

Bayesian Approaches to Motion-Based Image and Video Segmentation

Daniel Cremers*

Department of Computer Science
University of California, Los Angeles, USA
<http://www.cs.ucla.edu/~cremers>

Abstract. We present a variational approach for segmenting the image plane into regions of piecewise parametric motion given two or more frames from an image sequence. Our model is based on a conditional probability for the spatio-temporal image gradient, given a particular velocity model, and on a geometric prior on the estimated motion field favoring motion boundaries of minimal length.

We cast the problem of motion segmentation as one of Bayesian inference, we derive a cost functional which depends on parametric motion models for each of a set of domains and on the boundary separating them. The resulting functional can be interpreted as an extension of the Mumford-Shah functional from intensity segmentation to motion segmentation. In contrast to most alternative approaches, the problems of segmentation and motion estimation are jointly solved by continuous minimization of a *single* functional. Minimization results in an eigenvalue problem for the motion parameters and in a gradient descent evolution for the motion boundary. The evolution of the motion boundaries is implemented by a multiphase level set formulation which allows for the segmentation of an arbitrary number of multiply connected moving objects.

We further extend this approach to the segmentation of space-time volumes of coherent motion from video sequences. To this end, motion boundaries are represented by a set of surfaces in space-time. An implementation by a higher-dimensional multiphase level set model allows the evolving surfaces to undergo topological changes. In contrast to an iterative segmentation of consecutive frame pairs, a constraint on the area of these surfaces leads to an additional temporal regularization of the computed motion boundaries.

Numerical results demonstrate the capacity of our approach to segment objects based exclusively on their relative motion.

1 Introduction

The segmentation of images into meaningful areas can be driven by various low-level grouping criteria, such as edge information, color information or texture information. In the present work, we address the question of how to exploit motion information for the purpose of segmentation.

While traditionally researchers have suggested to first estimate a motion field and to subsequently segment the scene based on this motion field [35], the problem of motion

* Present Affiliation: Department of Computer Science, University of Bonn, Germany.

segmentation can be viewed as a chicken and egg problem: Reliable motion estimation algorithms generally require a region of support (ideally given by a segmentation of the moving object), while the computation of a segmentation assumes knowledge of the motion.

Many researchers have addressed this coupling of segmentation and motion estimation. Some have proposed to model motion discontinuities implicitly by non-quadratic robust estimators [2, 4, 21, 22, 25, 36]. Others tackled the problem of segmenting motion by treating the problems of motion estimation in disjoint sets and optimization of the motion boundaries separately [5, 14, 27, 29, 30]. Some approaches are based on Markov Random Field (MRF) formulations and optimization schemes such as stochastic relaxation by Gibbs sampling [20], split-and-merge techniques [15], deterministic relaxation [3], graph cuts [31] or expectation maximization (EM) (cf. [18, 37]). As pointed out in [37], exact solutions to the EM algorithm are computationally expensive and therefore suboptimal approximations are employed. An elegant method to directly compute both the segmentation and the motion models was recently proposed in [34], yet this approach differs from the above approaches in that it does not allow to impose smoothness of the estimated motion boundaries.

In the present paper, we propose a framework which allows to jointly solve the problems of segmentation and motion estimation by minimizing a single functional. We formulate the problem of motion segmentation in the framework of Bayesian inference. Related Bayesian formulations have been proposed in the discrete MRF framework (cf. [3]). Our formulation differs from the above approach in that it is continuous, uses a contour representation of the motion discontinuity set, can be optimized by a simple and fast gradient descent minimization and is based on a different (normalized) likelihood in the data term. The proposed functional can be interpreted as an extension of the Mumford-Shah model [24] from the case of gray value segmentation to the case of motion segmentation. Minimization leads to an eigenvalue problem for the motion parameters associated with each region, and to a gradient descent evolution for the boundary separating the regions.

This joint minimization of a *single* functional with respect to motion parameters and motion boundaries generates a pde-based solution to the above chicken and egg problem. The resulting boundary evolution can be interpreted in the way that neighboring regions compete for the boundary in terms of their motion energy. In analogy to the corresponding gray value model, which has been termed *Region Competition* [39], we therefore refer to this process as *Motion Competition*.

We propose a multiphase level set implementation of the motion competition functional, which is based on the corresponding gray value model of Chan and Vese [7]. The level set formulation permits the segmentation of several (possibly multiply connected) objects, based on their relative motion.

In order to impose temporal regularity of the estimated motion segmentation, we generalize the motion boundaries from contours in 2D to surfaces in 3D (space and time). An analogous level set implementation allows to compute several multiply-connected motion phases representing moving objects or regions over time. The present paper integrates and extends results presented in [8–12].

The paper is organized as follows. In Section 2, we formulate motion estimation as a problem of Bayesian inference. In Section 3, we consistently derive a variational framework for motion segmentation. In Section 6, we introduce a level set implementation of the proposed functional. In Section 7, we present an extension to space-time motion segmentation of videos. Numerical results for simulated ground truth data and real-world sequences are given in Section 8.

2 From Motion Estimation to Motion Segmentation

2.1 Motion Estimation as Bayesian Inference

Let $\Omega \subset \mathbb{R}^2$ denote the image plane and let $f : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ be a gray value image sequence. Denote the spatio-temporal image gradient of $f(x, t)$ by

$$\nabla_3 f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial t} \right)^t. \quad (1)$$

Let

$$v : \Omega \rightarrow \mathbb{R}^3, \quad v(x) = (u(x), w(x), 1)^t, \quad (2)$$

be the velocity vector at a point x in homogeneous coordinates.¹

With these definitions, the problem of motion estimation now consists in maximizing the conditional probability

$$\mathcal{P}(v | \nabla_3 f) = \frac{\mathcal{P}(\nabla_3 f | v) \mathcal{P}(v)}{\mathcal{P}(\nabla_3 f)}, \quad (3)$$

with respect to the motion field v . For a related Bayesian formulation of motion segmentation in the discrete case, we refer to [3].

2.2 A Normalized Velocity Likelihood

In the following, we will assume that the intensity of a moving point remains constant throughout time. Expressed in differential form, this gives a relation between the spatio-temporal image gradient and the homogeneous velocity vector, known as *optic flow constraint*:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x_1} \frac{dx_1}{dt} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dt} = v^t \nabla_3 f = 0. \quad (4)$$

The optic flow constraint has been extensively exploited in the motion estimation community. Following the seminal work of Horn and Schunck [16], researchers commonly estimate motion fields by minimizing functionals which integrate this constraint in a least-squares manner (while imposing a smoothness constraint on the velocity field). In this work, we propose an alternative geometric approach to interpret the optic flow constraint. As we will argue in the following, the resulting likelihood is more appropriate in the context of motion segmentation.

