



Motion Competition: A Variational Approach to Piecewise Parametric Motion Segmentation

DANIEL CREMERS* AND STEFANO SOATTO

Department of Computer Science, University of California, Los Angeles, CA 90095, USA

cremers@ucla.edu (<http://www.cs.ucla.edu/~cremers>)

soatto@ucla.edu

Received February 12, 2003; Revised May 18, 2004; Accepted June 3, 2004

First online version published in November, 2004

Abstract. We present a novel variational approach for segmenting the image plane into a set of regions of parametric motion on the basis of two consecutive 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.

Exploiting the Bayesian framework, we derive a cost functional which depends on parametric motion models for each of a set of regions and on the boundary separating these regions. 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. Minimizing this functional with respect to its dynamic variables results in an eigenvalue problem for the motion parameters and in a gradient descent evolution for the motion discontinuity set.

We propose two different representations of this motion boundary: an explicit spline-based implementation which can be applied to the motion-based tracking of a single moving object, and an implicit multiphase level set implementation which allows for the segmentation of an arbitrary number of multiply connected moving objects.

Numerical results both for simulated ground truth experiments and for real-world sequences demonstrate the capacity of our approach to segment objects based exclusively on their relative motion.

Keywords: image segmentation, variational methods, motion estimation, Bayesian inference, level set methods, multiphase motion, optic flow

1. Introduction

The estimation of motion from image sequences has a long tradition in computer vision. Two seminal variational methods were proposed by Horn and Schunck (1981) and by Lucas and Kanade (1981). Both of these methods are based on a least-squares criterion for the optic flow constraint, and some global or local smoothness assumption on the estimated flow field.

In practice, flow fields are generally not smooth. The boundaries of moving objects will correspond to discontinuities in the motion field. At these discontinuities, the smoothness assumption is strongly violated. Yet, one cannot simply drop the regularization term, since the problem of motion estimation is highly ill-posed. Ideally, one would like to enforce a regularity of the estimated motion field only in the areas corresponding to the different moving objects, allowing for discontinuities across the boundaries of objects. Yet this requires knowledge of the correct segmentation. In this sense, the segmentation of objects based on

*Daniel Cremers is now with Siemens Corporate Research, Princeton, NJ.

their motion has been considered a chicken and egg problem.

Many researchers have addressed this coupling of segmentation and motion estimation. Rather than first estimating local motion and subsequently segmenting or clustering regions with respect to the estimated motion (Wang and Adelson, 1994), some researchers have proposed to model motion discontinuities implicitly by non-quadratic robust estimators (Nagel and Enkelmann, 1986; Black and Anandan, 1996; Memin and Perez, 1998; Kornprobst et al., 1999; Weickert and Schnörr, 2001; Brox et al., 2004). Others tackled the problem of segmenting motion by treating the problems of motion estimation in disjoint sets and optimization of the motion boundaries separately (Schnörr, 1992; Black, 1994; Caselles and Coll, 1996; Odobez and Bouthemy, 1998; Paragios and Deriche, 2000; Farneback, 2001). Some approaches are based on Markov Random Field (MRF) formulations and optimization schemes such as stochastic relaxation by Gibbs sampling (Konrad and Dubois, 1992), split-and-merge techniques (Heitz and Bouthemy, 1993; Zheng and Blostein, 1995), deterministic relaxation (Bouthemy and Francois, 1993), graph cuts (Shi and Malik, 1998) or expectation maximization (EM) (cf. Jepson and Black, 1993; Ayer and Sawhney, 1995; Weiss, 1997). As pointed out in Weiss (1997), exact solutions to the EM algorithm are computationally expensive and therefore suboptimal approximations are employed.

In Cremers and Schnörr (2003), we presented a variational approach to motion segmentation with an explicit contour, in which motion estimation and boundary optimization are derived by minimizing a *single* energy functional by gradient descent. This approach had two drawbacks: Firstly, satisfactory results were only obtained upon applying two *posterior* normalizations to the terms driving the evolution of the motion boundary (see also (Farneback, 1999)), which is not consistent with the minimization of a single energy. Secondly, due to the explicit representation of this boundary, the segmentation of multiple moving objects is not straight-forward.

In the present paper, we propose an approach which will overcome these limitations. We formulate the problem of motion estimation in the framework of Bayesian inference. Related Bayesian formulations have been proposed in the discrete MRF framework (cf. Bouthemy and Francois, 1993). Our formulation differs from the above approach in that it is continuous,

uses a contour representation of the motion discontinuity set (spline or level set based), can be optimized by a simple and fast gradient descent minimization and is based on a different (normalized) likelihood in the data term. We propose a geometrically motivated model for the conditional probability of a spatio-temporal image gradient given a particular velocity vector, and a prior on the estimated motion field favoring motion boundaries of minimal length. We derive a novel variational formulation for segmenting the image plane into a set of disjoint regions of piecewise parametric motion. The proposed functional can be interpreted as an extension of the Mumford and Shah (1989) model 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* (Zhu and Yuille, 1996), we therefore refer to this process as *Motion Competition*.

We propose two implementations of this boundary set. The first one is based on a spline representation of the boundary. Although such a sparse representation of the object boundary is computationally efficient, explicit contour representations have a number of drawbacks. Firstly, they rely on a particular parameterization. During the evolution, some regridding mechanism needs to assure that control points do not overlap. Secondly, the explicit representation does not permit topological changes of the boundary such as splitting and merging. To overcome these limitations, we revert to an implicit level set based representation of the boundary. Level set based contour representations (Osher and Sethian, 1988) have become a popular framework in image segmentation (cf. Malladi et al., 1995; Caselles et al., 1995; Kichenassamy et al., 1995; Caselles and Coll, 1996; Paragios and Deriche, 2000; Chan and Vese, 2001; Yezzi and Soatto, 2003). 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 (2001).

This novel motion segmentation framework overcomes the drawbacks of our previous approach. Firstly, all normalizations comprised in the evolution equation are derived in a consistent manner by minimizing the proposed functional. Secondly, the level set formulation permits the segmentation of several (possibly multiply connected) objects, based on their relative motion. The present paper extends work which was presented on various conferences (Cremers, 2003a, 2003b; Cremers and Soatto, 2003).

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. We present two alternative implementations of the proposed functional: an explicit spline-based formulation in Section 6 and an implicit multiphase level set formulation in Section 7. Numerical results of the proposed framework for simulated ground truth data and real-world image sequences are given in Section 8. Sections 9 and 10 provide a conclusion and a discussion of the limitations of our framework.

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

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

with respect to the motion field v . For a related Bayesian formulation of motion segmentation in the discrete case, we refer to Boutheymy and Francois (1993).

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 (1981), 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.²

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:

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

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 (Bigün et al., 1991). For derivations of alternative likelihood functions from generative models of the image formation process and associated noise models, we refer to Nestares et al. (2000), Weiss and Fleet (2001) and Cremers and Yuille (2003).

To account for vanishing gradient, we regularize expression (5) by replacing

$$|\nabla_3 f(x)| \longrightarrow |\nabla_3 f(x)| + \epsilon \quad (6)$$

in the denominator. This guarantees that the probability is maximal if the gradient vanishes (in which case the optic flow constraint is also fulfilled), while not affecting the result for gradients much larger than ϵ . The regularizing constant ϵ can be interpreted as the noise scale for the gradient. As long as ϵ is chosen sufficiently small, we did not find a noticeable influence of its precise value in numerical implementations.

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 :

$$v(x) = \{v_i, \text{ if } x \in \Omega_i\} \quad (7)$$

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 $\mathcal{L}(C)$ of the boundary C separating these regions:

$$\mathcal{P}(v) \propto \exp(-v \mathcal{L}(C)) \quad (8)$$

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. Kass et al., 1988; Mumford and Shah, 1989). As we shall see in the next section, the choice of velocity representation in (7) combined with the prior in (8) will transform the motion estimation framework into one of motion segmentation.

3. Variational 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 Eq. (3) can be written as:

$$\begin{aligned} \mathcal{P}(\nabla_3 f | v) &= \prod_{x \in \Omega} \mathcal{P}(\nabla_3 f(x) | v(x))^h \\ &= \prod_{i=1}^n \prod_{x \in \Omega_i} \mathcal{P}(\nabla_3 f(x) | v_i)^h, \end{aligned} \quad (9)$$

where $h = dx$ denotes the pixel size of the discretization of Ω . The first step is based on the assumptions that the velocity affects the spatio-temporal gradient only

locally, and that the gradient measurements at different locations are independent. And the second step is based on the discretization of the velocity field given in (7).

With the prior probability (8), maximizing the conditional probability (3) with respect to the velocity field v therefore amounts to

$$\begin{aligned} \max_v \mathcal{P}(v | \nabla_3 f) \\ = \max_{v_i, C} \left\{ e^{-v \mathcal{L}(C)} \prod_{i=1}^n \prod_{x \in \Omega_i} \mathcal{P}(\nabla_3 f(x) | v_i)^h \right\}. \end{aligned} \quad (10)$$

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

$$\begin{aligned} E(C, \{v_i\}) \\ = - \sum_{i=1}^n \int_{\Omega_i} \log(\mathcal{P}(\nabla_3 f(x) | v_i)) dx + v \mathcal{L}(C). \end{aligned} \quad (11)$$

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

$$E(C, \{v_i\}) = \sum_{i=1}^n \int_{\Omega_i} \frac{(v_i^t \nabla_3 f(x))^2}{|v_i|^2 |\nabla_3 f(x)|^2} dx + v \mathcal{L}(C). \quad (12)$$

Let us make the following remarks about this functional:

- The functional (12) can be considered an extension of the piecewise constant Mumford and Shah (1989) functional 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.
- Note that 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 (1981), our measure (5) of orthogonality introduces an additional normalization with respect to the length of the two vectors. In Section 5.3, we will argue that these normalization are essential in the case of motion *segmentation*, where differently moving regions need to be compared.

- The functional (12) 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 (1995), such a scale parameter is fundamental to all segmentation approaches.

4. Piecewise Parametric Motion Segmentation

Minimizing functional (12) 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:

$$v_i(x) = M(x) p_i, \quad (13)$$

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

$$M(x) = \begin{pmatrix} x_1 & x_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & x_2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (14)$$

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 (13) into the optic flow constraint (4) gives a relation which—again interpreted geometrically—states that 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

$M^t \nabla_3 f$:

$$P(\nabla_3 f | p_i) \propto \exp\left(-\frac{(p_i^t M^t \nabla_3 f)^2}{|p_i|^2 |M^t \nabla_3 f|^2}\right). \quad (15)$$

The corresponding generalization of functional (12) from piecewise constant to piecewise parametric motion segmentation is given by:

$$E(C, \{p_i\}) = \sum_i \int_{\Omega_i} \frac{|p_i^t M^t \nabla_3 f|^2}{|p_i|^2 |M^t \nabla_3 f|^2} dx + \nu \mathcal{L}(C). \quad (16)$$

5. Energy Minimization

The functional (16) is of the form

$$E(C, \{p_i\}) = \sum_{i=1}^n \int_{\Omega_i} \frac{p_i^t T(x) p_i}{|p_i|^2} dx + \nu \mathcal{L}(C), \quad (17)$$

where, for notational simplification, we have introduced the matrix

$$T(x) = \frac{\nabla_3 f M^t M \nabla_3 f^t}{|M^t \nabla_3 f|^2}, \quad (18)$$

again regularized as done in (6).

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 $\{p_i\}$. Particular representations of the boundary C will be specified in Sections 6 and 7.

5.1. An Eigenvalue Problem for the Motion Parameters

The functional (17) can be further simplified:

$$E(C, \{p_i\}) = \sum_{i=1}^n \frac{p_i^t T_i p_i}{|p_i|^2} dx + \nu \mathcal{L}(C), \quad (19)$$

with a set of matrices

$$T_i = \int_{\Omega_i} T(x) dx, \quad (20)$$

with T given in (18). 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:

$$p_i = \arg \min_p \frac{p^t T_i p}{p^t p}. \quad (21)$$

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 in (20). 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. Biguen et al., 1991; Jepson and Black, 1993). Unal et al. (to appear) proposed in the context of tracking an alternative solution which integrates the motion information along the boundaries of the current segmentation.

5.2. Motion Competition

Conversely, for fixed motion models p_i , a gradient descent on the energy (17) for the boundary C results in the evolution equation:

$$\frac{\partial C}{\partial t} = -\frac{\partial E}{\partial C} = (e_j - e_k)n - v \frac{d\mathcal{L}(C)}{dC}, \quad (22)$$

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

$$e_i = \frac{p_i^t T p_i}{p_i^t p_i} = \frac{p_i^t \nabla_3 f M^t M \nabla_3 f^t p_i}{|p_i|^2 |M^t \nabla_3 f|^2} \quad (23)$$

is an energy density.

Note that we have neglected in the evolution Eq. (22) 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 Jehan-Besson et al. (2003) is the focus of ongoing research.

The two terms in the contour evolution (22) 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.³ 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 (22) with energy densities

$$e_i = \frac{v_i^t \nabla_3 f \nabla_3 f^t v_i}{|v_i|^2 |\nabla_3 f|^2} \quad (24)$$

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 (1981) type likelihood would induce contour driving terms which do not include the normalizing denominator:

$$e_i = v_i^t \nabla_3 f \nabla_3 f^t v_i. \quad (25)$$

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. An Explicit Spline-Based Implementation

In order to minimize the motion competition functional (17), we need to specify an appropriate representation for the boundary C . In the following, we will present two alternative representations: an explicit spline-based representation and an implicit level set representation.

In this section, we propose an implementation of the contour evolution (22) with a closed spline curve of the form:

$$C: [0, 1] \times \mathbb{R}^+ \rightarrow \Omega, \quad C(s, t) = \sum_{i=1}^N p_i(t) B_i(s), \quad (26)$$

with quadratic periodic B-spline basis functions B_i and control points $p_i = (x_i, y_i)^t$. This representation allows for a computationally efficient implementation of the contour evolution, since the evolution Eq. (22) reduces to an evolution equation for a small number of spline control points $\{p_i\}_{i=1,\dots,N}$. The number N of control points defines a trade-off between the spatial resolution of the contour approximation and the speed of computation.

One difficulty of explicit contour parameterizations is the fact that control points may cluster in one point. This causes the normal vector to become ill-defined and consequently the evolution along the normal becomes unstable. To prevent this behavior, we use the length measure

$$\mathcal{L}(C) = \int \left(\frac{\partial C}{\partial s} \right)^2 ds, \quad (27)$$

which corresponds to the elastic energy used in the classical snake approach (Kass et al., 1988). As discussed in Cremers et al. (2002), minimizing this constraint enforces an equidistant spacing of control points which strongly improves the numerical stability. The contour evolution then reads:

$$\frac{\partial C}{\partial t} = (e_+ - e_-)n - v \frac{\partial^2 C}{\partial s^2}, \quad (28)$$

where the indices ‘+’ and ‘-’ denote the two regions neighboring the respective contour point. Inserting the spline definition (26), we obtain the equation:

$$\sum_i \dot{p}_i B_i = (e_+ - e_-)n - v \sum_j p_j B_j'' \quad \forall s \in [0, 1], \quad (29)$$

where $\dot{\cdot}$ and \prime denote derivatives with respect to t and s , respectively.

By projecting Eq. (29) onto the basis functions $\{B_k\}_{k=1,\dots,n}$, we obtain a set of linear differential equations. The resulting evolution of spline control points $p_i(t)$ is given by:

$$\dot{p}_i = \sum_k \mathbf{B}^{-1}_{ik} \left[\int (e_+ - e_-)n B_k ds - v \sum_j p_j \tilde{B}_{jk} \right]. \quad (30)$$

Here \mathbf{B} denotes the matrix of basis function overlap integrals $B_{ik} = \int B_i B_k ds$, and $\tilde{B}_{kj} = \int B_j B_k'' ds$. The

first term in the brackets integrates the motion competition term over the part of the boundary affected by control point p_k , while the second term enforces an equidistant spacing of control points.

In practice, we iterate this gradient descent for the control points $p_i(t)$, in alternation with an update of the motion models according to (21).

7. A Multiphase Level Set Implementation

Although they are computationally efficient, explicit contour representations have a number of drawbacks. Firstly, one needs to take care of a regridding of control points which are not intrinsic to the contour. And secondly, the contour topology is fixed, such that no contour splitting or merging is possible, unless it is modeled explicitly by some (inevitably) heuristic method (cf. McInerney and Terzopoulos, 1995; Delingette and Montagnat, 2000).

An alternative are implicit level set representations of the boundary (Osher and Sethian, 1988). Level set based contour representations have become a popular framework in image segmentation (cf. Caselles et al. 1995; Kichenassamy et al., 1995; Chan and Vese, 2001), because they do not depend on a particular choice of parameterization, and because they do not restrict the topology of the evolving interface. 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 (2001), we will first present a two-phase level set model for the motion competition functional (17) with a single level set function ϕ . This model is subsequently extended to a multi-phase model with a vector-valued level set function.

7.1. 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 (17) be represented as the zero level set of a function $\phi : \Omega \rightarrow \mathbb{R}$:

$$C = \{x \in \Omega \mid \phi(x) = 0\}. \quad (31)$$

With the Heaviside step function

$$H(\phi) = \begin{cases} 1 & \text{if } \phi \geq 0 \\ 0 & \text{if } \phi < 0 \end{cases}, \quad (32)$$

the energy (17) can be embedded by the following *two-phase functional*:

$$\begin{aligned} E(p_1, p_2, \phi) = & \int_{\Omega} \frac{p_1^t T p_1}{|p_1|^2} H(\phi) dx \\ & + \int_{\Omega} \frac{p_2^t T p_2}{|p_2|^2} (1 - H(\phi)) dx \\ & + \nu \int_{\Omega} |\nabla H(\phi)| dx. \end{aligned} \quad (33)$$

The first two terms in (33) 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 (33) 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 (33) with respect to the motion vectors p_1 and p_2 results in the eigenvalue problem:

$$p_i = \arg \min_v \frac{v^t T_i v}{v^t v}, \quad (34)$$

for the matrices

$$\begin{aligned} T_1 &= \int_{\Omega} T(x) H(\phi) dx \quad \text{and} \\ T_2 &= \int_{\Omega} T(x) (1 - H(\phi)) dx. \end{aligned} \quad (35)$$

The solution of (34) is given by the eigenvectors corresponding to the smallest eigenvalues of T_1 and T_2 , normalized such that its last component is 1.

(b) *Evolution of the Level Set Function.*

Conversely, for fixed motion vectors, the gradient descent on the functional (33) for the level set function ϕ is given by:

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[\nu \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + e_2 - e_1 \right], \quad (36)$$

with the energy densities e_i given in (23). As suggested in Chan and Vese (2001), we implement the

Delta function $\delta(\phi) = \frac{d}{d\phi} H(\phi)$ by a smooth approximation of finite width τ :

$$\delta_{\tau}(s) = \frac{1}{\pi} \frac{\tau}{\tau^2 + s^2}. \quad (37)$$

Thereby the update of ϕ is not restricted to the areas of zero-crossing, but rather spread out over a band of width τ around it. Depending on the size of τ , this permits to detect interior motion boundaries. This will be demonstrated in Section 8.2.

7.2. 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.

One approach, investigated e.g. in Zhao et al. (1996) and Samson et al. (2000), is to represent each phase by a different level set function ϕ_i : $\Omega_i = \{x \in \Omega \mid \phi_i(x) \geq 0\}$. Although this approach permits to overcome the above limitations, it has two disadvantages: Firstly, it is computationally expensive to represent a large number of phases by a separate level set function for each phase. And secondly, one needs to suppress the formation of vacuum and overlap regions by introducing additional energy terms.

An elegant alternative to model multiple phases was proposed by Chan and Vese (2001). They introduce a more compact representation of up to n phases which needs only $m = \log_2(n)$ level set functions.⁴ Moreover, by definition, it generates a partition of the image plane and therefore does not suffer from overlap or vacuum formation. We will adopt this representation which shall be detailed in the following.

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

$$\Phi = (\phi_1, \dots, \phi_m) \quad (38)$$

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:

$$R = \{x \in \Omega \mid H(\Phi(x)) = \text{constant}\}. \quad (39)$$

In analogy to the corresponding level set formulation of the Mumford-Shah functional (Chan and Vese, 2001), we propose to replace the two-phase functional (33) by the *multiphase functional*:

$$E(\{p_i\}, \Phi) = \sum_{i=1}^n \int_{\Omega} \frac{p_i^t T p_i}{|p_i|^2} \chi_i(\Phi) dx + \nu \sum_{i=1}^n \int_{\Omega} |\nabla H(\phi_i)| dx, \quad (40)$$

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 (33).

For the purpose of illustration, we explicitly give the functional for the case of $n = 4$ phases:

$$\begin{aligned} E(\{p_i\}, \Phi) = & \int_{\Omega} \frac{p_{11}^t T p_{11}}{|p_{11}|^2} H_1 H_2 dx \\ & + \int_{\Omega} \frac{p_{10}^t T p_{10}}{|p_{10}|^2} H_1 (1 - H_2) dx \\ & + \int_{\Omega} \frac{p_{01}^t T p_{01}}{|p_{01}|^2} (1 - H_1) H_2 dx \\ & + \int_{\Omega} \frac{p_{00}^t T p_{00}}{|p_{00}|^2} (1 - H_1) (1 - H_2) dx \\ & + \nu \int_{\Omega} |\nabla H_1| dx + \nu \int_{\Omega} |\nabla H_2| dx, \end{aligned} \quad (41)$$

where we have used the short-hand notation

$$H_i \equiv H(\phi_i(x)). \quad (42)$$

Minimization of this functional with respect to the motion vectors $\{p_i\}$ for fixed Φ results in the eigenvalue problems:

$$p_i = \arg \min_v \frac{v^t T_i v}{v^t v}, \quad (43)$$

with four matrices T_i given by

$$\begin{cases} T_{11} = \text{mean}(T) & \text{in } \{\phi_1 \geq 0, \phi_2 \geq 0\} \\ T_{10} = \text{mean}(T) & \text{in } \{\phi_1 \geq 0, \phi_2 < 0\} \\ T_{01} = \text{mean}(T) & \text{in } \{\phi_1 < 0, \phi_2 \geq 0\} \\ T_{00} = \text{mean}(T) & \text{in } \{\phi_1 < 0, \phi_2 < 0\} \end{cases} \quad (44)$$

Conversely, for fixed velocity vectors, the evolution equations for the two level set functions are given by:

$$\begin{aligned} \frac{\partial \phi_1}{\partial t} = & \delta(\phi_1) \left[\nu \nabla \left(\frac{\nabla \phi_1}{|\nabla \phi_1|} \right) + (e_{01} - e_{11}) H_2 \right. \\ & \left. + (e_{00} - e_{10})(1 - H_2) \right], \end{aligned} \quad (45)$$

$$\begin{aligned} \frac{\partial \phi_2}{\partial t} = & \delta(\phi_2) \left[\nu \nabla \left(\frac{\nabla \phi_2}{|\nabla \phi_2|} \right) + (e_{10} - e_{11}) H_1 \right. \\ & \left. + (e_{00} - e_{01})(1 - H_1) \right], \end{aligned}$$

with the energy densities e_i defined in (23).

7.3. Redistancing

During their evolution according to Eqs. (36) or (45), the level set functions ϕ_i generally grow to very large positive or negative values in the respective areas of the input image corresponding to a particular motion hypothesis. At the zero crossings, they rise steeply, the gradient can become arbitrarily large. Indeed, there is nothing in the level set formulation of Chan and Vese (2001) 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.

Many people have advocated the use of a redistancing procedure in the evolution of level set functions to constrain the slope of ϕ to $|\nabla \phi| = 1$, see also Gomes and Faugeras (2000). In order to reproject the evolving level set function to the space of distance functions, we intermittently iterate several steps of the redistancing equation (Sussman et al., 1994):

$$\frac{\partial \phi}{\partial t} = \text{sign}(\hat{\phi}) (1 - |\nabla \phi|), \quad (46)$$

where $\hat{\phi}$ denotes the level set function before redistancing. Note that this transformation does not affect the motion boundaries given by the zero-crossing of ϕ .

Although this regularization is optional in the proposed level set model—see also Chan and Vese (2001)—it improves and accelerates the convergence of the boundary evolution. We applied the redistancing procedure in all experiments except the one in Section 8.2. Since the data term given by the image motion information dominates the evolution of the boundary, we found this simple redistancing process to be

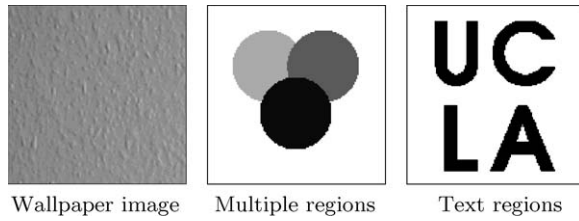


Figure 1. Data for ground truth experiments. Specific image regions of the wallpaper shot (left) are artificially translated to generate input data.

sufficiently accurate for our application. Therefore we did not revert to more elaborate iterative redistancing schemes such as the one presented in Sussman and Fatemi (1999).

8. Numerical Results

8.1. Ground Truth Experiments

In order to verify the spatial precision of the level set based motion segmentation approach introduced in Section 7, 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 Fig. 1. The input frames and the computed segmentations are (or will be made) available on the first author's web page.

We determined the spatio-temporal image gradient from two consecutive images and specified a particular initialization of the boundary. We minimized the functional (40) by alternating the three fractional steps of:

- updating the motion models for all phases by solving the corresponding eigenvalue problem (43),
- evolving the level set functions according to (36) or (45),
- and redistancing the level set functions according to (46).

For all experiments, we show the evolving motion boundaries (and in the first case also the corresponding motion estimates) superimposed onto the ground truth region information. It should be noted that in these experiments the objects cannot be distinguished from the background based on their appearance, as they corre-

sponds to homogeneously textured parts of the wallpaper. All results are obtained *exclusively* on the basis of the motion information.

8.1.1. Segmenting Several Motion Phases. In this experiment, we demonstrate an application of the four-phase model (41) 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 Fig. 1, left side, with three regions (shown in Fig. 1, right side) moving away from the center. The upper two regions move by a factor 1.4 faster than the lower region.

Figure 2 shows several steps in the minimization of the functional (41). 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

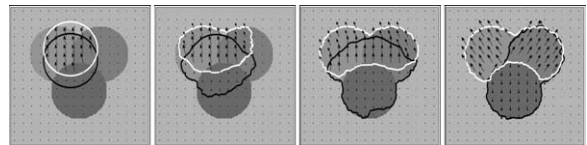


Figure 2. Segmenting multiple moving regions. The two input images show the wallpaper of Fig. 1, left side, with three circular regions moving away from the center. The magnitude of the velocity of the upper two regions is 1.4 times larger than that of the bottom region. Superimposed on the true region 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 four-phase model (41) with parameters $\nu = 0.05$, $\tau = 2$ with respect to the level set functions and the motion vectors. Note that in the final solution, the two boundaries clearly separate the four phases corresponding to the three moving regions and the static background.

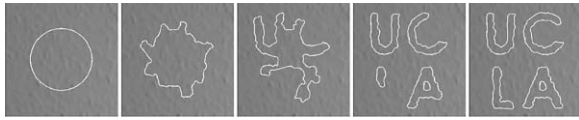


Figure 3. Accurate motion segmentation. Contour evolution obtained with functional (33) and parameter values $\nu = 0.06$, $\tau = 1$, superimposed on one of the two input frames. The input images show the text region (Fig. 1, right side) of the wallpaper moving to the right and the remainder moving 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.

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*.

8.1.2. Accurate Motion Segmentation Without Features.

In the previous examples, the moving regions were of fairly simple shape. The following example shows that one can generate spatially accurate segmentation results exploiting only motion information, even when using image sequences that exhibit little intensity variation or salient features. Figure 3 shows segmentation the contour evolution generated by minimizing functional (12). The input data consists of two wall paper images with the text region (Fig. 1, right side) moving to the right and the remainder of the 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 Harris corner points.

8.2. Detecting Interior Motion Boundaries

Many level set based boundary evolutions propagate the contour only locally and are therefore not able to detect boundaries which are not connected to the evolving interface. Due to the finite width τ of the Delta function approximation in (36), however, the embed-

ding surface is evolved in a band of width τ around its zero crossing. As a consequence, for sufficiently large values of τ this permits to also detect boundaries away from the evolving interface. We will demonstrate this property by segmenting a multiply-connected object based on its relative motion, once with an initialization which intersects the true motion boundaries and once with an initialization entirely outside the object of interest. The segmentation of the latter requires the detection of an interior boundary.

The two input images show a roll of adhesive tape moving on a newspaper. Figure 4, top row, shows the initial contour and the contour evolution obtained by minimizing the two-phase functional (33), superimposed on one of the two consecutive frames. The figures in the second row show the corresponding evolution of the embedding surface ϕ , underlying the contour evolution. It explains the change of contour topology from the third to the fourth image.

Note that ϕ is less negative in image regions where the newspaper is not sufficiently structured—these

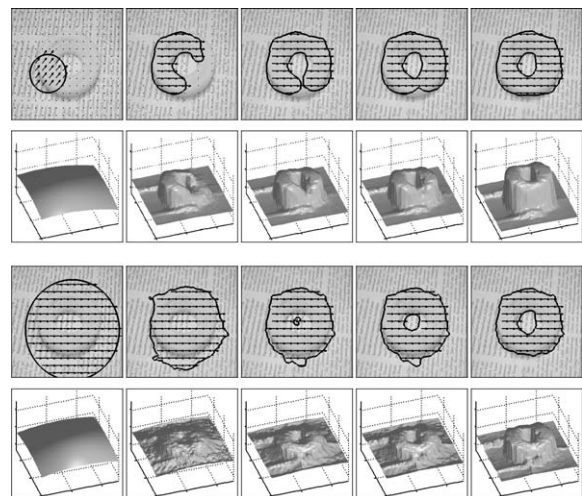


Figure 4. Motion segmentation with the two-phase functional (33) and parameters $\nu = 1.0$, $\tau = 100$ for two different initializations. 1st and 3rd row: One of the two input images (showing a roll of tape moving on a newspaper) with the evolving contour and the estimated motion superimposed. Note that the object of interest is hardly distinguishable from the background on the basis of its appearance. Yet, minimization of the functional (33) generates both a segmentation of the image plane and an estimate of the motion in each region. 2nd and 4th row: Corresponding evolution of the embedding surface ϕ . Note that in the upper example, the evolving surface induces a change of the contour topology from the third to the fourth image. The lower example demonstrates the detection of an interior boundary from the second to the third frame.

areas are less easily ascribed to one or the other motion hypothesis. In order to visualize this effect of a lack of intensity variation on the evolution of the level set function, we left out the redistancing procedure introduced in Section 7.3 since the latter enforces a slope of 1.

Figure 4, bottom rows, show the same segmentation for a different initialization. In this case, the moving ring lies entirely inside the initial contour. The numerical scheme is capable of detecting the interior motion boundary, as is apparent in the transition from the second to the third frame. Yet, there is a trade-off: While the ability of the scheme to detect interior boundaries improves with the smoothing scale τ of the delta function in (37), the accuracy of the segmentation clearly degrades for larger values of τ . Moreover, the capacity to detect interior boundaries may be suppressed by very strong redistancing.

Due to the region-based (rather than edge-based) formulation of our approach, the contour converges over fairly large distances. At the same time, the final segmentation is quite accurate, given that it is purely based on the motion information. Those image areas which do not show sufficient gray value structure to generate a motion estimate will be ascribed to one or the other motion hypothesis according to the boundary regularization.

These observations reflect the fact that motion-based image segmentation is fundamentally different from intensity-based segmentation: One can always ascribe an intensity to a given image pixel, but a velocity component can only be associated with it if there is sufficient gray value variation along a given direction. This fundamental limitation is commonly referred to as the *aperture problem*. In the motion competition framework it is dealt with by the boundary regularization and by the fact that motion is only estimated for entire regions (never locally).

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 analogue. 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).⁵ 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 5, top, shows the boundary evolution obtained by minimizing the two-phase model of Chan and Vese (2001) 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

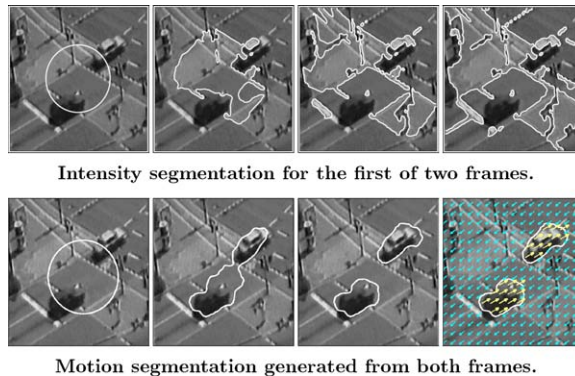


Figure 5. Intensity segmentation versus motion segmentation. Two consecutive input frames show two cars moving to the top right, and the background moving to the bottom left. Top row: Segmentation of the first frame from a traffic scene according to the two-phase level set model of the piecewise constant Mumford-Shah functional, as introduced by Chan and Vese (2001). 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 (33) with parameters $\nu = 1.5$, $\tau = 5$, 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.

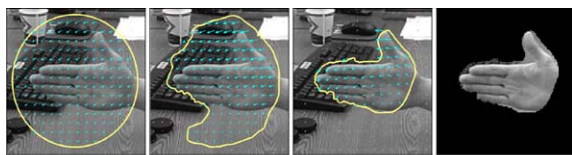
darker parts of the street, whereas the car in the upper right is split into its brighter and darker parts.

In this example, the cars and the street are moving according to different motion models. The motion competition framework exploits this property. Figure 5, bottom, show the contour evolution generated by minimizing the motion segmentation functional (33) and the corresponding motion estimates superimposed on the first frame.

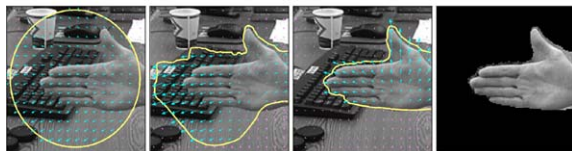
The contour evolution generated by motion competition is fundamentally different from the one generated by its gray value analogue. The energy minimization simultaneously generates a fairly accurate segmentation of the two cars and an estimate of the motion of cars and background. Minor discrepancies 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 (17) allows to segment piecewise affine motion fields. In particular, this class of motion models includes rotation and expansion/contraction. Figure 6 shows 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



Motion segmentation of a hand rotating around the wrist.



Motion segmentation of a hand moving toward the camera.

Figure 6. Piecewise affine motion segmentation. Functional (17) 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 ($v = 8 \cdot 10^{-5}$, $\tau = 2$).

or divergent motion. The images on the right demonstrate that the objects of interest can be extracted from a fairly complex background based exclusively on their motion.

8.5. An Application: Tracking with Multiple Motion

In the present section, we evaluate the explicit scheme introduced in Section 6 on the Avengers sequence.⁶ In this movie sequence, moving cars are captured by a moving camera.

Figure 7 shows the segmentation results obtained on the frames 18 through 34. We fixed an initial contour, as shown in Fig. 7, top left. For each pair of consecutive images in the sequence, we then determined the spatio-temporal derivative and performed a fixed number of steps in the minimization of energy (12) with $v = 2.0$, alternating the motion estimation (34) and the contour evolution (30). Despite the model hypothesis of constant motion per region, the segmentation is fairly robust to non-translatory motion. Once the car starts turning the segmentation slowly degrades—see the last images in Fig. 7.

Minimizing energy (12) simultaneously generates a segmentation of the image plane and an estimate for the motion in the separate regions. The motion estimated for the first two frames in the sequence is shown in Fig. 9. Both the car and the background

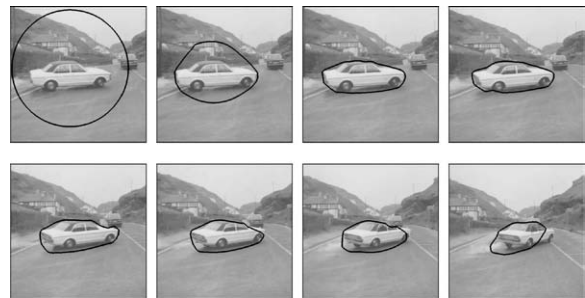


Figure 7. Motion segmentation for the frames 18–34 from the Avengers sequence: Contour evolution for the functional (12) with an explicit contour (using 100 spline control points and $v = 2.0$) initialized as shown in the top left image. The first three images show the evolution of the contour for the first pair of frames, the following images show the segmentation results obtained for consecutive frames. Both the car and the background are moving. Despite the model hypothesis of constant motion per region, the segmentation is fairly robust to non-translatory motion. Yet, the segmentation degrades once the car starts moving perpendicular to the viewing plane (bottom right). The algorithm runtime was approximately 2 seconds per frame.

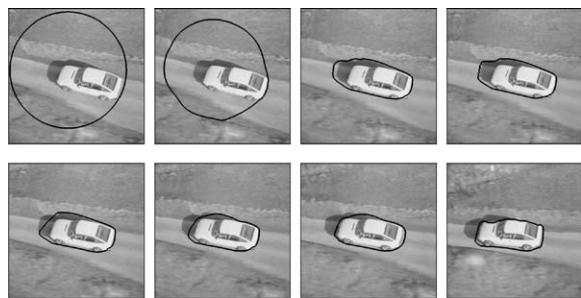


Figure 8. Motion segmentation for the frames 35–45 from the Avengers sequence. The contour is initialized as shown in the top left, then the minimization of (12) with $\nu = 2.0$ is iterated for several steps on each pair of consecutive frames (the first three images showing frame 35). Since the computation relies on only two consecutive frames, there is no hypothesis of motion continuity. Therefore the approach can be used for segmenting (temporally) discontinuous motion and for tracking.

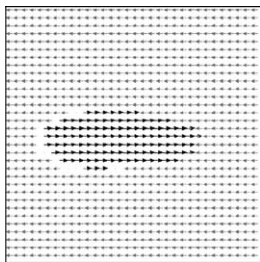


Figure 9. Motion estimate generated by minimizing energy (12) for the first two frames from Fig. 7. The corresponding segmentation is shown in Fig. 7, third image. Note that both the car and the background are moving at different velocities.

are moving, with velocities of different direction and magnitude.

Figure 8 shows similar results for the frames 35–45 of the Avengers sequence. Interestingly, the co-moving shadow is attributed to the car in the first few frames, and attributed to the background in subsequent frames. Since the street is visible within the shadow area, it is indeed unclear whether the shadow region should be associated with the car or with the street. The change from one hypothesis to the other may be due to subtle changes in brightness and background contrast.

9. Conclusion

Starting from a Bayesian formulation of motion estimation, we derived a novel variational framework for segmenting the image plane into a set of regions of parametric motion. 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 piecewise parametric motion field favoring motion boundaries of minimal length.

The functional depends on parametric velocity models for a set of regions and the boundary separating them. It can be considered as an extension of the Mumford-Shah functional from intensity segmentation to motion segmentation. The only free parameter in the functional is the fundamental scale parameter intrinsic to all segmentation schemes.

We presented two alternative implementations of the motion discontinuity set. The first is an explicit spline-based representation suited for tracking objects based on their relative motion. The evolution of the boundary reduces to an evolution of a small number of spline control points. Moreover, we proposed an implementation of the motion segmentation functional in a multiphase level set framework. The resulting model has the following properties:

- The minimization of a *single* functional with respect to its dynamic variables jointly solves the problems of motion estimation and motion segmentation. It generates a segmentation of the image plane in terms of piecewise parametric motion.
- The *implicit* representation of the motion discontinuity set does not depend on a particular choice of parameterization. It allows for topological changes of the boundary such as splitting or merging. The *multi-phase* formulation permits a segmentation of the image plane into several (possibly multiply-connected) motion phases.
- Minimizing the proposed functional is straightforward. It 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, the motion boundaries tend to converge over fairly large spatial distances.
- Segmentation and motion estimates are generated for two consecutive frames of an image sequence. Therefore the approach is in principle amenable to real-time implementations and tracking.
- Once the motion of objects deviates from the model hypothesis, then the segmentation gradually degrades.

We demonstrated these properties by experimental results, both on simulated ground truth data and on real-world sequence data.

10. Limitations and Ongoing Research

The proposed motion segmentation framework exhibits several limitations. We want to end our presentation by pointing out the major limitations and hint at possible solutions to address these. Some of these can be integrated into our framework and this is the focus of ongoing research. In particular:

- The model is based on the assumption of small motion. This means that motion should not exceed four pixels between frames. To address the segmentation of larger motion, one can revert to multi-scale and multi-resolution implementations of the proposed functional. This is a common practice in the motion community (cf. Odobez and Bouthemy, 1995; Memin and Perez, 2002; Brox et al., 2004).
- Our model is based on the assumption that objects do not change their brightness throughout time. This assumption may be violated due to lighting changes, speckle noise or non-Lambertian reflectance such as the specularities on a car. While we believe that our region-based formulation can handle some deviation from this assumption, it is possible to extend our framework by incorporating constraints on gradient constancy (cf. Brox et al., 2004).
- While the multi-phase level set scheme is in principle able to handle an arbitrary number of motion phases, it has several drawbacks. Firstly, it is a local scheme, therefore it is not well-suited to deal with new objects entering the scene in locations very far from the evolving boundary. Secondly, the computational complexity does not scale well with the number of regions—the number of necessary level set functions scales with the logarithm of the number of possible motion models. Certain phases may disappear during the evolution, yet further studies will investigate how well the proposed framework can deal with a larger number of motion phases. In particular, recent developments in the level set community may present alternative more efficient solutions (cf. Lie et al., 2003).
- Our framework allows the segmentation of images in terms of piecewise parametric motion fields. In numerous practical applications—such as the segmentation of a moving heart—standard parametric motion models (such as affine motion) are not applicable. A possible extension would be to learn parametric models of the specific motion from presegmented data and to impose these problem-

specific parametric models in the motion segmentation framework.

Acknowledgments

We thank Christoph Schnörr, Joachim Weickert, Alan Yuille, Song-Chun Zhu, Paolo Favaro, Etienne Memin, Jean-Marc Odobez, Thomas Brox and Andrés Bruhn for fruitful discussions. We also thank the three reviewers for detailed comments on the manuscript. This research was supported by ONR N00014-02-1-0720/N00014-03-1-0850 and AFOSR F49620-03-1-0095/E-16-V91-G2.

Notes

1. Since we are only concerned with two consecutive frames from a sequence, we will drop the time coordinate in the notation of the velocity field.
2. The optic flow constraint is violated if the brightness constancy assumption does not hold or if the inter-frame motion is larger than a few pixels. We will not treat these cases here. The first case can be addressed by additional assumptions such as gradient constancy (cf. Brox et al., 2004), the second case can be handled by multi-scale processing (cf. Odobez and Bouthemy, 1995; Brox et al., 2004). Our segmentation framework can be extended accordingly.
3. In particular, the functionals (12) and (17) are invariant to global scale transformations of the intensity: $f \rightarrow \gamma f$.
4. During the optimization certain phases may disappear such that the final segmentation may consist of less than n phases.
5. http://i21www.ira.uka.de/image_sequences/
6. We thank P. Bouthemy and his group for providing us with the image data from the Avengers sequence.

References

- Ayer, S. and Sawhney, H.S. 1995. Layered representation of motion video using robust maximum likelihood estimation of mixture models and MDL encoding. In *Proc. of the Int. Conf. on Comp. Vis.*, Boston, USA, pp. 777–784.
- Bigün, J., Granlund, G.H., and Wiklund, J. 1991. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE PAMI*, 13(8):775–790.
- Black, M.J. 1994. Recursive non-linear estimation of discontinuous flow fields. In *Proc. of the Europ. Conf. on Comp. Vis.*, J.O. Eklundh, (Ed.), vol. 800 of *LNCIS*, Springer-Verlag, pp. 138–145.
- Black, M.J. and Anandan, P. 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Comp. Vis. Graph. Image Proc.: IU*, 63(1):75–104.
- Bouthemy, P. and Francois, E. 1993. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *Int. J. of Computer Vision*, 10(2):157–182.
- Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. 2004. High accuracy optical flow estimation based on a theory for warping.

- In *European Conf. on Computer Vision*, T. Pajdla and V. Hlavac (Eds.), vol. 3024 of *LNCS*, Springer, Prague, pp. 25–36.
- Caselles, V. and Coll, B. 1996. Snakes in movement. *SIAM J. Numer. Anal.*, 33:2445–2456.
- Caselles, V., Kimmel, R., and Sapiro, G. 1995. Geodesic active contours. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, Boston, USA, pp. 694–699.
- Chan, T. and Vese, L. 2001. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277.
- Cremers, D. 2003a. A multiphase level set framework for variational motion segmentation. In *Int. Conf. on Scale Space Theories in Computer Vision*, L. Griffith (Ed.), vol. 2695 of *LNCS*, Isle of Skye, Springer, pp. 599–614.
- Cremers, D. 2003b. A variational framework for image segmentation combining motion estimation and shape regularization. In *IEEE Conf. on Comp. Vis. and Patt. Recog.*, C. Dyer and P. Perona (Eds.), vol. 1, pp. 53–58.
- Cremers, D. and Schnörr, C. 2003. Statistical shape knowledge in variational motion segmentation. *Image and Vision Computing*, 21(1):77–86.
- Cremers, D. and Soatto, S. 2003. Variational space-time motion segmentation. In *IEEE Int. Conf. on Computer Vision*, B. Triggs and A. Zisserman (Eds.), Nice, vol. 2, pp. 886–892.
- Cremers, D., Tischhäuser, F., Weickert, J., and Schnörr, C. 2002. Diffusion Snakes: Introducing statistical shape knowledge into the Mumford–Shah functional. *Int. J. of Computer Vision*, 50(3):295–313.
- Cremers, D. and Yuille, A.L. 2003. A generative model based approach to motion segmentation. In *Pattern Recognition*, B. Michaelis and G. Krell (Eds.), vol. 2781 of *LNCS*, Springer, Magdeburg, pp. 313–320.
- Delingette, H. and Montagnat, J. 2000. New algorithms for controlling active contours shape and topology. In *Proc. of the Europ. Conf. on Comp. Vis.*, D. Vernon (Ed.), vol. 1843 of *LNCS*, Springer, pp. 381–395.
- Farneback, G. 1999. *Spatial Domain Methods for Orientation and Velocity Estimation*. PhD thesis, Dept. of Electrical Engineering, Linköpings universitet.
- Farneback, G. 2001. Very high accuracy velocity estimation using orientation tensors, parametric motion, and segmentation of the motion field. In *ICCV*, vol. 1, pp. 171–177.
- Gomes, J. and Faugeras, O.D. 2000. Level sets and distance functions. In *Proc. of the Europ. Conf. on Comp. Vis.*, D. Vernon (Ed.), vol. 1842 of *LNCS*, Dublin, Ireland, Springer, pp. 588–602.
- Heitz, F. and Bouthemy, P. 1993. Multimodal estimation of discontinuous optical flow using markov random fields. *IEEE PAMI*, 15(12):1217–1232.
- Horn, B.K.P. and Schunck, B.G. 1981. Determining optical flow. *A.I.*, 17:185–203.
- Jehan-Besson, S., Barlaud, M., and Aubert, G. 2003. DREAM2S: Deformable regions driven by an eulerian accurate minimization method for image and video segmentation. *Int. J. of Computer Vision*, 53(1):45–70.
- Jepson, A. and Black, M.J. 1993. Mixture models for optic flow computation. In *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, New York, pp. 760–761.
- Kass, M., Witkin, A., and Terzopoulos, D. 1988. Snakes: Active contour models. *Int. J. of Computer Vision*, 1(4):321–331.
- Kichenassamy, S., Kumar, A., Olver, P.J., Tannenbaum, A., and Yezzi, A.J. 1995. Gradient flows and geometric active contour models. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, