mathcal{R}_A) = (1-\alpha)\underbrace{\int_0^1 g(p(\mathbf{I}(\partial\mathcal{R}_A(p))))|\partial\dot{\mathcal{R}}_A(p)|dp}_{Boundary\ Term} + \underbrace{\alpha\left\{\iint_{\mathcal{R}_A}\xi\ [p_A(\mathbf{I}(x,y))]\,dxdy + \iint_{\mathcal{R}_B}\xi\,[p_B(\mathbf{I}(x,y))]\,dxdy\right\}}_{Region\ Term} \tag{10}$$

where $\{g(), \xi()\}$ are positive monotonically decreasing functions (*e.g.* Gaussians).

Let us now try to interpret the above functional. The boundary term aims at finding the best minimal length curve that is attracted by the real region boundaries of $\mathcal{R}_A$. The region term uses this curve to define a partition where the interior curve region corresponds to $\mathcal{R}_A$, while the exterior to $\mathcal{R}_B$. This partition 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 initial curve (that can belong either to $\mathcal{R}_A$ or to $\mathcal{R}_B$) and we compute the Euler-Lagrange equations (using the Stokes theorem) [6, 44], then we should

deform the curve to this point using the following equation:

$$\frac{du}{dt} = \underbrace{[\alpha\left(\xi(p_A(\mathbf{I}(u))) - \xi(p_B(\mathbf{I}(u)))\right)}_{region-based\ force} + \underbrace{(1-\alpha)\left(g(u)\mathcal{K}(u) - \vec{\nabla} g(u)\cdot\vec{\mathcal{N}}(u)\right)]}_{boundary-based\ force}\vec{\mathcal{N}}(u) \tag{11}$$

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 information 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 effect 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 $D(x,y;t)$ the current difference frame: $[D = I - R]$. An obvious way to provide the boundary-based information is to seek high gradient values in the current or in the difference

frame. The use of an edge-based detector on the input frame could provide reliable boundary-based information for the moving objects, but it will provide similar information for the high contrast parts of the static background. On the other hand, the use of the difference frame eliminates the static background edges, thus the boundary information can be determined by the the high gradient values of the difference frame (the absolute values are considered $|D(x,y;t)|$), and is captured using a Gaussian function

$$g(\mathbf{B}(x,y)|\sigma_B) = \frac{1}{\sigma_B\sqrt{2\pi}} e^{-\frac{\mathbf{B}^2(x,y)}{2\sigma_B^2}} \tag{12}$$

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

$$\mathbf{E_B}(\partial\mathcal{P}(\mathcal{R})) = \sum_{i=1}^{N} \int_0^1 g(\mathbf{B}(\partial\mathcal{R}_i(p_i))|\sigma_B) \left|\partial\dot{\mathcal{R}}_i(p_i)\right| dp_i \tag{13}$$

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. The interpretation of this term is evident, since we ask for curves that are attracted by the real object boundaries.

3.2 Motion Detection Module

Besides, if we assume that the difference $\{\mathbf{D}\}$ frame is a selection of independent points, then it is composed of two populations [36]. The *static* one contains the background points while the *mobile* one contains the points that belong to moving objects and usually preserve different illumination properties with respect to the corresponding background illumination properties. Additionally, we could assume that the *mobile* population can be decomposed into many different sub-populations with respect to the different intensity properties preserved by the moving objects. These assumptions can be easily projected to a statistical model, where the observed density function of the difference frame can be decomposed into two main statistical components, the static and the mobile one,

$$p_D(d) = P_{static}\ p_S(d|\Theta_S)\ +\ P_{mobile}\ p_M(d|\Theta_M) \tag{14}$$

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. Besides, we can consider that the conditional probability density function with respect to the *mobile* component is a collection of sub-components that express the different illumination properties of the observed objects (in terms of the difference frame). Thus, we can assume that a mixture density can be used

to model the statistical behavior of the *mobile* component, which is given by

$$p_D(d) = p_M(d|\Theta_M) = \sum_{i=1}^{C_N} P_{iM}\ p_M(d|\Theta_{iM}) \tag{15}$$

where P_{iM} is the *a priori* probability of the i component and Θ_{iM} the unknown density function parameters. By combining the last two equations, we obtain the following mixture model

$$p_D(d) = P_{static}\ p_S(d|\Theta_S)\ +\ P_{mobile} \sum_{i=1}^{C_N} P_{iM}\ p_M(d|\Theta_{iM}) \tag{16}$$

