Geodesic Active Regions and Level Set Methods for Motion Estimation and Tracking

Nikos Paragios [a] and Rachid Deriche [b]

[a] *Ecole Nationale des Ponts et Chaussees*
6-8 Avenue Blaise Pascal, 77455 Champs sur Marne Cedex, France
mailto: nikos.paragios@computer.org

[b] *I.N.R.I.A.*
2004 Route des Lucioles
BP 93, 06902 Sophia Antipolis Cedex, France

Abstract

Motion analysis in computer vision is a well studied problem with numerous applications. In particular, the tasks of optical flow estimation and tracking are of increasing interest. In this paper, we propose a level set approach to address both aspects of motion analysis. Our approach relies on the propagation of smooth interfaces to perform tracking while using an incremental estimation of the motion models. Implicit representations are used to represent moving objects, and capture their motion parameters. Information from different sources like a boundary attraction term, a background subtraction component and a visual consistency constraint are considered. The Euler-Lagrange equations within a gradient descent method lead to a flow that deforms a set of initial curve towards the object boundaries as well an incremental robust estimator of their apparent motion. Partial extension of the proposed framework to address desne motion estimation and the case of moving observer is also presented. Promising results demonstrate the performance of the method.

Key words:
Tracking, Optical Flow, Level Set Method, Implicit Representations, Background Subtraction, Constant Brightness Constraint, Robust Estimators, Affine Motion.

1 Introduction

Recent advances on the sensors side have made the use of computer vision techniques quite attractive to a number of domains. Image and film restoration,

Preprint submitted to Elsevier Science *6 August 2004*

video-based surveillance, medical imaging, post-production & cinematography, visual inspection, etc. are application domains where tracking is of significant interest. The task of tracking consists of recovering the position and the deformations of an object of interest in an image sequence. Object boundaries are a suitable feature space for tracking.

Boundary-driven methods rely on the generation of a strength image that is often equivalent with the extraction of prominent edges and their consistent detection in time. Objects of interest are represented with parametric structures (curves) and tracking is performed by seeking the lowest potential of a cost function that exploits image characteristics along the parametric representation of the object [27,29,33]. Snakes [25], deformable templates [28] and active shape/appearance models [14] are parametric techniques that were successfully used to address tracking. Their main strength is robustness since tracking within such an approach is equivalent with the recovery of (a small number) the model parameters. On the other hand dealing with non-rigid objects, local deformations and changes of topology are issues that cannot be addressed in an efficient manner from such methods.

Geometric flows - an alternative to the snake-driven models - have the advantage of being model-free methods and therefore can deal (to some extend) with such limitations [6]. The constant brightness assumption [20,30] is also a constraint heavily considered for tracking. Objects are represented using a certain visual representation that is assumed to be constant in the temporal domain [13]. Tracking is performed through a matching process according to a similarity metric [4] on the visual representation space. To this end, given an initial region of interest (moving object), one can seek for a region across time with similar characteristics. Correlation is the simplest technique to tackle such an objective while one can find in the literature more advanced mathematical formulations that aim to recover dense motion flow [34] using more ellaborated matching criteria.

The use of parametric mechanisms like the Kalman filter [52] to describe object trajectories is a more elaborated technique that integrates prediction capabilities in the tracking process. Constraints imposed by the linearity of such a method were addressed through the consideration of more advanced prediction models like particle filters [22,53]. The prediction mechanism consists of recovering a model that could eventually describe the object displacement over time according to prior observations. Such a mechanism is used to provide an initial guess on the position of the object at the next frame, which is then optimally determined using to the characteristics of the observed image (new observations). Upon convergence of the process, the tracking result is used to improve the performance of the prediction model.

Optical flow estimation is a vital component of motion analysis. It consists of recovering the displacement of the object at the image plane from one image to the next. Tracking algorithms that are based on the assumption of global correspondence between the target features in the temporal domain implicitly address motion estimation. The use of parametric models to describe the motion of the object is the most frequent approach to implement such a condition [36]. Such a consideration refers to a fair compromise between low complexity and reasonable solution to the correspondence problem. Prior art in the domain consists of motion models of increasing complexity; rigid, similarity, affine, projective as well as quadratic models were considered in the motion estimation process. While such a method fails to account for non-planar objects, one can claim that approximate tracking could be an efficient solution in various application domains (video-based surveillance).

Medical image analysis is an example where accurate tracking is a strict requirement. Therefore the use of global parametric motion models is not a plausible solution. Anatomical structures exhibit an important degree of local deformations. Techniques focussing on the recovery of the dense motion field are suitable to cope with objects undergoing notable local deformations. Local motion estimates are equivalent with seeking for a smooth motion field that provides pixel-wise intensity correspondences for the object region. The constant brightness assumption [20,30] that valid for lambertian and non-specular surfaces is a well explored constraint to perform such an estimation. Such an approach has two important limitations; (i) the number of available constraints is lesser compared to the number of variables to be recovered, (ii) these techniques are computational expensive.

To conclude, model-based tracking techniques can be very efficient when dealing with a small degree of deformations [24] while complexity becomes an issue when objects undergo significant structural changes. The case of non-rigid objects and even further the ones that involve changes of topology are examples where the use of non-parametric (model-free) methods could be beneficial. Similar scepticism arises when considering the use of optical flow. Parametric models can cope in an efficient and robust manner with certain objects. On the other hand the complete recovery of a dense displacement field is problematic even if beneficial for the cases of non-rigid or heavily deformed objects. It is a natural conclusion that upon ideal conditions non-parametric object representations are to be combined with dense motion estimation. Such an approach though could suffer from stability due to the ill-poseness of the motion estimation problem.

In this paper, we propose the use of a model-free approach for tracking that recovers motion through parametric models. Such a framework could cope with important local deformations while being robust to the presence of noise and preserving the constant brightness assumption in a global manner. Our method represents objects using implicit representations [38]. The level set method

[38] is a well-known technique for tracking moving interfaces, a problem that arises in numerous concepts of computer vision [37]. Our approach is based on the geodesic active region model [40] a variational framework to cope with frame partition problems in computer vision. Such a concept is based on the propagation of regular curves to perform image partition under the influence of boundary-based and region-driven terms. The use of implicit representations within such a framework leads to an intrinsic, implicit and parameter free approach that can cope with topological changes.

Towards introducing a new concept while taking advantage of prior art in tracking, the proposed method consists of a boundary detection, a grouping criterion, a background subtraction and a visual consistency module. Edge-driven object detection is a standard technique to image segmentation, object extraction and tracking. Region-based object extraction aims at separating the intensity properties of the object from a static background. To this end, continuous density functions are used to describe the global appearance characteristics of the moving targets and discriminate them from the static part of scene. Background subtraction is an efficient module to change detection. In the case of static camera - an assumption initially considered in this paper - the essence of these methods is to build a model that can account for the static properties of the scene. Such a concept can be used in the form of a region-component to detect and track object boundaries from either sides [45]. The use of the constant brightness assumption for moving objects couples tracking with the motion estimation problem. To this end, we propose a visual consistency term that represents motion displacement with linear models (affine). The estimation of such models depends on the position of the object (tracking), while better estimation of the optical flow improves the position of the object to be tracked. These components are integrated within a variational level-set framework.

Prior art in the domain consists of the use of the level set method for segmentation of temporal sequences, and the estimation of the motion field [6,10,23,32,39,47,57]. In [10] dense optical flow was determined only at the object boundaries and used to perform tracking from one image to the next. Such an approach suffers from robustness since object boundaries often refer to depth discontinuities where motion estimation is most challenging. In [42,47] an approach that integrates background subtraction within edge-driven tracking is proposed. While promising results are reported, the method does not address estimation of the displacement field. In [6] a geometric flow is introduced that aims at deforming the appearance of an image region towards its corresponding in the next image. The speed of this flow is then projected to the contour space to perform tracking. Similar to [42,47] such a method does not address motion estimation. Last, but not least in [57] tracking is considered within segmentation of consecutive frames where the propagation of an implicit surface in several time instances according to some motion models is used to revocer the object projections over time.

Opposite to that, recent techniques [23,32,39] address to some extent motion estimation. In [23] global displacement is recovered through the calculation of the object's centroid, while local motion is then determined through the constant brightness assumption of the object iso-photes. In [39] motion estimation and tracking are addressed in two separate steps. Motion estimates are used within a prediction mechanism within a region-driven tracker. Last, but not least in [32] explicit motion estimation is replaced with a local search step for the best visual correspondences. The result of such step reflects the speed of the flow that performs tracking within a level set formulation. One can also refer to more recent approaches like [15,19,58].