¹ For the moment, we are only concerned with two consecutive frames from a sequence. Therefore we will drop the time coordinate in the notation of the velocity field.

Except for locations where the spatio-temporal gradient vanishes, the constraint (4) states that the homogeneous velocity vector must be orthogonal to the spatio-temporal image gradient. Therefore we propose to use a measure of this orthogonality as a conditional probability on the spatio-temporal image gradient. Let α be the angle between the two vectors then:

$$\mathcal{P}(\nabla_3 f(x) | v(x)) \propto e^{-\cos^2(\alpha)} = \exp\left(-\frac{(v(x)^t \nabla_3 f(x))^2}{|v(x)|^2 |\nabla_3 f(x)|^2}\right). \quad (5)$$

By construction, this probability is independent of the length of the two vectors and monotonically increases the more orthogonal the two vectors. A normalization with respect to the length of the velocity vector only has been proposed in the context of motion estimation [1]. For derivations of alternative likelihood functions from generative models of the image formation process and associated noise models, we refer to [13, 26, 38].

2.3 A Geometric Prior on the Velocity Field

We discretize the velocity field v by a set of disjoint regions $\Omega_i \subset \Omega$ with constant velocity v_i :

$$v(x) = \{v_i, \text{ if } x \in \Omega_i\} \quad (6)$$

An extension to piecewise parametric motion is presented in Section 4. We now assume the prior probability on the velocity field to only depend on the length $|C|$ of the boundary C separating these regions:

$$\mathcal{P}(v) \propto \exp(-\nu |C|) \quad (7)$$

In particular, this means that we do not make any prior assumptions on the velocity vectors v_i . Such a prior would necessarily introduce a bias favoring certain velocities. Priors on the length of separating boundaries are common in the context of variational segmentation (cf. [6, 19, 24]). For alternative more object-specific priors in the context of motion segmentation, we refer to [10]. As we shall see in the next section, the choice of velocity representation in (6) combined with the prior in (7) will transform the motion estimation framework into one of motion segmentation.

3 A Variational Framework for Motion Segmentation

With the above assumptions, we can use the framework of Bayesian inference to derive a variational method for motion segmentation. The first term in the numerator of equation (3) can be written as:

$$\mathcal{P}(\nabla_3 f | v) = \prod_{x \in \Omega} \mathcal{P}(\nabla_3 f(x) | v(x))^h = \prod_{i=1}^n \prod_{x \in \Omega_i} \mathcal{P}(\nabla_3 f(x) | v_i)^h, \quad (8)$$

where $h = dx$ denotes the grid size of the discretization of Ω .² The first step is based on the assumptions that gradient measurements are spatially independent and that the

² The introduction of the grid size h ensures the correct continuum limit.

velocity affects the spatio-temporal gradient only locally.³ And the second step is based on the discretization of the velocity field given in (6).

With the prior probability (7), maximizing the conditional probability (3) with respect to the velocity field v therefore amounts to

$$\max_v \mathcal{P}(v | \nabla_3 f) = \max_{v_i, C} \left\{ e^{-\nu |C|} \prod_{i=1}^n \prod_{x \in \Omega_i} \mathcal{P}(\nabla_3 f(x) | v_i)^h \right\}. \quad (9)$$

Equivalently one can minimize the negative logarithm of this expression, which is given by the energy functional:

$$E(C, \{v_i\}) = - \sum_{i=1}^n \int_{\Omega_i} \log(\mathcal{P}(\nabla_3 f(x) | v_i)) dx + \nu |C|. \quad (10)$$

With the conditional probability (5) on the spatio-temporal gradient, this gives:

$$E(C, \{v_i\}) = \sum_{i=1}^n \int_{\Omega_i} \frac{(v_i^t \nabla_3 f(x))^2}{|v_i|^2 |\nabla_3 f(x)|^2} dx + \nu |C|. \quad (11)$$

Let us make the following remarks about this functional:

- The functional (11) can be considered an extension of the piecewise constant Mumford-Shah functional [24] from the case of gray value segmentation to the case of motion segmentation. Rather than having a constant f_i modeling the intensity of each region Ω_i , we now have a velocity vector v_i modeling the motion in each region Ω_i .
- Gradient descent minimization with respect to the boundary C and the set of motion vectors $\{v_i\}$, jointly solves the problems of segmentation and motion estimation. In our view, this aspect is crucial, since these two problems are tightly coupled. Many alternative approaches to motion segmentation tend to instead treat the two problems separately by first (globally) estimating the motion and then trying to segment the estimated motion into a set of meaningful regions.
- The integrand in the data term differs from the one commonly used in the optic flow community for motion estimation: Rather than minimizing the deviation from the optic flow constraint in a least-squares manner, as done e.g. in the seminal work of Horn and Schunck [16], measure (5) introduces an additional normalization with respect to the length of the two vectors. In Section 5.3, we will argue that these normalizations are essential in the case of motion *segmentation*, where differently moving regions are compared.

³ Both of these assumptions are known to be inaccurate. More elaborate modeling of spatial correlations might lead to improved segmentation schemes.

- The functional (11) contains one free parameter ν , which determines the relative weight of the length constraint. Larger values of ν will induce a segmentation of the image motion on a coarser scale. As argued by Morel and Solimini [23], such a scale parameter is fundamental to all segmentation approaches.

4 Piecewise Parametric Motion Segmentation

Minimizing functional (11) generates a segmentation of the image plane into domains of piecewise constant motion. In order to cope with more complex motion regions, one can extend this approach to piecewise parametric motion. An extension of the geometric reasoning of Section 2.2 to parametric motion models is as follows.

The velocity on the domain Ω_i is allowed to vary according to a model of the form:

$$v_i(x) = M(x) p_i, \quad (12)$$

where M is a matrix depending only on space and time and p_i is the parameter vector associated with each region. A particular model which allows for expansion, contraction, rotation and shearing is the case of *affine motion* given by the matrix

$$M(x) = \begin{pmatrix} x_1 & x_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & x_2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (13)$$

and a parameter vector $p_i = (a_i, b_i, c_i, d_i, e_i, f_i, 1)$ for each region Ω_i .

Inserting model (12) into the optic flow constraint (4) gives a relation which – again interpreted geometrically – states that the the vector $M^t \nabla_3 f$ must either vanish or be orthogonal to the vector p_i . We therefore model the conditional probability that the point $x \in \Omega$ belongs to the domain Ω_i by a quantity which only depends on the angle between p_i and $M^t \nabla_3 f$:

$$P(\nabla_3 f | p_i) \propto \exp\left(-\frac{(p_i^t M^t \nabla_3 f)^2}{|p_i|^2 |M^t \nabla_3 f|^2}\right). \quad (14)$$

The corresponding generalization of functional (11) from piecewise constant to piecewise parametric motion segmentation is given by:

$$E(C, \{p_i\}) = \sum_i \int_{\Omega_i} \frac{|p_i^t M^t \nabla_3 f|^2}{|p_i|^2 |M^t \nabla_3 f|^2} dx + \nu |C|. \quad (15)$$

5 Energy Minimization

The functional (15) is of the form

$$E(C, \{p_i\}) = \sum_{i=1}^n \int_{\Omega_i} \frac{p_i^t T(x) p_i}{|p_i|^2} dx + \nu |C|, \quad (16)$$

where, for notational simplification, we have introduced the matrix

$$T(x) = \frac{\nabla_3 f M^t M \nabla_3 f^t}{|M^t \nabla_3 f|^2}. \quad (17)$$

This functional is minimized by alternating the two fractional steps of optimizing with respect to the motion parameters $\{p_i\}$ for fixed boundary C , and iterating the gradient descent with respect to C for fixed parameters $\{p_i\}$.

5.1 An Eigenvalue Problem for the Motion Parameters

The functional (16) can be further simplified:

$$E(C, \{p_i\}) = \sum_{i=1}^n \frac{p_i^t T_i p_i}{|p_i|^2} dx + \nu |C|, \quad \text{where } T_i = \int_{\Omega_i} T(x) dx, \quad (18)$$

with T given in (17). For fixed boundary C , i.e. fixed regions Ω_i , minimizing this functional with respect to the motion parameters $\{p_i\}$ results in a set of eigenvalue problems of the form:

$$p_i = \arg \min_p \frac{p^t T_i p}{p^t p}. \quad (19)$$

The parametric motion model p_i for each region Ω_i is therefore given by the eigenvector corresponding to the smallest eigenvalue of the matrix T_i defined above. It is normalized, such that the third component is 1. Similar eigenvalue problems arise in motion estimation due to normalization with respect to the velocity magnitude (cf. [1, 18]).

5.2 Motion Competition

Conversely, for fixed motion models p_i , a gradient descent on the energy (16) for the boundary C results in the evolution equation:

$$\frac{\partial C}{\partial t} = -\frac{\partial E}{\partial C} = (e_j - e_k) n - \nu \frac{d|C|}{dC}, \quad (20)$$

where the indices ‘ j ’ and ‘ k ’ refer to the regions adjoining the contour, n denotes the normal vector on the boundary pointing into region Ω_j , and

$$e_i = \frac{p_i^t T p_i}{p_i^t p_i} = \frac{p_i^t \nabla_3 f M^t M \nabla_3 f^t p_i}{|p_i|^2 |M^t \nabla_3 f|^2} \quad (21)$$

is an energy density.

Note that we have neglected in the evolution equation (20) higher-order terms which account for the dependence of the motion parameters p_i on the regions Ω_i . An Eulerian accurate shape optimization scheme as presented for example in [17] is the focus of ongoing research.

The two terms in the contour evolution (20) have the following intuitive interpretation:

- The first term is proportional to the difference of the energy densities e_i in the regions adjoining the boundary: The neighboring regions compete for the boundary in terms of their motion energy density, thereby maximizing the motion homogeneity. For this reason we refer to this process as *Motion Competition*.
- The second term minimizes the length \mathcal{L} of the separating motion boundary.

5.3 Effect of the Normalization

In Section 2.2 we argued that the proposed likelihood (5) (in contrast to the commonly used least-squares formulation) does not introduce a bias with respect to the magnitude of the velocity or the image gradient.⁴ As a direct consequence, the respective contour evolutions differ, as we will detail for the case of piecewise constant motion.

The proposed motion likelihood (5) results in a contour evolution of the form (20) with energy densities

$$e_i = \frac{v_i^t \nabla_3 f \nabla_3 f^t v_i}{|v_i|^2 |\nabla_3 f|^2} \quad (22)$$

This means that the term driving the contour evolution does not depend on the magnitude of the spatio-temporal gradient and it does not depend on the magnitude of the respective velocity models.

In contrast, a Horn-and-Schunck type likelihood [16] would induce contour driving terms which do not include the normalizing terms in the denominator:

$$e_i = v_i^t \nabla_3 f \nabla_3 f^t v_i. \quad (23)$$

This lack of normalization has two effects on the boundary evolution and resulting segmentation: Firstly the motion boundary will propagate much faster in areas of high gradient. Secondly the evolution direction and speed will be affected by the magnitude of velocities: regions with larger velocity will exert a stronger pull on the motion boundary.

6 A Multiphase Level Set Implementation

A few years after its introduction in [28], the level set based evolution of contours was adopted as a framework for image segmentation (cf. [6, 7, 19]). In contrast to explicit boundaries, the level set representation does not depend on a particular choice of parameterization. During the evolution of the boundary one avoids the issues of control point regridding. Moreover, the topology of the evolving interface is not constrained. This permits splitting and merging of the contour during evolution and therefore makes level set representations well suited for the segmentation of several objects or multiply connected objects.

Based on a corresponding gray value model of Chan and Vese [7], we will first present a two-phase level set model for the motion competition functional (16) with a single level set function ϕ . This model is subsequently extended to a multi-phase model on the basis of a vector-valued level set function.

⁴ In particular, the functionals (11) and (16) are invariant to global scale transformations of the intensity: $f \rightarrow \gamma f$.

6.1 The Two-Phase Model

In this subsection, we restrict the class of permissible motion segmentations to two-phase solutions, i.e. to segmentations of the image plane for which each point can be ascribed to one of two velocity models p_1 and p_2 . The general case of several velocity models $\{p_i\}_{i=1,\dots,n}$ will be treated in the next subsection.

Let the boundary C in the functional (16) be represented as the zero level set of a function $\phi : \Omega \rightarrow \mathbb{R}$:

$$C = \{x \in \Omega \mid \phi(x) = 0\}. \quad (24)$$

With the Heaviside step function

$$H(\phi) = \begin{cases} 1 & \text{if } \phi \geq 0 \\ 0 & \text{if } \phi < 0 \end{cases}, \quad (25)$$

the energy (16) can be embedded by the following *two-phase functional*:

$$\begin{aligned} E(p_1, p_2, \phi) &= \int_{\Omega} \frac{p_1^t T p_1}{|p_1|^2} H(\phi) dx + \int_{\Omega} \frac{p_2^t T p_2}{|p_2|^2} (1 - H(\phi)) dx \\ &+ \nu \int_{\Omega} |\nabla H(\phi)| dx. \end{aligned} \quad (26)$$

The first two terms in (26) enforce a homogeneity of the estimated motion in the two phases, while the last term enforces a minimal length of the region boundary given by the zero level set of ϕ .

The two-phase functional (26) is simultaneously minimized with respect to the velocity models p_1 and p_2 , and with respect to the embedding level set function ϕ defining the motion boundaries. To this end, we alternate the two fractional steps:

(a) Updating the Motion Models.

For fixed ϕ , minimization of the functional (26) with respect to the motion vectors p_1 and p_2 results in the eigenvalue problem:

$$p_i = \arg \min_v \frac{v^t T_i v}{v^t v}, \quad (27)$$

for the matrices

$$T_1 = \int_{\Omega} T(x) H(\phi) dx \quad \text{and} \quad T_2 = \int_{\Omega} T(x) (1 - H(\phi)) dx. \quad (28)$$

The solution of (27) is given by the eigenvectors corresponding to the smallest eigenvalues of T_1 and T_2 .

(b) Evolution of the Level Set Function.

Conversely, for fixed motion vectors, the gradient descent on the functional (26) for the level set function ϕ is given by:

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[\nu \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + e_2 - e_1 \right], \quad (29)$$

with the energy densities e_i given in (21). As suggested in [7], we implement the Delta function $\delta(\phi) = \frac{d}{d\phi}H(\phi)$ by a smooth approximation of finite width σ :

$$\delta_\sigma(s) = \frac{1}{\pi} \frac{\sigma}{\sigma^2 + s^2}. \quad (30)$$

Thereby the update of ϕ is not restricted to the areas of zero-crossing, but rather spread out over a band of width σ around it. Depending on the size of σ , this permits to detect interior motion boundaries.

6.2 The General Multiphase Model

Compared to the explicit contour representation, the above level set representation permits to segment several, possibly multiply connected, moving regions. Yet, the representation of the motion boundary with a *single* level set function ϕ permits to model motion fields with only two phases (i.e. it permits only two different velocity models). Moreover, one cannot represent certain geometrical features of the boundary, such as triple junctions, by the zero level set of a single function ϕ . There are various ways to overcome these limitations by using multiple level set functions.

An elegant solution to model multiple phases was proposed by Chan and Vese in [7]. Rather than representing each phase by a separate level set function, they introduce a more compact representation of up to n phases which needs only $m = \log_2(n)$ level set functions.⁵ Moreover, by definition, the suggested approach generates a partition of the image plane and therefore does not suffer from overlap or vacuum formation, a difficulty which commonly arises when modeling each region by its own level set function. We will therefore adopt this representation of Chan and Vese to implement multiple motion phases, as detailed in the following.

Consider a set of m level set functions $\phi_i : \Omega \rightarrow \mathbb{R}$, let

$$\Phi = (\phi_1, \dots, \phi_m) \quad (31)$$

be a vector level set function and let $H(\Phi) = (H(\phi_1), \dots, H(\phi_m))$ be the associated vector Heaviside function. This function maps each point $x \in \Omega$ to a binary vector and therefore permits to encode a set of $n = 2^m$ phases Ω_i defined by:

$$R = \{x \in \Omega \mid H(\Phi(x)) = \text{constant}\}. \quad (32)$$

In analogy to the corresponding level set formulation of the Mumford-Shah functional [7], we propose to replace the two-phase functional (26) by the *multiphase functional*:

$$E(\{p_i\}, \Phi) = \sum_{i=1}^n \int_{\Omega} \frac{p_i^t T p_i}{|p_i|^2} \chi_i(\Phi) dx + \nu \sum_{i=1}^n \int_{\Omega} |\nabla H(\phi_i)| dx, \quad (33)$$

where χ_i denotes the indicator function for the region Ω_i . Note, that for $n = 2$, this is equivalent to the two-phase model introduced in (26).

For further details regarding the minimization of this multiphase model, we refer to [8].

⁵ During the optimization certain phases may disappear such that the final segmentation may consist of less than n phases.

6.3 Redistancing

During their evolution, the level set functions generally grow to very large positive or negative values in the respective areas of the input image corresponding to a particular motion hypothesis. Indeed, there is nothing in the level set formulation of Chan and Vese [7] which prevents the level set functions from growing indefinitely. In numerical implementations, we found that a very steep slope of the level set functions can even inhibit the flexibility of the boundary to displace. In order to reproject the evolving level set function to the space of distance functions, we intermittently iterate several steps of the redistancing equation [32]:

$$\frac{\partial \phi}{\partial t} = \text{sign}(\hat{\phi}) (1 - |\nabla \phi|), \quad (34)$$

where $\hat{\phi}$ denotes the level set function before redistancing. This transformation does not affect the motion boundaries given by the zero-crossing of ϕ . It merely enforces the gradient to be of magnitude 1.

7 Motion-Based Space-Time Segmentation of Videos

The above framework allows to segment images into regions of parametric motion based on two consecutive frames from an image sequence. Given an entire video sequence, one can apply the proposed method iteratively to consecutive frame pairs – see [9] for details. In practical applications, the inferred contour tends to jitter over time.

Rather than processing the sequence frame by frame, one can impose temporal regularity of the inferred segmentation by casting the motion segmentation problem as one of identifying volumes $D_i \subset \Omega \times [0, \tau]$ of coherent motion in space-time, where τ denotes the length of the sequence. Rather than minimizing (16) with respect to a boundary C , one minimizes by the functional

$$E(S, \{p_i\}) = \sum_{i=1}^n \int_{D_i} \frac{p_i^t T(x, t) p_i}{|p_i|^2} dx dt + \nu |S|, \quad (35)$$

with respect to a surface $S \subset \Omega \times [0, \tau]$ separating the phases D_i . The constraint on the area $|S|$ of the surface imposes regularity of the segmentation both in space and in time.

The parametric representation of velocity fields (12) can be directly extended into the temporal domain, thereby allowing to consider the case of *accelerated motion* with

$$M(x, t) = \begin{pmatrix} 1 & 0 & t & 0 & 0 \\ 0 & 1 & 0 & t & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (36)$$

and $p_i = (u, w, a_u, a_w, 1)$ modeling an accelerated motion in each domain. Combinations of models of spatial and temporal variation are conceivable to capture accelerated rotations and other kinds of motion.

For the extension of the multiphase framework of Section 6.2 to propagate motion surfaces in space-time, the reader is referred to [11].

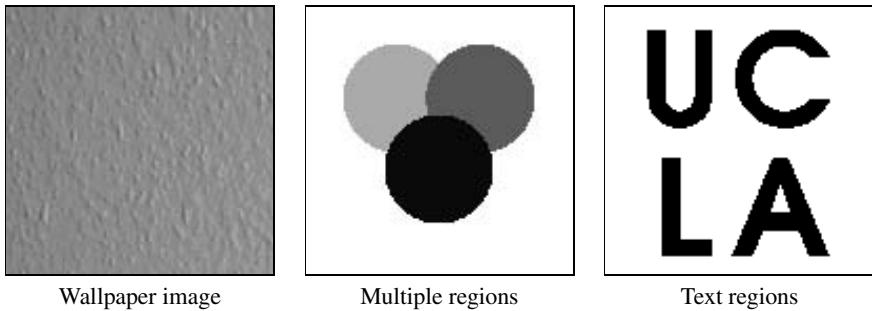


Fig. 1. Data for ground truth experiments. Specific image regions of the wallpaper shot (left) are artificially translated to generate input data.

8 Numerical Results

In the following, we will present numerical results demonstrating various properties of the proposed framework for motion segmentation.

For all experiments we determined the spatio-temporal image gradient from two consecutive images and specified a particular initialization of the boundary (or surface in the case of space-time segmentation). We subsequently minimized the functionals (33) or (35) by alternating the three fractional steps of:

- updating the motion models for all phases by solving the corresponding eigenvalue problem (27),
- evolving the level set functions by iterating the appropriate gradient descent pdes – e.g. equation (29) in the two-phase case,
- and redistancing the level set functions according to (34).

For all experiments, we show the evolving motion boundaries (and in some cases also the corresponding motion estimates) superimposed onto one of the frames. It should be noted that all results are obtained *exclusively* on the basis of the motion information.

8.1 Accurate Motion Segmentation Without Features

In order to verify the spatial precision of the motion competition approach, we performed a number of ground truth experiments in the following way. We took a snapshot of homogeneously structured wallpaper. We artificially translated certain image regions according to specific motion models. The input image and the respective image regions are highlighted (in various shades of gray) in Figure 1.

In the first example, we show that one can generate spatially accurate segmentation results exploiting only motion information, even for image sequences that exhibit little intensity variation or salient features. Figure 2 shows the contour evolution generated by minimizing functional (11). The input data consists of two wall paper images with the text region (Figure 1, right side) moving to the right and the remainder of the

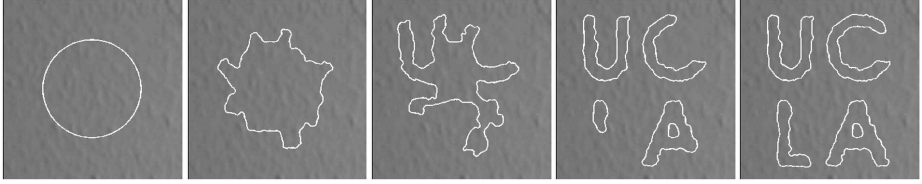


Fig. 2. Accurate motion segmentation. Contour evolution obtained with functional (26) and parameter values $\nu = 0.06$, $\sigma = 1$, superimposed on one of the two input frames. The input images show the text region (Figure 1, right side) of the wallpaper moving to the right and the remainder moving to the left. The motion competition framework generates highly accurate segmentations, even if the input images exhibit little in terms of salient features. Due to the region-based formulation, the initial contour does not need to be close to the final segmentation. We found that alternative initializations generate essentially identical segmentation results. The contour evolution took approximately 10 seconds in Matlab.

image plane moving to the left. Even for human observers the differently moving regions are difficult to detect – similar to a camouflaged lizard moving on a similarly-textured ground. The gradient descent evolution superimposed on one of the two frames gradually separates the two motion regions without requiring salient features such as edges or corner points.

8.2 Segmenting Several Motion Phases

In this experiment, we demonstrate an application of the multi-phase model (33) to the segmentation of up to four different regions based on their motion information. The input data consists of two images showing the wallpaper from Figure 1, left side, with three regions (shown in Figure 1, right side) moving away from the center. The upper two regions move by a factor 1.4 faster than the lower region.

Figure 3 shows several steps in the minimization of the functional (33) for two level set functions. Superimposed onto the ground truth region information are the evolution of the zero level sets of the two embedding functions ϕ_1 (black contour) and ϕ_2 (white contour), and the estimated piecewise constant motion field indicated by the black arrows.

Note that the two contours represent a set of four different phases:

$$\begin{aligned}\Omega_1 &= \{x \in \Omega \mid \phi_1 \geq 0, \phi_2 \geq 0\}, & \Omega_2 &= \{x \in \Omega \mid \phi_1 \geq 0, \phi_2 < 0\}, \\ \Omega_3 &= \{x \in \Omega \mid \phi_1 < 0, \phi_2 \geq 0\}, & \Omega_4 &= \{x \in \Omega \mid \phi_1 < 0, \phi_2 < 0\}.\end{aligned}$$

Upon convergence, these four phases clearly separate the three moving regions and the static background. The resulting final segmentation of the image, which is not explicitly shown here, is essentially identical to the ground truth region information. Note that the segmentation is obtained purely on the basis of the *motion information*: In the input images, the different regions cannot be distinguished from the background on the basis of their *appearance*.

8.3 Intensity Segmentation Versus Motion Segmentation

All image segmentation models are based on a number of more or less explicitly stated assumptions about the properties which define the objects of interest. The motion competition model is based on the assumption that objects are defined in terms of homogeneously moving regions. It extends the Mumford-Shah functional of piecewise constant intensity to a model of piecewise parametric motion.

In this example, we will show that despite this formal similarity the segmentations generated by the motion competition framework are very different from those of its gray value analog. The task is to segment a real-world traffic scene showing two moving cars on a differently moving background. We used two consecutive images from a sequence recorded by D. Koller and H.-H. Nagel (KOGS/IAKS, University of Karlsruhe).⁶ The sequence shows several cars moving in the same direction, filmed by a static camera. In order to increase the complexity of the sequence, we artificially induced a background motion by selecting a subarea of the original sequence and shifting one of the two frames, thereby simulating the case of a moving camera.

Figure 4, top, shows the boundary evolution obtained by minimizing the two-phase model of Chan and Vese [7] for the first of the two frames. The segmentation process progressively separates bright and dark areas of the image plane. Yet, since the objects of interest are not well-defined in terms of homogeneous gray value, the final segmentation inevitably fails to capture them. The dark car in the lower left is associated with the darker parts of the street, whereas the car in the upper right is split into its brighter and darker parts.

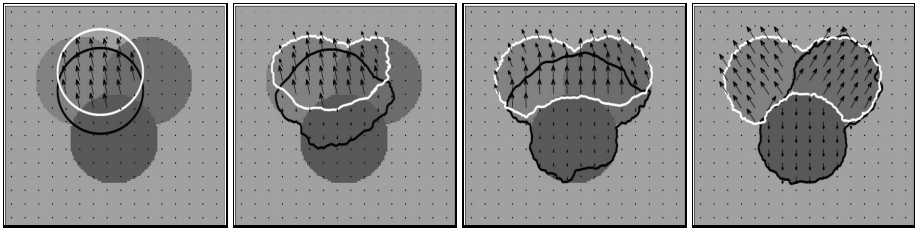
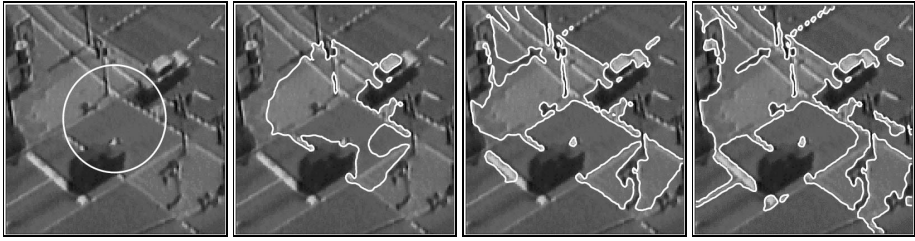


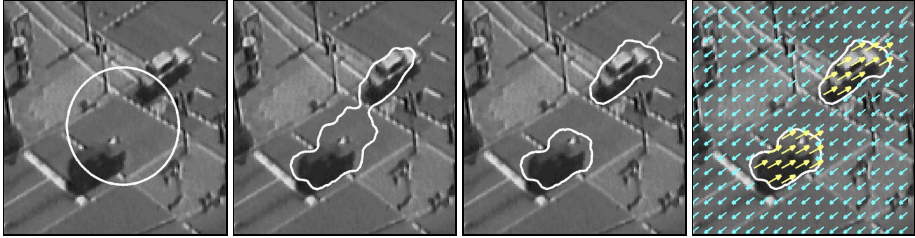
Fig. 3. Segmenting multiple moving regions. The two input images show the wallpaper of Figure 1, left side, with three circular regions moving away from the center. The magnitude of the velocity of the upper two regions is 1.4 times larger than that of the bottom region. Superimposed on the true region information are the evolving zero level sets of ϕ_1 (black contour) and ϕ_2 (white contour), which define four different phases. The simultaneously evolving piecewise constant motion field is represented by the black arrows. Both the phase boundaries and the motion field are obtained by minimizing the multiphase model (33) with parameters $\nu = 0.05$, $\sigma = 2$ with respect to the level set functions and the motion vectors. In the final solution, the two boundaries clearly separate four phases corresponding to the three moving regions and the static background.

In this example, the cars and the street are moving according to different motion models. The motion competition framework exploits this property. Figure 4, bottom,

⁶ http://i21www.ira.uka.de/image_sequences/



Intensity segmentation for the first of two frames.



Motion segmentation generated from both frames.

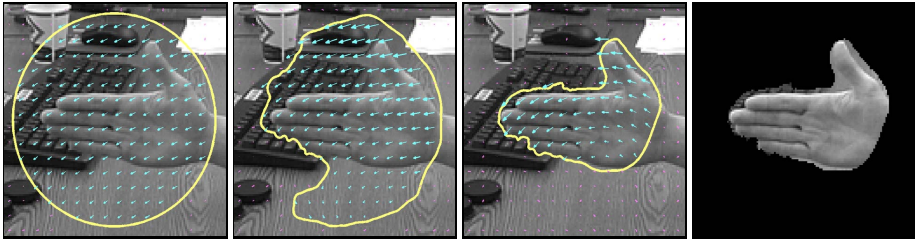
Fig. 4. Intensity segmentation versus motion segmentation. Two consecutive input frames show two cars moving to the top right, and the background moving to the bottom left. **Top row:** Segmentation of the first frame from a traffic scene according to the two-phase level set model of the piecewise constant Mumford-Shah functional, as introduced by Chan and Vese [7]. The assumption of homogeneous intensity is clearly not appropriate to segment the objects of interest. **Bottom:** Motion segmentation of the same traffic scene. By minimizing the motion competition functional (26) with parameters $\nu = 1.5$, $\sigma = 5$, one obtains a fairly accurate segmentation of the two cars and an estimate of the motion of cars and background. Since the objects of interest are better defined in terms of homogeneous motion than in terms of homogeneous intensity, the segmentation is more successful than the one obtained by the analogous gray value model. Until convergence, the contour evolution took 41 seconds in Matlab on a 2.4 GHz computer.

show the contour evolution generated by minimizing the motion segmentation functional (26) and the corresponding motion estimates superimposed on the first frame.

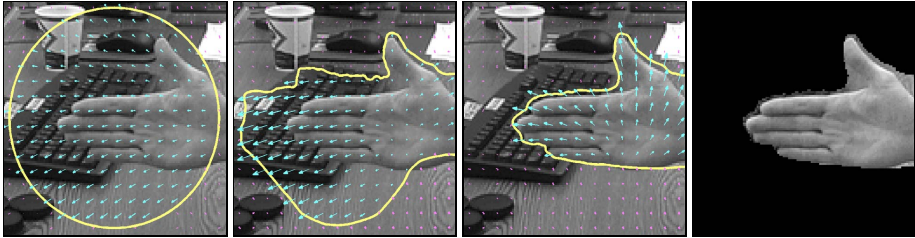
The contour evolution generated by motion competition is fundamentally different from the one generated by its gray value analog. The energy minimization simultaneously generates a fairly accurate segmentation of the two cars and an estimate of the motion of cars and background. Minor discrepancies of the final segmentation may be due to several factors, in particular the weak gray value structure of the street, which prevents reliable motion estimation, and the reflections on the cars which violate the Lambertian assumption.

8.4 Segmentation by Piecewise Affine Motion

The functional (16) allows to segment piecewise affine motion fields. In particular, this class of motion models includes rotation and expansion/contraction. Figure 5 shows



Motion segmentation of a hand rotating around the wrist.



Motion segmentation of a hand moving toward the camera.

Fig. 5. Piecewise affine motion segmentation. Functional (16) allows to segment objects based on the model of affine motion. The above images show contour evolutions obtained for two image pairs showing a hand rotating (top) and moving toward the camera (bottom). Minor discrepancies of the final segmentation (right) are probably due to a lack of gray value variation of the table. Both results were obtained with the same parameter values ($\nu = 8 \cdot 10^{-5}$, $\sigma = 2$). Again certain areas of little intensity variation do not provide sufficient motion information to be reliably associated with one or the other motion model.

contour evolutions obtained for a hand in a cluttered background rotating (in the camera plane) and moving toward the camera. The energy minimization allows to segment the object and estimate its rotational or divergent motion.

The images on the right of Figure 5 demonstrate that the objects of interest can be extracted from a fairly complex background based exclusively on their motion. Applications of such motion-based segmentation schemes to video editing and MPEG compression are conceivable.

8.5 Spatio-temporal Motion Segmentation

While the previous results were obtained using only two consecutive frames from a sequence, we will now present an application of the segmentations in space-time obtained by minimizing a multiphase implementation of the functional (35) for several frames of the flower garden sequence [35], which shows a static scene filmed by a moving camera. Figure 6 shows the evolution of the surfaces separating the motion phases in space-time (top rows). The lower rows depict the corresponding temporal slices of these surfaces associated with the frames 2, 5 and 8. During energy minimization, the surfaces propagate to the final segmentation both in space and in time. The final segmentation clearly separates foreground, midplane and background. The simultaneously evolving piecewise constant motion field is depicted in Figure 7.

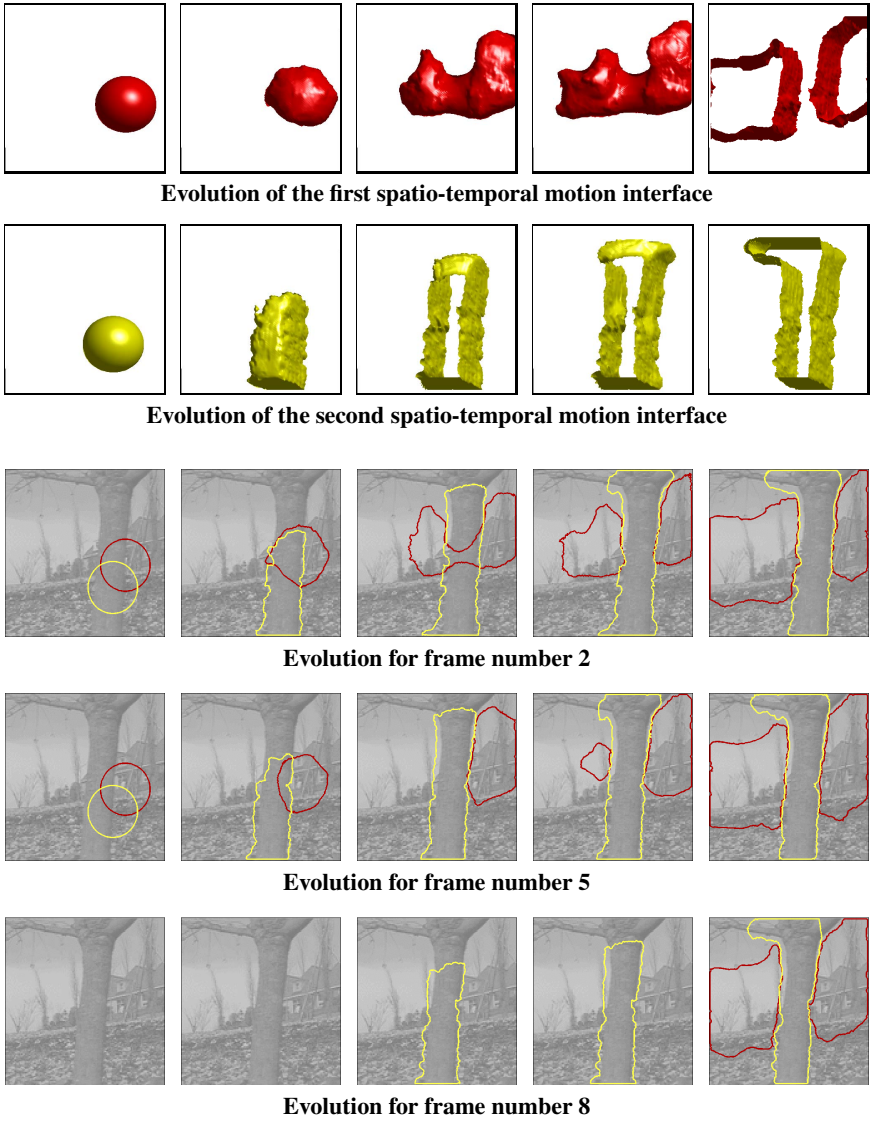


Fig. 6. Spatio-temporal sequence segmentation with the multiphase model. The top rows show the evolution of the spatio-temporal surfaces given by the zero level sets of two embedding functions ϕ_1 and ϕ_2 . The lower rows show various temporal slices of these surfaces, corresponding to the 2nd, 5th and 8th frame of the sequence. The evolving surfaces propagate both in space and time during minimization of the energy (35). In the final segmentation the phases clearly separate foreground, midplane and background. For better visibility, the simultaneously estimated piecewise constant motion field is shown separately in Figure 7.

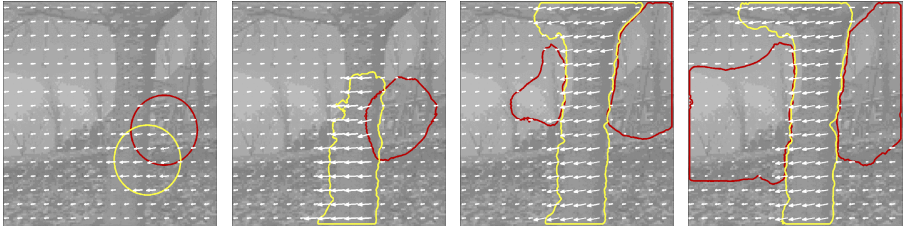


Fig. 7. Temporal slice through the evolving surfaces shown in Figure 6. The final segmentation separates the tree in the foreground, the grass in the midplane and the houses and smaller trees in the background. Boundary and the motion estimates are obtained by simultaneously minimizing an appropriate cost functional defined on the spatio-temporal image derivatives. Unlike most alternative approaches to layer extraction, no preprocessing (such as local disparity estimation, camera calibration and prior rectification of individual frames [33]) is applied to the image data.

9 Conclusion

We derived a variational framework for segmenting the image plane (or the space-time volume of an image sequence) into a set of phases of parametric motion. The proposed functional depends on parametric velocity models for a set of phases and the boundary separating them. The only free parameter in the functional is the fundamental scale parameter intrinsic to all segmentation schemes.

The motion discontinuity set is implemented by a multiphase level set formulation (for contours in 2D or surfaces in 3D). The resulting model has the following properties:

- The minimization of a *single* functional jointly solves the problems of segmentation and motion estimation. It generates a segmentation of the image plane (or the space-time volume) in terms of piecewise parametric motion.
- An extension to surfaces in space-time allows to segment moving regions over time, providing an additional temporal regularity of the segmentation.
- *Implicit multiphase representations* allow for topological changes of the evolving boundaries. They permit a segmentation of the image plane (or the space-time volume) into several (possibly multiply-connected) motion phases.
- Local minimization of the proposed functional results in an eigenvalue problem for the motion vectors, and an evolution equation of the level set functions embedding the motion boundary.
- Due to the region-based homogeneity criterion rather than an edge-based formulation, motion boundaries converge over fairly large spatial distances.
- Segmentation and motion estimates can be generated from two consecutive frames of an image sequence. Therefore the approach is in principle amenable to real-time implementations and tracking.

Acknowledgments

The author thanks C. Schnörr, S. Soatto and A. Yuille for fruitful discussions.

References

1. J. Bigün, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE PAMI*, 13(8):775–790, 1991.
2. M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Comp. Vis. Graph. Image Proc.: IU*, 63(1):75–104, 1996.
3. P. Bouthemy and E. Francois. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *Int. J. of Comp. Vis.*, 10(2):157–182, 1993.
4. T. Brox, A. Bruhn, N. Papenbergh, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and V. Hlavac, editors, *European Conf. on Computer Vision*, volume 3024 of *LNCS*, pages 25–36, Prague, 2004. Springer.
5. V. Caselles and B. Coll. Snakes in movement. *SIAM J. Numer. Anal.*, 33:2445–2456, 1996.
6. V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, pages 694–699, Boston, USA, 1995.
7. T. Chan and L. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.
8. D. Cremers. A multiphase level set framework for variational motion segmentation. In L. Griffith, editor, *Int. Conf. on Scale Space Theories in Computer Vision*, volume 2695 of *LNCS*, pages 599–614, Isle of Skye, 2003. Springer.
9. D. Cremers. A variational framework for image segmentation combining motion estimation and shape regularization. In C. Dyer and P. Perona, editors, *IEEE Conf. on Comp. Vis. and Patt. Recog.*, volume 1, pages 53–58, June 2003.
10. D. Cremers and C. Schnörr. Motion Competition: Variational integration of motion segmentation and shape regularization. In L. van Gool, editor, *Pattern Recognition*, volume 2449 of *LNCS*, pages 472–480, Zürich, Sept. 2002. Springer.
11. D. Cremers and S. Soatto. Variational space-time motion segmentation. In B. Triggs and A. Zisserman, editors, *IEEE Int. Conf. on Computer Vision*, volume 2, pages 886–892, Nice, Oct. 2003.
12. D. Cremers and S. Soatto. Motion Competition: A variational framework for piecewise parametric motion segmentation. *Int. J. of Comp. Vis.*, 2004. To appear.
13. D. Cremers and A. L. Yuille. A generative model based approach to motion segmentation. In B. Michaelis and G. Krell, editors, *Pattern Recognition*, volume 2781 of *LNCS*, pages 313–320, Magdeburg, Sept. 2003. Springer.
14. G. Farneback. Very high accuracy velocity estimation using orientation tensors, parametric motion, and segmentation of the motion field. In *ICCV*, volume 1, pages 171–177, 2001.
15. F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. *IEEE PAMI*, 15(12):1217–1232, 1993.
16. B.K.P. Horn and B.G. Schunck. Determining optical flow. *A.I.*, 17:185–203, 1981.
17. S. Jehan-Besson, M. Barlaud, and G. Aubert. DREAM2S: Deformable regions driven by an eulerian accurate minimization method for image and video segmentation. *Int. J. of Comp. Vis.*, 53(1):45–70, 2003.
18. A. Jepson and M.J. Black. Mixture models for optic flow computation. In *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, pages 760–761, New York, 1993.
19. S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi. Gradient flows and geometric active contour models. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, pages 810–815, Boston, USA, 1995.
20. J. Konrad and E. Dubois. Bayesian estimation of motion vector fields. *IEEE PAMI*, 14(9):910–927, 1992.
21. P. Kornprobst, R. Deriche, and G. Aubert. Image sequence analysis via partial differential equations. *J. Math. Im. Vis.*, 11(1):5–26, 1999.

22. E. Memin and P. Perez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Trans. on Im. Proc.*, 7(5):703–719, 1998.
23. J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhäuser, Boston, 1995.
24. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
25. H.H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE PAMI*, 8(5):565–593, 1986.
26. O. Nestares, D.J. Fleet, and D.J. Heeger. Likelihood functions and confidence bounds for total-least-squares problems. In *Proc. Conf. Computer Vis. and Pattern Recog.*, volume 1, pages 760–767, Hilton Head Island, SC, June, 2000.
27. J.-M. Odobez and P. Bouthemy. Direct incremental model-based image motion segmentation for video analysis. *Signal Proc.*, 66:143–155, 1998.
28. S. J. Osher and J. A. Sethian. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. *J. of Comp. Phys.*, 79:12–49, 1988.
29. N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE PAMI*, 22(3):266–280, 2000.
30. C. Schnörr. Computation of discontinuous optical flow by domain decomposition and shape optimization. *Int. J. of Comp. Vis.*, 8(2):153–165, 1992.
31. J. Shi and J. Malik. Motion segmentation and tracking using normalized cuts. In *Intl. Conf. on Comp. Vision*, Bombay, India, 1998.
32. M. Sussman, Smereka P., and S. J. Osher. A level set approach for computing solutions to incompressible twophase flow. *J. of Comp. Phys.*, 94:146–159, 1994.
33. P. H. S. Torr, R. Szeliski, and P. Anandan. An integrated bayesian approach to layer extraction from image sequences. *IEEE PAMI*, 23(3):297–303, 2002.
34. R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (gpca): an analytic solution to segmentation of mixtures of subspaces. In *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, volume 1, pages 621–628, Madison, 2003.
35. J.Y.A. Wang and E.H. Adelson. Representating moving images with layers. *IEEE Trans. on Image Processing*, 3(5):625–638, 1994.
36. J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE–based computation of image motion. *Int. J. of Comp. Vis.*, 45(3):245–264, 2001.
37. Y. Weiss. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, pages 520–527, Puerto Rico, 1997.
38. Y. Weiss and D.J. Fleet. Velocity likelihoods in biological and machine vision. In M.S. Lewicki R.P.N. Rao, B.A. Olshausen, editor, *Probabilistic Models of the Brain: Perception and Neural Function*, pages 81–100. MIT Press, 2001.
39. S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE PAMI*, 18(9):884–900, 1996.