We assume that these probability density functions follow Gaussian or Laplacian law. As for the unknown parameters of this model, some constraints are imposed by the problem. It is quite obvious that differences between background values appear because of the presence of noise, and as a consequence, the conditional probability density function with respect to the static case is zero-mean. Additionally, we can assume that the mobile mixture model contains a zero-mean density function due to the fact that some moving objects may preserve similar intensity properties with respect to the background. The estimation of the unknown parameters of this model $\{(P_i, \Theta_i) : i \in \{0, ..., C_N\}\}$ is done using the maximum likelihood principle.

Then, it is evident that the regions that are associated to moving objects $\{\mathcal{R}_i : i \in [1, N]\}$ should 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 composed of static points. Taking this into account, the energy expression for the motion detection **region-based** module is defined as:

$$\mathbf{E_D}(\partial\mathcal{P}(\mathcal{R})) = \iint_{\mathcal{R}_0} g\left(p_S(\mathbf{D}(x,y;t))|\sigma_D\right) dxdy\ +\ \sum_{i=1}^{N} \iint_{\mathcal{R}_i} g\left(p_M(\mathbf{D}(x,y;t))|\sigma_D\right) dxdy \tag{17}$$

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 intensity properties.

3.3 Intensity Module

Since we assume a background reference frame $[R(x, y)]$, we can extract some intensity-based information related to it. This can be done by applying some invariant operators to this frame. 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

- $\vec{\mathbf{p}}_0(\vec{\mathbf{x}}) = \left(p_{0|x}(x), p_{0|\eta\eta}(x_{\eta\eta}), p_{0|\xi\xi}(x_{\xi\xi})\right)$

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:

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

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,

$$\mathbf{E_I}(\partial\mathcal{P}(\mathcal{R})) = \sum_{i=0}^{N} \sum_{o \in F} w_o \iint_{\mathcal{R}_i} g\left(p_{i|o}(\mathbf{I}(x,y;t))|\sigma_I\right) dxdy \tag{18}$$

where w_i are positive weights that take into account the operator uncertainties and $[g(.|\sigma_I)]$ is a Gaussian function.

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 $t-1$ $(s + A(s))$. We use the inverse motion model.

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)

$$E(A) = \iint_{O} \left[I(x,y;t) - I(x + A_x(x,y), y + A_y(x,y); t-1)\right]^2 dxdy \tag{19}$$

where $A(x,y) = [A_x(x,y), A_y(x,y)]$. These approaches have been extensively studied and used [13]. 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. Besides, 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,

$$E(\partial\mathcal{P}(\mathcal{R}), A) = \iint_{\mathcal{R}_0} [\mathbf{I}(x,y;t) - \mathbf{R}(x,y)]^2 \, dxdy \; + \sum_{i=1}^{N} \iint_{\mathcal{R}_i} [\mathbf{I}(x,y;t) - \mathbf{I}(x + A_{ix}(x,y), y + A_{ix}(x,y); t-1)]^2 \, dxdy \tag{20}$$

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

$$\mathbf{E_C}(\partial\mathcal{P}(\mathcal{R}), A) = \iint_{\mathcal{R}_0} h\,[\mathbf{I}(x,y;t) - \mathbf{R}(x,y)|\sigma_C]\, dxdy \; + \sum_{i=1}^{N} \iint_{\mathcal{R}_i} h\,(\mathbf{I}(x,y;t) - \mathbf{I}(x + A_{ix}(x,y), y + A_{ix}(x,y); t-1)|\sigma_C)\, dxdy \tag{21}$$

The interpretation of this term is quite obvious. 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 minimized. 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 minimized.

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, given by

$$A(x,y) = \begin{bmatrix} A_x(x,y) \\ A_y(x,y) \end{bmatrix} = \begin{bmatrix} \mathbf{a}_0 & \mathbf{a}_1 & \mathbf{a}_2 \\ \mathbf{b}_0 & \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ y \end{bmatrix} \tag{22}$$

3.5 The Complete Model

Following [31, 33, 32, 35, 44] we fuse the different functionals, by defining a Geodesic Active Region Tracking model:

$$E(\partial\mathcal{P}(\mathcal{R}), A) = \alpha\ \mathbf{E_B}(\partial\mathcal{P}(\mathcal{R})) + \beta\ \mathbf{E_D}(\partial\mathcal{P}(\mathcal{R})) + \gamma\ \mathbf{E_I}(\partial\mathcal{P}(\mathcal{R})) + \delta\ \mathbf{E_C}(\partial\mathcal{P}(\mathcal{R}), A) \quad (23)$$

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

$$\begin{aligned}\mathbf{e}(u) =& I(u;t) - I(x + A_x(u), y + A_y(u); t-1) \\ & I(u;t) - A_x(u)I_x(u;t-1) - A_y(u)I_y(u;t-1) - I(u;t-1)\end{aligned} \quad (24)$$

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

$$E(A_i) = \iint_{\mathcal{R}_i} \left[\mathbf{I}(x, y; t) - \mathbf{I}(x + A_{ix}(x, y), y + A_{ix}(x, y); t-1)\right]^2 dxdy \quad (25)$$

We propose an incremental way to update the motion parameters estimation 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,

$$E(A_i, \Delta A_i) = \iint_{\mathcal{R}_i} \left[\mathbf{I}(x, y; t) - \mathbf{I}\left(x + [A_{ix} + \Delta A_{ix}](x, y)\ \ , y + [A_{ix} + \Delta A_{iy}](x, y); t-1\right)\right]^2 dxdy \quad (26)$$

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

$$E(A_i, \Delta A_i) = \iint_{\mathcal{R}_i} \left[\mathbf{I}(x,y;t) - \nabla \mathbf{I} \left([\mathbf{1} + A_i]^T(x,y); t-1 \right) \begin{bmatrix} \Delta A_{ix}(x,y) \\ \Delta A_{iy}(x,y) \end{bmatrix} \right. \\ \left. - \mathbf{I}([\mathbf{1} + A_i]^T(x,y); t-1) \right]^2 dxdy \tag{27}$$

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,

$$\begin{bmatrix} \iint_{\mathcal{R}_i} \mathbf{I}_x^2 & \iint_{\mathcal{R}_i} x\mathbf{I}_x^2 & \iint_{\mathcal{R}_i} y\mathbf{I}_x^2 & \iint_{\mathcal{R}_i} \mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} x\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} y\mathbf{I}_x\mathbf{I}_y \\ \iint_{\mathcal{R}_i} x\mathbf{I}_x^2 & \iint_{\mathcal{R}_i} x^2\mathbf{I}_x^2 & \iint_{\mathcal{R}_i} xy\mathbf{I}_x^2 & \iint_{\mathcal{R}_i} x\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} x^2\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} xy\mathbf{I}_x\mathbf{I}_y \\ \iint_{\mathcal{R}_i} y\mathbf{I}_x^2 & \iint_{\mathcal{R}_i} xy\mathbf{I}_x^2 & \iint_{\mathcal{R}_i} y^2\mathbf{I}_x^2 & \iint_{\mathcal{R}_i} y\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} xy\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} y^2\mathbf{I}_x\mathbf{I}_y \\ \iint_{\mathcal{R}_i} \mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} x\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} y\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} \mathbf{I}_y^2 & \iint_{\mathcal{R}_i} x\mathbf{I}_y^2 & \iint_{\mathcal{R}_i} y\mathbf{I}_y^2 \\ \iint_{\mathcal{R}_i} x\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} x^2\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} xy\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} x\mathbf{I}_y^2 & \iint_{\mathcal{R}_i} x^2\mathbf{I}_y^2 & \iint_{\mathcal{R}_i} xy\mathbf{I}_y^2 \\ \iint_{\mathcal{R}_i} y\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} xy\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} y^2\mathbf{I}_x\mathbf{I}_y & \iint_{\mathcal{R}_i} y\mathbf{I}_y^2 & \iint_{\mathcal{R}_i} xy\mathbf{I}_y^2 & \iint_{\mathcal{R}_i} y^2\mathbf{I}_y^2 \end{bmatrix} \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} = \begin{bmatrix} \iint_{\mathcal{R}_i} \mathbf{e}\mathbf{I}_x \\ \iint_{\mathcal{R}_i} x\mathbf{e}\mathbf{I}_x \\ \iint_{\mathcal{R}_i} y\mathbf{e}\mathbf{I}_x \\ \iint_{\mathcal{R}_i} \mathbf{e}\mathbf{I}_y \\ \iint_{\mathcal{R}_i} x\mathbf{e}\mathbf{I}_y \\ \iint_{\mathcal{R}_i} y\mathbf{e}\mathbf{I}_y \end{bmatrix} \tag{28}$$

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.

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:

$$\frac{du}{dt} = \Bigg(\alpha \underbrace{\Big[g(\mathbf{B}(u)|\sigma_B)\ \mathcal{K}(u) \ - \ \vec{\nabla} g(\mathbf{B}(u)|\sigma_B) \cdot \vec{\mathcal{N}}(u) \Big]}_{boundary\ force\ \triangleq\ \mathbf{f}_{B1}+\mathbf{f}_{B2}} + \beta \underbrace{\left[g\left(p_M(\mathbf{D}(u))|\sigma_D\right) - g\left(p_S(\mathbf{D}(u))|\sigma_D\right) \right]}_{motion\ detection\ force\ \triangleq\ \mathbf{f}_D} + \gamma \underbrace{\sum_{o \in F} w_o \left[g\left(p_{\mathbf{ro}}(\mathbf{I}(u))|\sigma_I\right) - g\left(p_{0o}(\mathbf{I}(u))|\sigma_I\right) \right]}_{intensity\ segmentation\ force\ \triangleq\ \mathbf{f}_I} + \delta \underbrace{\left[h\left(\mathbf{I}(u;t) - \mathbf{R}(u)|\sigma_C\right) \ - \ h\left(\mathbf{I}(u;t) - \mathbf{I}(u + A_r(u); t-1)|\sigma_C\right) \right]}_{visual\ consistency\ force\ \triangleq\ \mathbf{f}_C} \Bigg) \vec{\mathcal{N}}(u) \tag{29}$$

where $\mathcal{K}$ is the Euclidean curvature and $\vec{\mathcal{N}}$ 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 contains information regarding the boundaries of the different moving objects 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 cure when it is located inside a moving object. The decision about the curve position is based on the observed values of the difference frame and the conditional probability density functions that have been assigned to the *static* and the *mobile* case.

- 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. The decision about the curve position is based on the observed intensities, and the intensity-based object/background conditional density functions.

- 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. Besides, this force as well as the intensity-based force ensures the tracking operation by creating a consistent correspondence between the object position over the time.

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 [30] 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

$$\frac{d}{dt}\Phi(p,t) = \mathcal{F}(p)\,||\nabla\Phi(p,t)|| \tag{30}$$

then the corresponding moving front evolves according to:

$$\frac{d}{dt}C(p,t) = \mathcal{F}(p)\vec{\mathcal{N}} \tag{31}$$

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

$$\frac{d\Phi}{dt}(u) = \Big(\alpha(\mathbf{f}_{B1} + \hat{\mathbf{f}}_{B2}) + \beta\ \mathbf{f}_D + \gamma\ \mathbf{f}_I + \delta\ \mathbf{f}_C\Big)\,||\nabla\Phi(u)|| \tag{32}$$

where $\Big[\hat{\mathbf{f}}_{B2} = \vec{\nabla} g(\mathbf{B}(u)|\sigma_B)\cdot\vec{\nabla}\Phi(u)\Big]$ and the geometric properties are estimated directly from the level set frame.

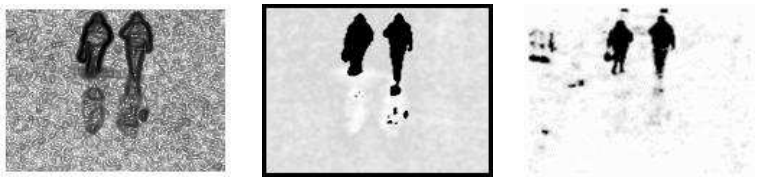
The Level Set Equation is implemented using the **Hermes** algorithm [33] that proposes a fast way to deform the initial curve towards the global minimum of the objective function. In our case the equation which deforms the initial curve can be rewritten in a more general form as:

$$\Phi^{t+1}_{(x,y)} = \Phi^{t}_{(x,y)} + \mathcal{V}(x,y,\Phi)dt \tag{33}$$

where $\mathcal{V}(x,y,\Phi)$ is the propagation speed function, depending on geometric features and image features. Since the speed $\mathcal{V}(x,y,\Phi)$ is basically estimated according to image characteristics, there are some points for which the front evolves faster compared to the others. The key idea on which the Hermes approach is based is to evolve the front locally according to the speed values of its points. The algorithm at each step selects the point with the highest absolute propagation speed from a set of actual curve points, and deforms the level-set image locally.

4.3 From Theory to Practice

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



(a) (b) (c)

Figure 3: Tracking Modules for the Walking in Sweden Sequence [**fig.** (7), second frame]. (a) Boundary, (b) Motion Detection, (c) Visual Consistency.

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 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 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.** (3); the white corresponds to positive, while the black to negative propagation velocities]. Besides, the motion detection module is quite reliable [**fig.** (3.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.** (3.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

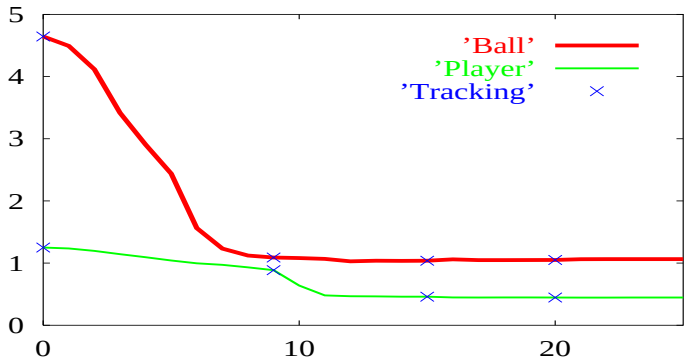


Figure 4: Motion Estimation: Mean Square Error.

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 with respect to the motion estimation [**fig.** (4)][1] and to the tracking [**fig.** (5,6,7)][2].

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,*

[1] This graph corresponds to the motion estimates between the first two frames of the Soccer sequence [**fig.** (6)]. 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.

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 some light geometric-based features (object representations) that will increase the robustness and 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).

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/

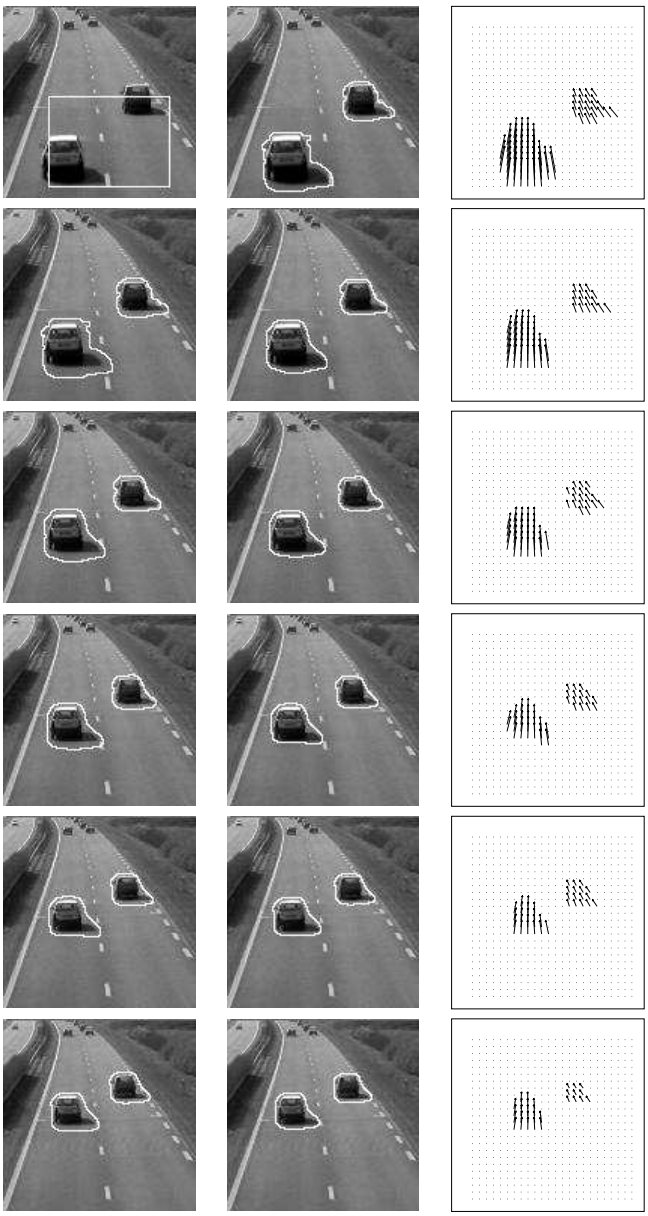
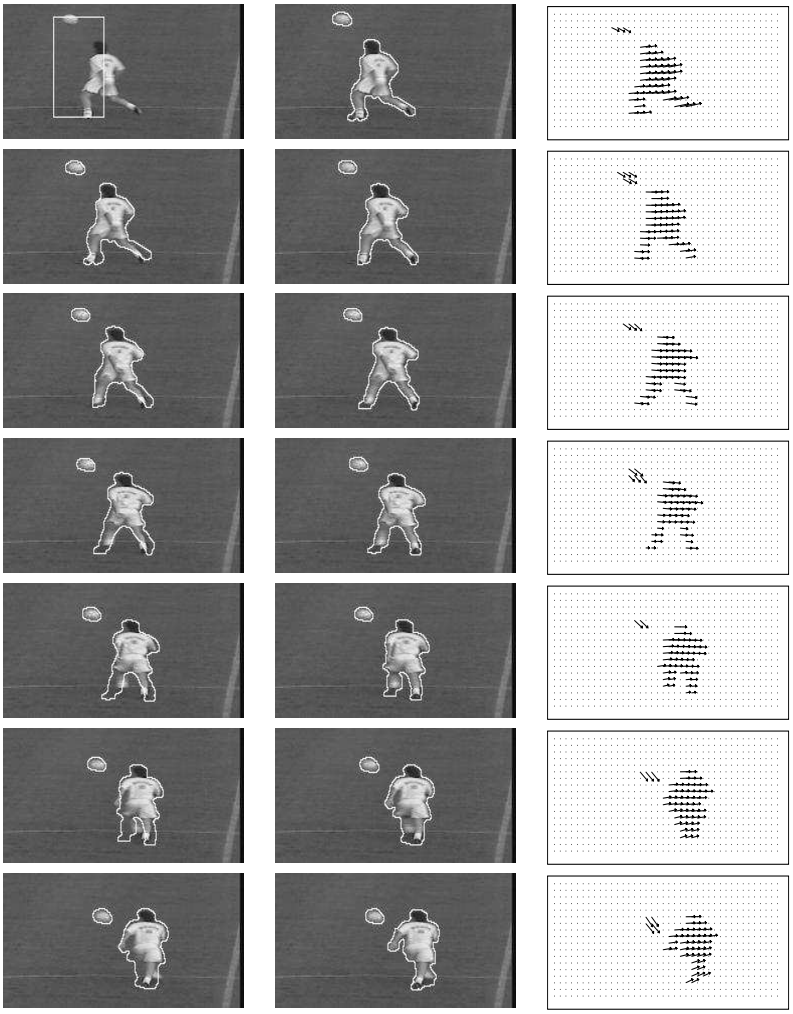


Figure 5: Highway Sequence:(a) Initial Curve , (b) Final Curve, (c) Motion Estimation.



(a) (b) (c)

Figure 6: Soccer Sequence:(a) Initial Curve , (b) Final Curve, (c) Motion Estimation.

References

[1] A. Blake and M. Isard. *Active Contours*. Springer-Verlag Press, 1997.

[2] N. Brady and N. O'Connor. Object detection and tracking using an em-based motion estimation and segmentation framework. In *International Conference on Image*

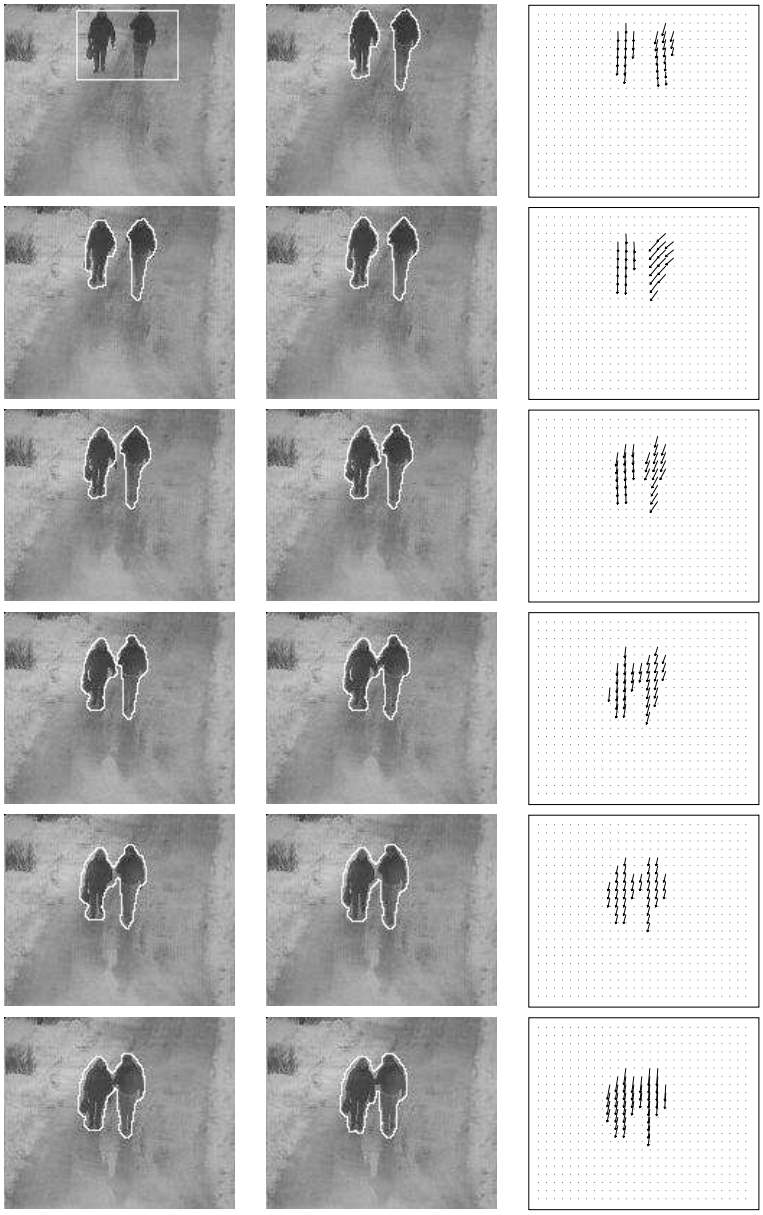


Figure 7: Walking Sequence:(a) Initial Curve , (b) Final Curve, (c) Motion Estimation.

Processing, 1996.

[3] T. Broida, S. Chandrashekhar, and R. Chellappa. Recursive 3-d motion estimation from a monocular image sequence. *IEEE Trans. Aerospace and Electronic Systems*, 26:639–656, 1990.

[4] T. Broida and R. Chellappa. Estimation of objects motion parameters from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:90–99, 1986.

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

[6] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22:61–79, 1997.

[7] J-C. Cheng and J. Moura. Tracking Human Walking in Dynamic Scenes. In *International Conference on Image Processing*, 1997.

[8] D. Cohen. On active contour models and balloons. *CVGIP: Image Understanding*, 53:211–218, 1991.

[9] O. Faugeras and R. Keriven. Variational principles, Surface Evolution, PDE's, level set methods and the Stereo Problem. *IEEE Transactions on Image Processing*, 7:336–344, 1998.

[10] M. Gelgon and P. Bouthemy. A region-level graph labeling approach to motion-based segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 514–519, Puerto Rico, USA, 1997.

[11] M. Gelgon and P. Bouthemy. A region-level motion-based graph representation and labeling for tracking a spatial image partition. *Pattern Recognition*, page to appear, 1999.

[12] D. Gravila and L. Davis. 3-D Model-based Tracking of Humans in Action: a Multi-view Approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, San Fransisco, USA, 1996.

[13] B. Horn. Non-correlation methods for stereo matching. *Photogrammetric Engineering and Remote Sensing*, 49:535–536, 1983.

[14] M. Isard and A. Blake. Contour Tracking by Stochastic Propagation of Conditional Density. In *European Conference on Computer Vision*, volume I, pages 343–356, Cambridge, UK, 1996.

[15] M-P. Jolly, S. Lakshmanan, and A. Jain. Vehicle segmentation and classification using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:293–308, 1996.

[16] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–332, 1988.

[17] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *International Conference on Computer Vision*, pages 810–815, Boston, USA, 1995.

[18] R. Kimmel and A. Bruckstein. Shape from shading: Level set propagation and viscosity solutions. *International Journal of Computer Vision*, 16:107–133, 1995.

[19] D. Koller, K. Daniilidis, and H-H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10:257–281, 1993.

[20] D. Koller, J. Weber, and J. Malik. Robust Multiple Car Tracking with Occlusion Reasoning. In Springer-Verlag, editor, *European Conference on Computer Vision*, volume I, pages 189–196, Stockhol, Sweden, 1994.

[21] A. Kumar, A. Tannenbaum, and G. Balas. Optical-flow: A curve evolution approach. *IEEE Transactions on Image Processing*, 5:598–610, 1996.

[22] F. Leymarie and M. Levine. Tracking deformable objects in the plane using an active contour model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:617–634, 1993.

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

[24] R. Malladi and J. Sethian. A Real-Time Algorithm for Medical Shape Recovery. In *International Conference on Computer Vision*, pages 304–310, Bombay, India, 1998.

[25] R. Malladi, J. Sethian, and B. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:158–175, 1995.

[26] D. Metaxas and D. Terzopoulos. Constrained deformable superquadrics and nonrigid motion tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 337–343, 1991.

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

[28] N. Navab, R. Deriche, and O. Faugeras. Recovering 3D motion and structure from stereo and 2D token tracking cooperation. In *iccv*, pages 513–517, Osaka, Japan, 1990.

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

[30] 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.

[31] N. Paragios and R. Deriche. Detecting Multiple Moving Targets Using Deformable Contours. In *IEEE International Concerence on Image Processing*, pages 183–186, Santa Barbara,USA, 1997.

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

[33] N. Paragios and R. Deriche. A PDE-based Level Set approach for Detection and Tracking of moving objects. In *International Conference on Computer Vision*, pages 1139–1145, Bombay, India, 1998.

[34] N. Paragios and R. Deriche. Geodesic Active Contours for Supervised Texture Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, Colorado, USA, 1999.

[35] N. Paragios and R. Deriche. Unifying Boundary and Region-based Information for Geodesic Active Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, Colorado, USA, 1999.

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

[37] F. Pla and J. Badenas. Segmentation based on region-tracking in image sequences for traffic monitoring. In *International Conference on Pattern Recognition*, Queensland, Australia, 1998.

[38] J. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 612–617, Boston, USA, 1995.

[39] J. Shi and J. Malik. Motion segmentation and tracking using normalized cuts. In *International Conference on Computer Vision*, pages 1154–1160, Bombay, India, 1998.

[40] B. Southall, J. Marchant, T. Hague, and B. Buxton. Model based tracking for navigation and segmentation. In *European Conference on Computer Vision*, Freighburg, Germany, 1998.

[41] G. Tziritas and C. Labit. *Motion Analysis for Image Sequence Coding.* Elsevier Science, 1994.

[42] T. Vieville. Estimation of 3D-motion and structure from tracking 2D-lines in a sequence of images. In *European Conference on Computer Vision*, pages 281–292, Antibes, France, 1990.

[43] H. Yahia, G. Mazars, and J. Berroir. Segmentation and motion tracking of deformable templates with level sets characterized by particle systems. In *International Conference on Pattern Recognition*, Queensland, Australia, 1998.

[44] S. Zhu and A. Yuille. Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:884–900, 1996.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399