The reminder of this paper is organized as follows; in section 2 introduce the geodesic active region model, the corresponding level set variant and some notation. Motion estimation and tracking are addressed in section 3 while some preliminary ideas for the case of moving observer are presented in section 4. Discussion is part of section 5.

2 Implicit Representations

Evolving interfaces (curves) according to some flow is a popular technique [25] to address object extraction. The flow that governs the propagation is either recovered through the minimization of an objective function [48], or defined according to the application context (geometric flows [9,31]). Snake-based segmentation approaches often refer to the propagation of curves from an initial position towards the desired image characteristics. Such flows consist of internal and external terms.

Level Set Methods

In order to introduce the level set representations, one can consider a parametric curve

$$\partial\mathcal{R}(c) : [0,1] \to \mathcal{R} \times \mathcal{R}$$

that evolves according to a given motion equation in the normal direction $\mathcal{N}$ [1] :

$$\frac{d}{d\tau}\partial\mathcal{R}(c) = \mathcal{F}(\partial\mathcal{R}(c))\ \mathcal{N}$$

where $\mathcal{F}$ is a scalar function on the local properties of the curve (e.g. curvature).

[1] The tangential component of the flow does affect only the internal parameterization of the curve.

One can implement this flow using a Lagrangian approach. The contour is represented in discrete form using a selection of control points and the curve position is updated by solving the above equations for each control point. Such technique -*in the most general case* - cannot change the topology of the evolving curve. The estimation of the local geometric characteristics of the curve within such framework is quite challenging. Consequently re-parameterization of the evolving curve is often required.

The level set method [38] - initially introduced in the area of fluid dynamics [17,18] - is an emerging technique to cope with various applications in imaging, vision and graphics [37]. Such methods rely on representing the evolving curve with the zero-level of an surface $\phi : [x, y, \phi(x, y)]$:

$$\phi(\partial\mathcal{R}(c)) = 0$$

Such representation is implicit, intrinsic and parameter free. Driven from the above condition, we can evolve the surface in such a way that the zero-level yields always to the deforming curve. Taking the derivatives of ϕ with respect to time, one can obtain the flow guiding the propagation of ϕ:

$$\frac{d}{d\tau}\phi + \mathcal{F}\ |\nabla\phi| = 0$$

Thus, we have established a connection between the family of evolving curves $\partial\mathcal{R}$ and the family of evolving surfaces ϕ. Such propagation schema can account for topological changes and provides natural support on the estimation of the local geometric properties of the curve. Techniques related with the introduction of the level set method in imaging and vision were initially reported in [9,31] and then spread across various applications [37]. Such tool was considered as an efficient numerical approximation technique to implement curve propagation according to various flows.

Geodesic Active Region

The Geodesic Active Region [48] refers to a variational framework able to deal with frame partition problems in imaging and vision. It was initially introduced in [44] for supervised texture segmentation, exploited in [43] to motion estimation and tracking and extended in [46] to deal with the task of un-supervised image segmentation. In order to facilitate the introduction of the model, the bi-modal case will be considered. To this end, the following definitions/assumptions regarding *a priori* knowledge required to introduce such model are considered;

- Let I be the input image composed of two classes (h_A, h_B),
- Let $\mathcal{P}(\mathcal{R}) = \{\mathcal{R}_A, \mathcal{R}_B = \Omega - \mathcal{R}_A\}$ be a partition of the image domain Ω into two non-overlapping regions,
- Let $\partial\mathcal{R}$ be the boundaries of this partition.

- Let us assume prior knowledge on the partition position, namely the density function p_C that measures the likelihood of a given pixel being at the boundaries,
- Let us assume prior knowledge on the expected region properties of h_A, h_B namely the densities $p_r(), p_B()$ that correspond to the conditional likelihood of a given intensity coming from the (h_A, h_B) hypotheses.

Recovering the optimal partition is equivalent with accurately extracting the boundaries between $\mathcal{R}_A$ and $\mathcal{R}_B$. This can be done using the geodesic active contour model [11,26], thus minimizing

$$E(\partial\mathcal{R}) = \int_0^1 \underbrace{g\left(\underbrace{p_C(I(\partial\mathcal{R}(c)))}_{boundary\ probability}\right)}_{boundary\ attraction} \underbrace{\left|\partial\dot{\mathcal{R}}(c)\right|}_{regularity} dc$$

where $\partial\mathcal{R}$ is a parameterization of the partition boundaries in a planar form and g is a positive, monotonicaly decreasing function with minimal values at the image locations with the desired features (high gradient).

The visual properties of the h_A, h_B hypotheses are additional cues to perform segmentation. To this end, one would like to recover a consistent frame partition between *the observed data, the associated hypotheses and their expected properties*. One can consider the *posterior* probability as criterion to derive such partition; Let $[p_S(\mathcal{P}(\mathcal{R})|I)]$ be the *posterior* partition density function with respect to $\mathcal{P}(\mathcal{R})$ given the input image I. This density function can be written according to the Bayes rules as follows:

$$p_S(\mathcal{P}(\mathcal{R})|I) = \frac{p(I|\mathcal{P}(\mathcal{R}))}{p(I)} p(\mathcal{P}(\mathcal{R}))$$

where

- $p(I|\mathcal{P}(\mathcal{R}))$ is the *posterior* segmentation probability for the image I, given the partition $\mathcal{P}(\mathcal{R})$,
- $p(\mathcal{P}(\mathcal{R}))$ is the probability of the partition $\mathcal{P}(\mathcal{R})$ among the space of all possible partitions of the image domain,
- and $p(I)$ is the probability of having as input the image I among the space of all possible images.

If we assume that all the partitions are equally probable $\left[p(\mathcal{P}(\mathcal{R})) = \frac{1}{Z}\right]$ (Z is the number of possible partitions), then one can ignore the constant terms $p(I)$, $p(\mathcal{P}(\mathcal{R}))$ and we can rewrite the density function as:

$$p_S(\mathcal{P}(\mathcal{R})|I) = \ p(I|\{\mathcal{R}_A, \mathcal{R}_B\})$$

Besides, one can further consider no correlation between the region labeling where the region probabilities depend on the observation set within the region. Such assumption can further simplify the form of the posterior probability;

$$p_S(\mathcal{P}(\mathcal{R})|I) = p([I|\mathcal{R}_A] \cap [I|\mathcal{R}_B]) = p(I|\mathcal{R}_A)\, p(I|\mathcal{R}_B)$$

where $p(I|\mathcal{R}_A)$ is the *a posterior* probability for the region $\mathcal{R}_A$ given the corresponding image intensities (*resp.* $p(I|\mathcal{R}_B)$). Last, but not least, independence on the pixel level can be considered to replace the region posterior with joint probability among the region pixels:

$$p(I|\mathcal{R}_X) = \prod_{s \in \mathcal{R}_X} p_X(I(s))$$

where $X \in \{A, B\}$. Such assumptions can lead to the following conditional (I) *posterior* partition probability for $\mathcal{P}(\mathcal{R})$;

$$p_S(\mathcal{P}(\mathcal{R})|I) = \prod_{s \in \mathcal{R}_A} p_A(I(s)) \prod_{s \in \mathcal{R}_B} p_B(I(s)).$$

Optimal grouping is equivalent with recovering the partition tha corresponds to the highest posterior. The optimization of the *posterior* probability is equivalent to the minimization of the corresponding [**-log**()] function;

$$E(\partial\mathcal{P}(\mathcal{R})) = -\underbrace{\int\int_{\mathcal{R}_A} \mathbf{log}\left[\underbrace{p_A(I(x,y))}_{h_A\ probability}\right] dxdy}_{\mathcal{R}_A\ fitting\ measurement} - \underbrace{\int\int_{\mathcal{R}_B} \mathbf{log}\left[\underbrace{p_B(I(x,y))}_{h_B\ probability}\right] dxdy}_{\mathcal{R}_B\ fitting\ measurement}$$

Such component is defined using the partition determined by the curve and aims at maximizing the *posterior* segmentation probability given the input image. It aims at separating the image regions according to their intensity properties.

The Geodesic Active Region framework consists of integrating these two different frame partition modules;

$$\begin{aligned} E(\partial\mathcal{P}(\mathcal{R})) =&(1-\alpha) \underbrace{\int_0^1 g\left(p_C(I(\partial\mathcal{R}(c))\right|\partial\dot{\mathcal{R}}(c)|dc}_{boundary\ term} \\ &- \sum_{X\in\{A,B\}} \alpha \underbrace{\int\int_{\mathcal{R}_X} \mathbf{log}\left[p_X(I(x,y))\right] dxdy}_{region\ term} \end{aligned}$$

where α is a positive constant that balances the contributions of the two terms $[0 \leq \alpha \leq 1]$.

The use of the level set function [60] as a direct optimization space for grouping was a step further towards the establishment of these techniques in imaging and vision. To this end, one can consider the distance transform $D(s, \partial\mathcal{R})$ as embedding function for $\partial\mathcal{R}$:

$$\phi(x, y) = \begin{cases} D(s, \partial\mathcal{R}) \ , & s \in \mathcal{R} \\ 0 \ , & s \in \partial\mathcal{R} \\ -D(s, \partial\mathcal{R}) \ , & s \in \Omega - \mathcal{R} \end{cases}$$

and the Dirac and Heaviside distributions:

$$\delta_\alpha(\phi) = \begin{cases} 0 & , |\phi| > \alpha \\ \frac{1}{2\alpha}\left(1 + \cos\left(\frac{\pi\phi}{\alpha}\right)\right) & , |\phi| < \alpha \end{cases}$$

$$\mathcal{H}_\alpha(\phi) = \begin{cases} 1 & , \phi > \alpha \\ 0 & , \phi < -\alpha \\ \frac{1}{2}\left(1 + \frac{\phi}{\alpha} + \frac{1}{\pi}\sin\left(\frac{\pi\phi}{\alpha}\right)\right) & , |\phi| < \alpha \end{cases}$$

and use them to introduce an image partition objective function [51,55].

The geodesic active contour [12,26] can now be defined in such a framework as

$$E_{geodesic}(\phi) = \int\int_\Omega \delta_\alpha(\phi) b(;) |\nabla\phi| \ d\Omega$$

where the arbitrary metric function $b : \mathcal{R}^+ \to [0, 1]$ is given by $b(;) = g\left(p_C(I(\partial\mathcal{R}(c)))\right)$.

The use of regional/global information modules [48] aim at separating the object from the background and can lead to adaptive balloon forces. Such criterion can be easily derived from the *Heaviside* distribution;

$$E_{regional}(\phi) = \underbrace{\int\int_\Omega \mathcal{H}_\alpha(\phi) r_O(;) \ d\Omega}_{class\ A} + \underbrace{\int\int_\Omega (1 - \mathcal{H}_\alpha(\phi)) r_B(;) \ d\Omega}_{class\ B}$$

according to the global descriptors $r_O : \mathcal{R}^+ \to [0, 1], r_B : \mathcal{R}^+ \to [0, 1]$ where according to the Geodesic Active Region model $r_A(;) = -\mathbf{log}\left[p_A(I(x, y))\right]$ and $r_B(;) = -\mathbf{log}\left[p_B(I(x, y))\right]$.

3 Optical Flow Estimation & Tracking

Tracking can be viewed in a simplistic form as object detection in static images. Therefore - without loss of generality - frameworks that perform grouping can be considered to tackle this application. Implicit representations offer the ability to deform an initial curve while being able to measure certain properties inside as well as outside the object. Therefore, their use to perform tracking as well motion estimation where optical flow measueres are updated/determined on the fly is quite prominent.

We consider the *N-Partition* case for the Geodesic Active Region model where the position of N objects is to be recovered for a given time instant t along with their motion models. Such models establish correspondences with their position at time instant $t-1$. Therefore the following notation can be adopted:

- N moving objects are visible in a sequence of T frames $[I_1, ..., I_T]$,
- N curves represented using level set functions $\phi_i \in [\phi_1, ..., \phi_N]$ are used to track these objects,
- $N \times T$ parametric motion models describe the transformation of the objects from one frame to the next; $\mathcal{A}^{i,t}$ is associated with object i and describes its 2D apparent motion between frames $t-1$ and t.

Boundary & Smoothness Component

Strong discontinuities between the moving objects and the static background - often assumed to be homogeneous - is a well explored tracking assumption. To this end, the outcome of standard edge-detection processing techniques [8,16] was used within snake-driven methods to perform tracking;

$$E_{boundary}(\phi_1, ..., \phi_N) = \sum_{i=1}^{N} \iint_{\Omega} \delta_\alpha(\phi_i) g(|\nabla I(t)|) |\nabla \phi_i| d\Omega$$

for a given frame t. One can consider replacing the attraction term with more sophisticated terms that can better account for the object boundaries [45] and go beyond the simplistic assumption of uniform background. Such objective function involves N level set functions (one for each object) to perform tracking. The calculus of variations and a gradient descent method can be used to obtained a minimum;

$$\frac{d}{d\tau}\phi_i = \delta_a(\phi) \mathrm{div} \left(g(|\nabla I(t)|) \frac{\nabla \phi_i}{|\nabla \phi_i|} \right)$$

Such a flow deforms an initial contour towards the object boundaries while being constrained by the curvature. Such flow is single-directional and reaches the object boundaries from one side. Optimal results are recovered when the initial curve is either interior to the object or encloses it. One can overcome this limitation by considering directional data terms like the ones introduced in [49,54].

Background Subtraction tracking

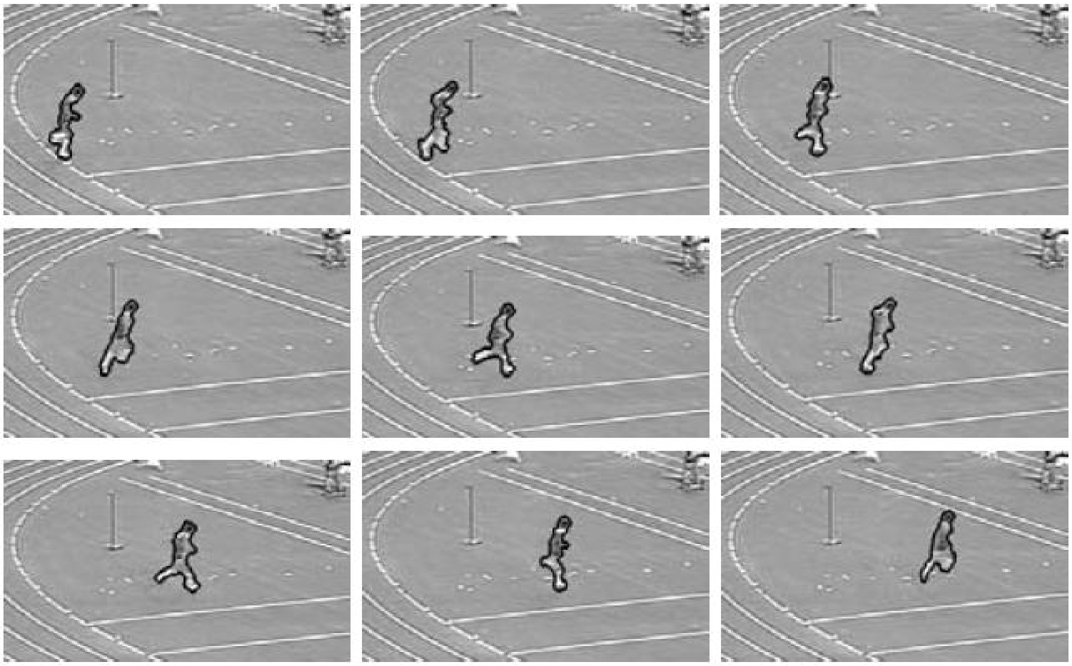

Fig. 1. Geodesic Active Tracking, Jumber Sequence using boundary attraction, background subtraction & visual grouping; results are presented in a raster scan format.

Background subtraction [35] and change detection [50] are basic components of motion analysis for static sequences. The basic assumption behind these modules is that a representation of the background can be recovered and maintained. Statistical tests can be employed to separate the pixels that belong to the moving objects from the static ones. The outcome of this process can be used then as feature space to perform tracking [47].

Global modeling of the (inter) frame-difference is a computational efficient method to detect moving objects [50] that does not require significant prior history. Static versus and non-static hypothesis can be represented using two zero-mean exponential functions. Such distributions can be derived from the empirical distribution of the difference frame; Let B be the background reference frame, and $D(t)$ the difference frame at moment t;

$$D(x, y; t) = B(x, y; t) - I(x, y; t)$$

One can assume that such distribution is a mixture model of two components [50], one that corresponds to the static hypothesis p_{st} (noise) and one that corresponds to the mobile hypothesis p_{mb}. Such assumption can lead to the following continuous form for the observed distribution:

$$p_D(d) = P_{st}\ p_{st}(d) + P_{mb}\ p_{mp}(d)$$

where P_{st}, P_{mb} are the a priori probabilities. The use of two exponential functions has been efficiently considered in the past [50]. Such simplification is valid when the static hypothesis is far more popular than the mobile one.

Otherwise, more complicated models that can account for multiple populations within the mobile component are to be considered. Maximum Likelihood principles can be used to recover the parameters of this mixture model. The analysis and modeling of the difference frame provides a fast and reliable way to perform background subtraction through the densities of the two conflicting classes (p_{st}, p_{mp}). One can expect that the moving objects are composed of mobile pixels. Therefore, tracking is equivalent with grouping pixels that do not refer to the background hypothesis. Furthermore, the density of the static hypothesis can be used to define a grouping metric for the background pixels.

Binary decisions from the background subtraction process can cause non-reversible errors in the tracking process. Decisions taken using Bayes rule compare the probabilities of the two conflicting hypotheses given the input image. Cases where both hypotheses do equally well or bad can produce important errors through a binary classification module. One can overcome this limitation by introducing a background subtraction tracking term in the form of a continuous region-defined component [45];

$$E_{detection}(\phi_1, ..., \phi_N) =$$
$$\underbrace{-\sum_{i=1}^{N} \iint_{\Omega} \mathcal{H}_\alpha(\phi_i) \log(p_{mb}(I(t)) d\Omega}_{moving\ objects} \underbrace{- \iint_{\Omega} \left[\prod_{i=1}^{N} (1 - \mathcal{H}_\alpha(\phi_i))\right] \log(p_{st}(I(t)) d\Omega}_{static\ background}$$

This terms assumes that moving objects have different visual properties compared to the static background. In areas where such objects are present, the pdf for the object (mobile) hypothesis is much stronger than the static one. Similar interpretation is valid for the background subtraction component. One can replace the $-\log$ with more appropriate terms that exhibit stable behavior. A gradient descent method is a naturally way to recover a solution that minimizes the previously defined objective function;

$$\frac{d}{d\tau}\phi_i = \delta_\alpha(\phi_i) \log \left(\frac{p_{st}(I(t)}{p_{mb}(I(t))} \right)$$

where we assume the absence of occlusions for the moving objects. The interpretation of such flow is quite clear; the evolving curve shrinks when located on the background and expands otherwise (inside a moving object). One can consider this term as an adaptive balloon force. In [45], boundary and background subtraction components were combined to perform tracking [Fig. (1,2)] with encouraging results. In the absence of background model, similar analysis on the inter-frame difference frame can be used to separate the static from the mobile image components.

Visual Consistency

Visual consistency through motion recovery for the moving objects in the

Fig. 2. Geodesic Active Tracking, Oxford Sequence using boundary attraction, background subtraction & visual grouping; results are presented in a raster scan format.

temporal domain is a standard constraint to perform tracking. Most of the 2-D motion estimation approaches are based on the measurement of the apparent motion of the intensity patterns over time [1,3]. Such methods assume that the image brightness along the motion trajectory is constant [20]. However, changes on the object pose, global/local illumination conditions, etc. violate the brightness constancy constraint, a core assumption during flow recovery. Furthermore, the motion vectors satisfying the image brightness constraint are not unique and external factors like surface reflections properties, sensor noise and distortions can cause changes not related with motion.

Motion can be determined either using global motion models or by considering correspondences pixel-wise. Global motion models assume the existence of a valid transformation for the entire object. Opposite to that, local motion (optical flow) is estimated independently pixel-wise. Robustness is the main advantage of global motion components while their inability to deal with local deformations is a strong limitation. One can claim that for a sufficiently small field of view and planar moving objects, the image velocity field (projection of the real 3D motion) can be approximated by a linear model $A(x, y) = (A_x(x, y), A_y(x, y))$ while in the absence of motion (static background), the brightenss remains constant in time;

$$\begin{cases} I(x, y; t) = I(x, y; t+1); \text{ background} \\ I(x, y; t) = I((x, y) + A(x, y); t+1); \text{ planar object} \end{cases}$$

Parametric motion estimation [5,7,34,56] is a computationally efficient method to recover optical flow. Low complexity and robust estimates are its main strengths. One can claim that when the assumptions imposed by the model are satisfied by the moving object, efficient motion estimates can be recovered. On the other hand, such approaches cannot deal with local object deformations or objects that exhibit depth discontinuities (non-planar). One can consider transformations that involve limited number of parameters like rigid, or more complicated ones that can account for more complex scenes and motions. In latest case the use of affine, homographic or quadratic models can be used to approximate the motion of the target. We consider affine transformations [43], a compromise between low complexity, stability and fairly good approximation of the motion field. Such model consists of six motion parameters:

$$\mathcal{A}(x,y) = \begin{bmatrix} \mathcal{A}^x(x,y) \\ \mathcal{A}^y(x,y) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}$$

The most common way to derive motion estimates is the optical flow constraint [20]. In the absence of global illumination changes, one can assume that the observed intensities at the position (x, y) in the frame $t+1$, and at $(x, y) + A(x, y)$ in the frame t are the same. This constraint relies on minimizing the *sum of squared differences* (SSD)

$$E(\mathcal{A}) = \iint_\Omega (I(\mathcal{A}; t+1) - I(t))^2 d\Omega$$

Global (affine) motion models introduce certain limitations in the estimation process. Local deformations as well as depth changes do not satisfy the brightness constraint and can perturb the estimation of the motion parameters. At the same time, least estimators (sum of square differences) are sensitive to the presence of noise and outliers [21]. A simple way to deal with such limitation is to consider local deformations as outliers of the estimation process and ignore them during the estimation of the motion parameters. Towards this direction within the considered application one would like recover N parametric motion models $\mathcal{A}_i$ that create visual correspondences for each object in the temporal domain for the moving targets. Furthermore one can assume the absence of motion for the static background, resulting on the following functional:

$$E_{motion}((\phi_1, \mathcal{A}_1), ..., (\phi_N, \mathcal{A}_N)) = \underbrace{\sum_{i=1}^{N} \iint_\Omega \mathcal{H}_\alpha(\phi_i) \rho\left(I(\mathcal{A}_i; t+1) - I(t)\right) d\Omega}_{moving\ object\ visual\ consistency}$$
$$+ \underbrace{\iint_\Omega \left[\prod_{i=1}^{N} (1 - \mathcal{H}_\alpha(\phi_i))\right] \rho\left(I(t+1) - I(t)\right) d\Omega}_{background\ visual\ consistency}$$

where ρ is a bounded error function. We consider the *fair* error estimator given

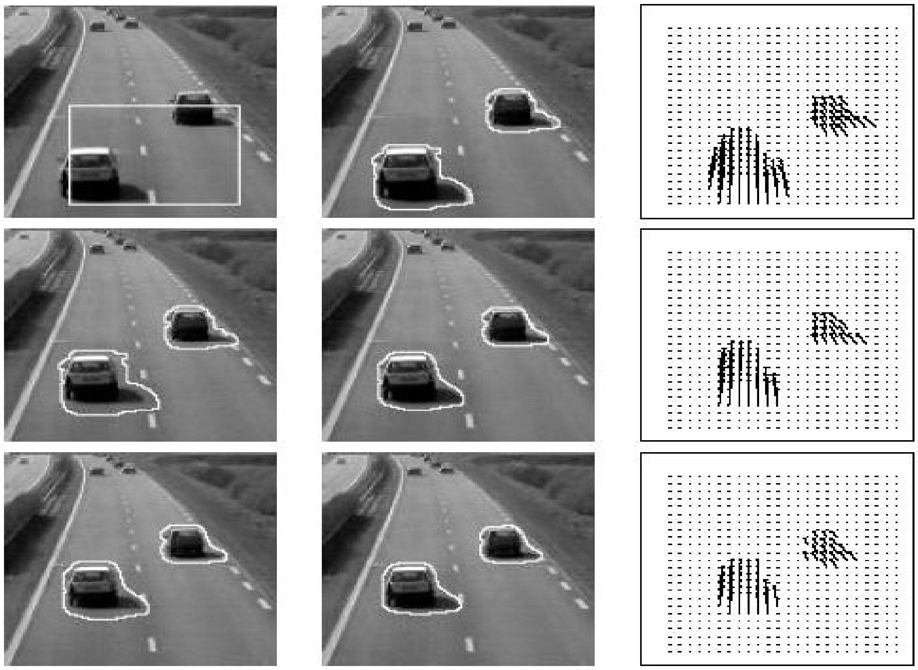

Fig. 3. Motion Estimation & Tracking, Highway Sequence (part 1); (first column) image t & tracking result at t, (second column) image $t+1$ & tracking result at $t+1$, & (third row) optical flow estimates.

by

$$\rho_{fair}(r) = c^2 \left[\frac{|r|}{c} - \log\left(1 + \frac{|r|}{c}\right) \right]$$

Such objective function couples motion estimation and tracking. The unknown variables are the motion parameters and the targets position at frame $t+1$. This functional implicitly assumes the absence of occlusions between the moving objects. In order to interpret the proposed term, we will consider the motion transformation known. In that case, the lowest potential of the objective function $E_{motion}((\phi_i, \mathcal{A}_1), ..., (\phi_N, \mathcal{A}_N))$ refers to an image region composed of pixels that satisfy the visual constancy constraint with the target position in the previous frame. On the other hand, for known objects positions, the lowest potential of the objective function corresponds to an optimal motion transformation $\mathcal{A}_i$ that creates pixel-wise visual correspondences for the target in the temporal domain.

One can optimize this functional with respect to the targets positions (ϕ_i) and the optical flow estimates ($\mathcal{A}_i$) using a gradient decent method;

$$\begin{aligned}
\frac{d}{d\tau}\phi_i =& \delta_\alpha(\phi_i) \left(\rho\left(I(t+1) - I(t)\right) - \rho\left(I(\mathcal{A}_i; t+1) - I(t)\right)\right) \\
\frac{d}{d\tau}a^i_{kl} =& \iint_\Omega \mathcal{H}_\alpha(\phi_i)\psi(I(\mathcal{A}_i; t+1) - I(t)) \\
& \qquad \left(\frac{\partial I(\mathcal{A}_i; t+1)}{\partial x}, \frac{\partial I(\mathcal{A}_i; t+1)}{\partial y}\right)\left(\frac{\partial \mathcal{A}^x_i}{\partial a^i_{kl}}, \frac{\partial \mathcal{A}^y_i}{\partial a^i_{kl}}\right)
\end{aligned}$$

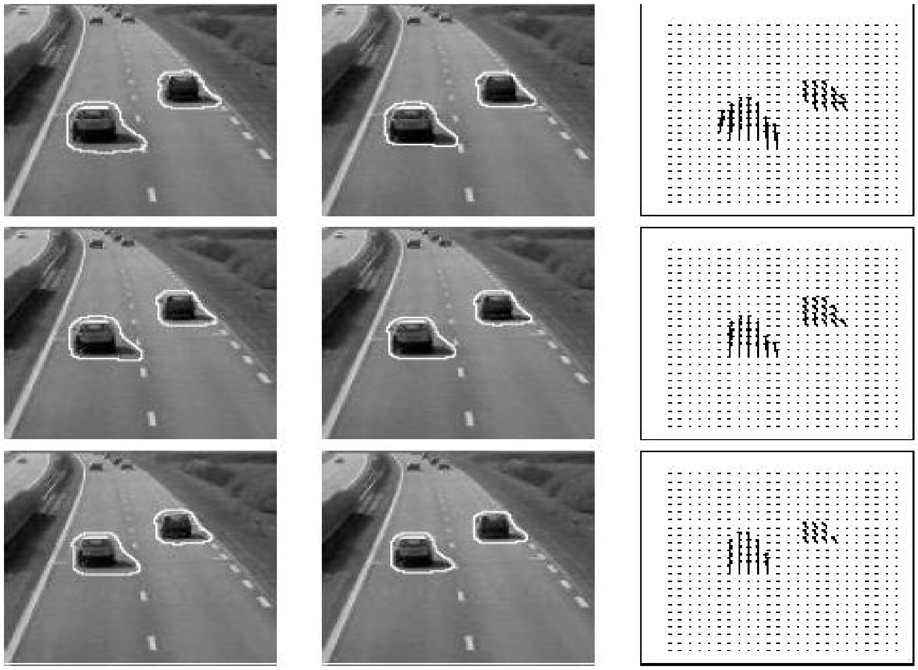

Fig. 4. Motion Estimation & Tracking, Highway Sequence (part 2); (first column) image t & tracking result at t, (second column) image $t+1$ & tracking result at $t+1$, & (third row) optical flow estimates.

where $\phi(r) = \rho'(r)$ is the derivative of ρ known as influence function and a_{kl} is the (k, l) parameter of the motion model $[\mathcal{A}_i]$. One can interpret the obtained motion equation as follows;

- i **Level Set Flow**: a force that aims to move the i curve towards the direction that decreases the visual correspondence error. A relative comparison between the background and the motion hypotheses is used to determine the propagation direction. If the error produced by the static hypothesis is greater than the one of the object (given the current estimation of the motion), then the contour expands to include this pixel in the object hypothesis and vice-versa.
- i **Motion Estimation Flow**: an iterative mechanism to update the i motion estimates given the current position of the object. Such updates are driven by a term that tends to improve the quality of visual correspondences between the current and the previous frame position of the objects.

Tracking and motion estimation can be jointly recovered in an iterative manner. However, one can claim that the use of a gradient descent method to estimate the motion parameters is not the most prominent solution. Such parameters have different rate of update and the use of the same time step can cause discrepancies on the estimation. Eventually, the system can become unstable due to the different convergence rates of the motion model components. Such limitation can be addressed by updating the motion estimates using a closed form solution [43]. To this end, the motion estimation task is reformu-

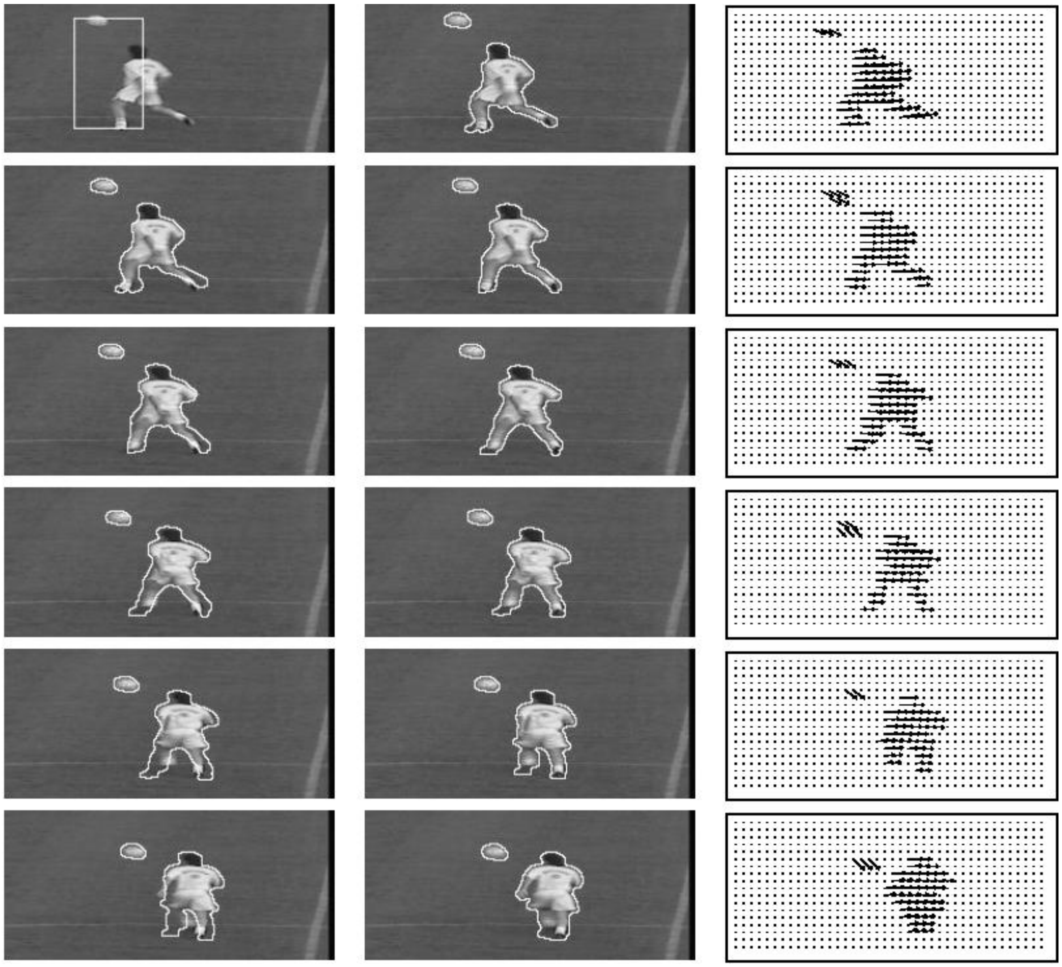

Fig. 5. Motion Estimation & Tracking, Footballeur Sequence; (first column) image t & tracking result at t, (second column) image $t+1$ & tracking result at $t+1$, & (third row) optical flow estimates.

lated as follows; given a current estimate of the motion model $\mathcal{A}$, recover a complementary affine model

$$\Delta\mathcal{A}(x,y) = \begin{bmatrix} \Delta\mathcal{A}^x(x,y) \\ \Delta\mathcal{A}^y(x,y) \end{bmatrix} = \begin{bmatrix} \delta a_{11} & \delta a_{12} \\ \delta a_{21} & \delta a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \delta a_{13} \\ \delta a_{23} \end{bmatrix}$$

that when combined with the current estimates improves the matching between the object positions in two consecutive frames. Such an approach is demonstrated in [Fig. (7)]. Using the notation introduced earlier, one can redefine the motion-related component of the objective function;

$$E_{motion}((\phi_i, \Delta\mathcal{A}_1), ..., (\phi_N, \Delta\mathcal{A}_N)) = \iint_\Omega \left[\prod_{i=1}^{N}(1 - \mathcal{H}_\alpha(\phi_i))\right] \rho\left(I(t+1) - I(t)\right) d\Omega$$
$$+ \sum_{i=1}^{N} \iint_\Omega \mathcal{H}_\alpha(\phi_i) \rho\left(I(\mathcal{A}_i + \Delta\mathcal{A}_i; t+1) - I(t)\right) d\Omega$$

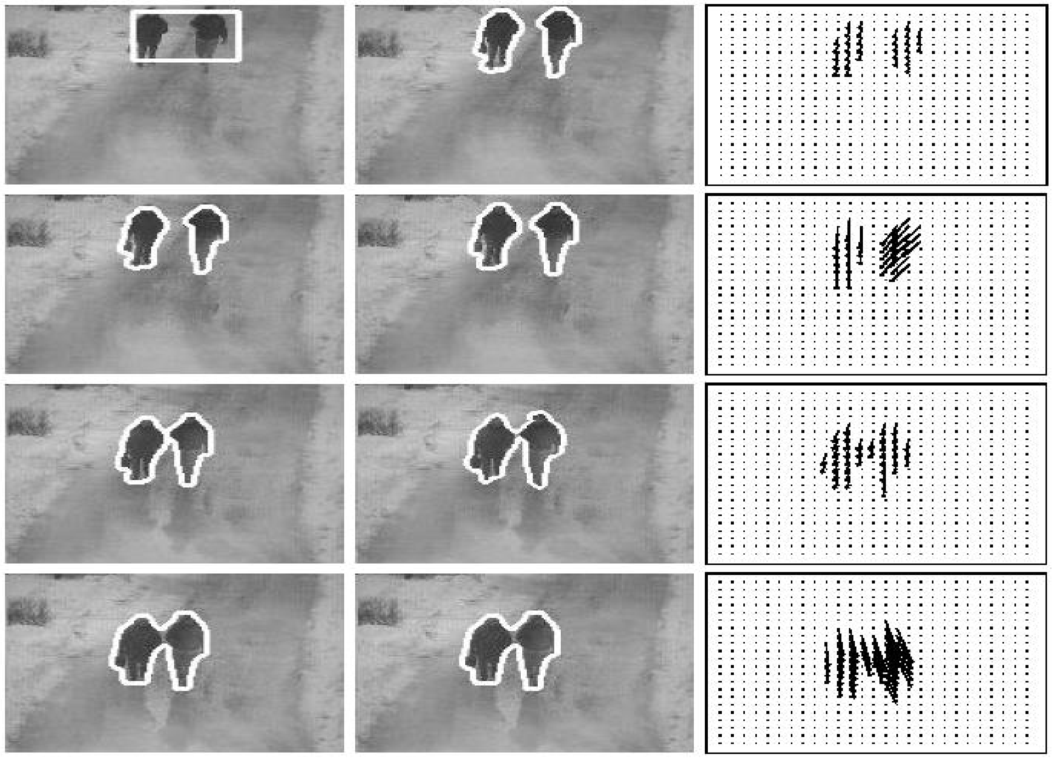

Fig. 6. Motion Estimation & Tracking, Swedish Pedestrians Sequence; (first column) image t & tracking result at t, (second column) image $t+1$ & tracking result at $t+1$, & (third row) optical flow estimates.

The optimal solution of such objective function with respect to $\Delta\mathcal{A}_i$ can now be recovered using the following constraints:

$$n \in [1,2],\ m \in [1,3]$$
$$\frac{\partial}{\partial \delta a^i_{mn}} E_{motion}((\phi_i, \Delta\mathcal{A}_1), ..., (\phi_N, \Delta\mathcal{A}_N)) = 0$$

leading to a linear system with respect to $[\delta a_{11}, \delta a_{12}, \delta a_{13}]$ and $[\delta a_{21}, \delta a_{22}, \delta a_{23}]$ that has a solution in a close form. We perform this motion estimation step until the motion model converges. Such mechanism can be used to recover the optimal estimates of the motion model according to the latest position of the object as defined from the tracking module (ϕ_i). One can consider performing such motion correction step after each iteration. Motion correction is required when the object position (tracking) changes significantly. The use of the proposed tracking framework will gradualy update the object position from one iteration to the next until tracking is optimized. Therefore, updating the motion model as shown in [Fig. (7)] in each iteration is not necessary [43].

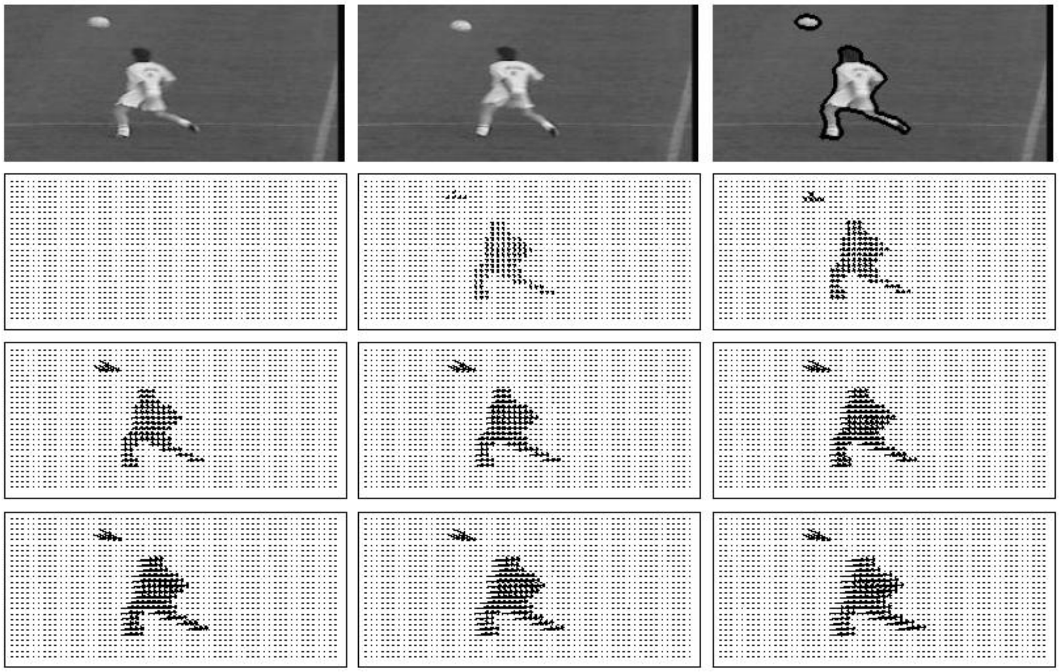

Fig. 7. Incremental Motion Estimation & Tracking, Footballeur Sequence.

4 Complete Recovery of the Apparent Motion & Mobile Observer

Complete recovery of the aparent motion is a most prominent solution when dealing with non-rigid objects. In that case, we assume that existence of a $(U(x,y),V(x,y)) = (u,v)$ field in the image plane as follows;

$$\begin{cases} (x,y) \in \Omega : (U(x,y),V(x,y)) \\ I(x,y;t) = I(u(x,y),v(x,y);t+1) \end{cases}$$

The constant brighness assumption can be considered to define an objective function that can recover the vector field (u,v) in the pixel level;

$$E(U,V) = \iint_\Omega \rho(I((u,v);t+1) - I(t))d\Omega$$

However the recovery of the optical flow using the above constraint is an ill-posed problem. The number of unknown variables is larger than the number of constraints. A common technique to overcome this limitation is to consider additional smoothness constraints on the flow;

$$E(U,V) = \iint_\Omega \rho(I((u,v);t+1) - I(t)) + \epsilon\zeta\left(|U| + |V|\right)d\Omega$$

where ζ is a regularization term that penalizes discontinuities on the optical flow and ϵ a positive constant that balances importance of the two terms.

This framework can be used to recover dense optical flow and perform tracking.

Opposite to the affine case, theoretically such functional can deal with the case of moving camera as well. However, such simplistic estimation component may fail to deal with complex 3D scenes and do not account for discontinuities on the optical flow that are quite natural along the object boundaries.

Mobile Observer

One can further explore this framework to address the case of mobile observer. Within such a scenario, the objective would be to separate the additive motion due to observer displacement from the motion of the non-static objects of the scene. While the real 3D motion of the observer can be determined in a unique manner, its projection to the image plane introduces certain difficulties. Depth discontinuities as well as depth layers would reflect to different motion observation in the image plane.

The use of unique parametric models to describe the observer motion is not an adequate solution. However, one can assume that such a model could be used to determine the observed motion at each layer. To this end, - *under the assumption that image has been partitioned into layers that refer to constant depth* - one can consider an affine model to describe their apparent motion. Under such consideration, non-static objects of the scene will appear in the form of motion layers.

Motion segmentation [2,5,56] was studied in a rigorous manner during the past decade. The objective of our approach is to extend the motion estimation and tracking framework to perform motion separation. Without loss of generality one can assume that the number of layers is known. Such a condition can be met by fitting parametric motion models to the dense motion field and then performing a clustering step in the space of their parameters. Such clustering step can also be used to determine the initial motion parameters of each layer.

The outcome of such a procedure will be the number of depth layers that then are to be recovered along with their motion parameters:

$$\begin{aligned} E_{motion}((\phi_i, \Delta\mathcal{A}_1), ..., (\phi_N, \Delta\mathcal{A}_N)) = \\ \sum_{i=1}^{N} \iint_{\Omega} \mathcal{H}_\alpha(\phi_i)\rho\left(I(\mathcal{A}_i + \Delta\mathcal{A}_i; t+1) - I(t)\right)) \, d\Omega \end{aligned}$$

where the notion of static background has been eliminated, as well as to some extend the notion of moving objects. Such a term is comparable with the one earlier presented for the case of multiple objects. Opposite to the case of static observer, the term of background subtraction cannot be used (unless a motion compensation step is first applied). On the other hand one can consider separating the regions according to their intensity properties that is equivalent

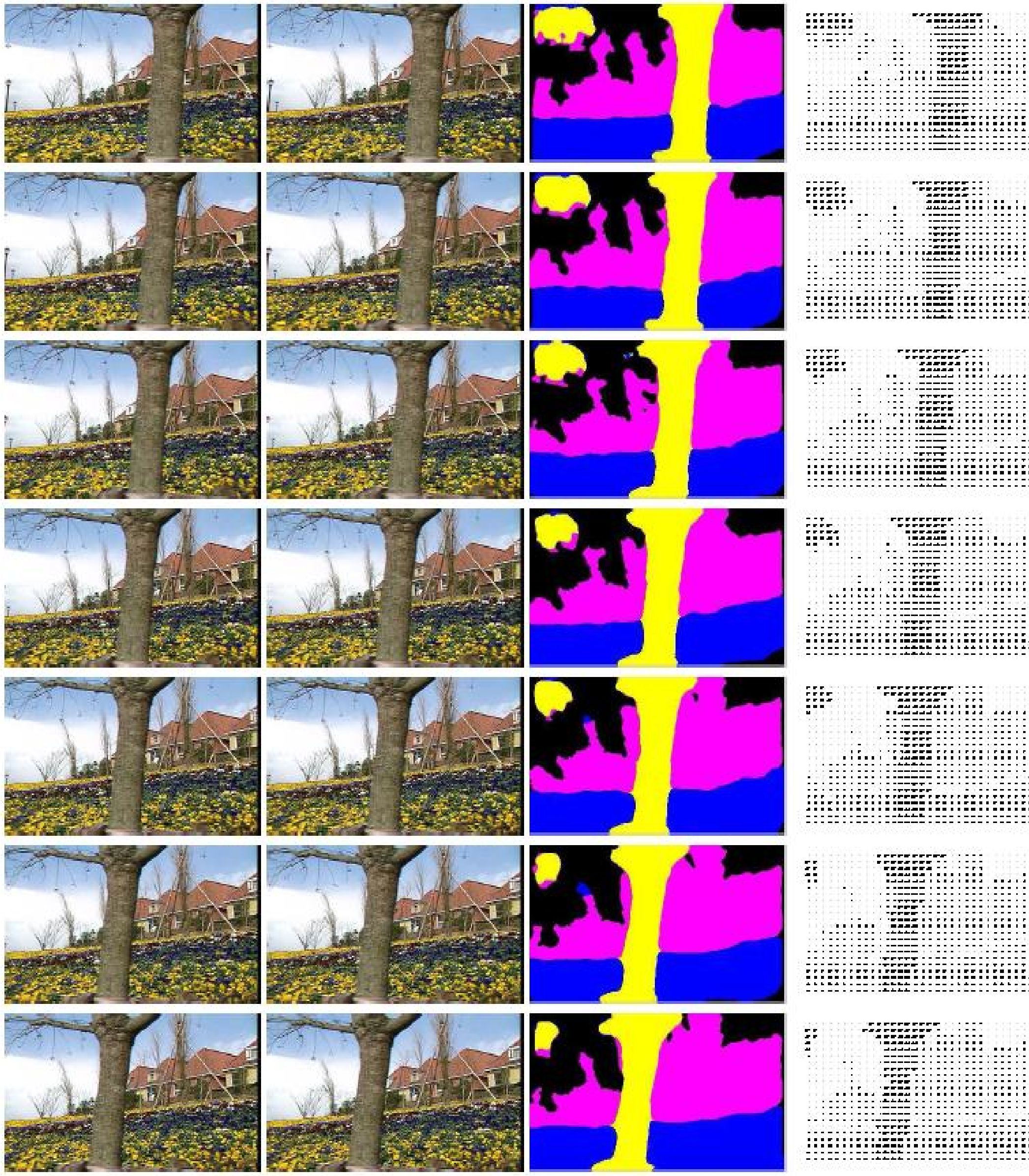

Fig. 8. Implicit Representations for Recovery and Reconstruction of Motion Layers; (first column) image t, (second column) image $t+1$, (third column) segmented layers & (fourth row) motion estimates.

with

$$E_{segmentation}(\phi_1, ..., \phi_N) = -\sum_{i=1}^{N} \iint_{\Omega} \mathcal{H}_{\alpha}(\phi_i) \log\left(p_i(I)\right) \, d\Omega,$$

where $I = I(;\tau)$ and $[p_i, i = 1, \cdots, N]$ are the non-parametric approximations of the intensity distribution of the different motion layers. The integral of the objective function measures the quality of fitting between the actual observations and the expected properties of each motion layer. More details on this component can be found at [59] while some preliminary results are reported in [Fig. (8)].

5 Discussion

In this paper a variational formulation to deal with tracking and motion estimation was reported as well as potential extensions of the methods for motion reconstruction in layers. The base of our approach was the Geodesic Active Region model. Several visual cues were integrated within an objective function that separates moving objects from the static parts of the scene, and tracks them in consecutive frames. To this end, we have proposed an edge-driven tracking module, a change detection background/foreground separation component and a visual consistency term that couples motion with tracking.

Implementation Issues

The proposed framework can be used to detect, track and recover the trajectories of rigid as well as non-rigid objects. The number of moving objects can be determined either by the user or by processing the first frame [2]. To this end, the background subtraction module can be used. Upon convergence of the background subtraction flow, a connected component analysis method can determine the number of moving objects in the scene. Then, a level set function can describe the motion of each object that is a computational intensive procedure. An elegant way to reduce complexity without reducing the model capacity is to couple the level set functions as proposed in [55] where N level set functions can be used to detect and track 2^N objects.

Promising experimental results demonstrate the potentials of such selection for different outdoor sequences with respect to the motion estimation [Fig. (3,4,5,6)] task and the tracking [Fig. (1,2,3,4,5,6)]. The incremental estimation of the apparent motion is a promising solution to the optical flow recovery problem. However, it suffers from being a global method that considers a linear model to recover the object motion. The use of robust estimators will lead to reasonable handling of the non-rigid parts. Hopefully within the estimation process such parts will be considered as outliers and will not perturb the motion estimates for the rigid part of the object. On the other hand, such errors will propagate to the tracking component of our technique and may cause certain discrepancies.

Future Directions

Dense motion recovery that also accounts for the uncertainties of the estimation process is an interesting extension of the proposed framework. The outcome of such a procedure will be the use optical flow in a qualitative manner within tracking. Use of prior knowledge on the geometric form of the objects

[2] Similar process has to be considered to deal with objects that appear in the camera view in some later time.

to be tracked is also a natural extension of our approach [41] with numerous applications. Last, but not least one can consider similar formulation to perform stereo reconstruction, or recovering the disparities between two different views.

Acknowledgements: The first author would like to thank Rong Zhang and Dimitris Metaxas from Rutgers University for fruitful discussions during the partial extension of the proposed framework to deal with the case of a mobile observer that is joint work currently under development.

References

[1] J. Aggarwal and N. Nandhakumar. On the computation of motion from sequences of images - A Review. *Procedings of the IEEE*, pages 917–935, 1988.

[2] S. Ayer and H. Sawhney. Layered Representation of Motion Video Using Robust Maximum-Likelihood Estimation of Mixture Models and MDL Encoding. In *IEEE International Conference in Computer Vision*, pages 777–784, Caibridge, USA, 1995.

[3] J. Barron, D. Fleet, and S. Beauchemin. Performance of Optical Flow Techniques. *International Journal of Computer Vision*, 12:43–77, 1994.

[4] B. Bascle and R. Deriche. Region Tracking through Image Sequences. In *IEEE International Conference in Computer Vision*, pages 302–307, 1995.

[5] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical Model-Based Motion Estimation. In *European Conference on Computer Vision*, pages 237–252, 1992.

[6] M. Bertalmio, G. Sapiro, and G. Randall. Morphing Active Contours. In *International Conference on Scale-Space Theories in Computer Vision*, pages 46–57, 1999.

[7] M. Black and A. Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:973–986, 1996.

[8] J. Canny. A computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:769–798, 1986.

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

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

[11] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. In *IEEE International Conference in Computer Vision*, pages 694–699, 1995.

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

[13] D. Comaniciu, V. Ramesh, and P. Meer. Real time tracking of non-rigid objects using Mean Shift. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 142–149, 2000.

[14] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.

[15] D. Cremers and S. Soatto. Motion Competition: A Variational Framework for Piecewise Parametric Motion Segmentation. *International Journal of Computer Vision*. to appear.

[16] R. Deriche. Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector. *International Journal of Computer Vision*, 1:167–187, 1987.

[17] A. Dervieux and F. Thomasset. A finite element method for the simulation of rayleigh-taylor instability. *Lecture Notes in Mathematics*, 771:145–159, 1979.

[18] A. Dervieux and F. Thomasset. Multifluid incompressible flows by a finite element method. In W. Reynolds and R.W. MacCormack, editors, *Seventh International Conference on Numerical Methods in Fluid Dynamics*, volume 141 of *Lecture Notes in Physics*, pages 158–163, June 1980.

[19] I. Dydenko, D. Friboulet, and I. Magnin. A Variational Framework for Affine Registration and Segmentation with Shape Prior: Application in Echocardiographic Imaging. In *IEEE Workshop in Variational and Level Set Methods*, pages 209–217, 2003. Faugeras, O. and Paragios, N. (eds).

[20] B. Horn and B. Schunck. Determinating Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.

[21] P. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

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

[23] S. Jehan-Besson, M. Barlaud, and G. Aubert. Video Object Segmentation Using Eulerian Region-Based Active Contours. In *IEEE International Conference in Computer Vision*, volume I, pages 353–360, 2001.

[24] 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.

[25] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. In *IEEE International Conference in Computer Vision*, pages 261–268, 1987.

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

[27] 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.

[28] P. Lipson, A. Yuille, D. Okeefe, J. Cavanaugh, J. Taaffe, and D. Rosenthal. Deformable Templates for feature extraction from Medical Images. In *European Conference on Computer Vision*, pages 413–417, 1990.

[29] D. Lowe. Robust Model-based Motion Tracking through the Integration of Search and Estimation. *International Journal of Computer Vision*, 8:113–122, 1992.

[30] B. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[31] R. Malladi, J. Sethian, and B. Vemuri. Evolutionary fronts for topology independent shape modeling and recovery. In *European Conference on Computer Vision*, pages 1–13, 1994.

[32] A. Mansouri. Region Tracking via Level Set PDEs without Motion Computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:947–961, 2002.

[33] 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.

[34] F. Meyer and P. Bouthemy. Region-Based Tracking Using Affine Motion Models in Long Image Sequences. *CVGIP: Image Understanding*, 60:119–140, 1994.

[35] A. Mittal and D. Huttenlocher. Scene Modeling for Wide Area Surveillance and Image Synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2160–2167, 2000.

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

[37] S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer Verlag, 2003.

[38] 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.

[39] C. Papin, P. Bouthemy, E. Mémin, and G. Rochard. Tracking and characterization of highly deformable cloud structures. In *European Conference on Computer Vision*, pages II: 428–442, 2000.

[40] N. Paragios. *Geodesic Active Regions and Level Set Methods: Contributions and Applications in Artificial Vision*. PhD thesis, I.N.R.I.A./University of Nice-Sophia Antipolis, 2000. http://www.inria.fr/RRRT/TU-0636.html.

[41] N. Paragios. Shape-based Segmentation and Tracking in Cardiac Image Analysis. *IEEE Transactions on Medical Imaging*, pages 773– 776, 2003.

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

[43] N. Paragios and R. Deriche. Geodesic Active Regions for Motion Estimation and Tracking. In *IEEE International Conference in Computer Vision*, pages 688–674, 1999.

[44] N. Paragios and R. Deriche. Geodesic Active Regions for Supervised Texture Segmentation. In *IEEE International Conference in Computer Vision*, pages 926–932, 1999. Previous: INRIA Research Report, RR 3440, June 1998, http://www.inria.fr/RRRT/RR-3440.html.

[45] N. Paragios and R. Deriche. Unifying Boundary and Region-based Information for Geodesic Active Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages II:300–305, 1999.

[46] N. Paragios and R. Deriche. Coupled Geodesic Active Regions for Image Segmentation: A Level Set Approach. In *European Conference in Computer Vision*, pages II:224–240, 2000. Previous: INRIA Research Report, RR 3783, October 1999, http://www.inria.fr/RRRT/RR-3783.html.

[47] N. Paragios and R. Deriche. Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:266–280, 2000.

[48] N. Paragios and R. Deriche. Geodesic Active Regions: A New Framework to Deal with Frame Partition Problems in Computer Vision. *Journal of Visual Communication and Image Representation*, 13:249–268, 2002.

[49] N. Paragios, O. Mellina-Gottardo, and V. Ramesh. Gradient Vector Flow Fast Geodesic Active Contours. In *IEEE International Conference in Computer Vision*, pages I:67–73, 2001.

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

[51] C. Samson, L. Blanc-Feraud, G. Aubert, and J. Zerubia. A Level Set Model for Image Classification. In *International Conference on Scale-Space Theories in Computer Vision*, pages 306–317, 1999.

[52] D. Terzopoulos and R. Szeliski. Tracking with Kalman Snakes. In A. Blake and A. Yuille, editors, *Active Vision*, pages 3–20. MIT Press, 1992.

[53] K. Toyama and A. Blake. Probabilistic Tracking in a Metric Space. In *IEEE International Conference in Computer Vision*, pages 50–59, 2001.

[54] A. Vasilevskiy and K. Siddiqi. Flux Maximizing Geometric Flows. In *IEEE International Conference in Computer Vision*, pages I: 149–154, 2001.

[55] L. Vese and T. Chan. A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. *International Journal of Computer Vision*, 50:271–293, 2002.

[56] J. Wang and E. Adelson. Representing Moving Images with Layers. *IEEE Transactions on Image Processing*, 3:625–638, 1994.

[57] A. Yezzi, L. Zollei, and T. Kapur. A Variational Framework for Joint Segmentation and Registration. In *IEEE Mathematical Methods in Biomedical Image Analysis*, pages 44–51, 2001.

[58] A. Yilmaz, X. Li, and B. Shah. Contour based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, November 2004.

[59] R. Zhang, N. Paragios, and D. Metaxas. Implicit Representations for recovery and reconstruction of motion layers. In *IEEE Workshop in Variational and Level Set Methods*, pages 209–217, 2003.

[60] H-K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational Level Set Approach to Multiphase Motion. *Journal of Computational Physics*, 127:179–195, 1996.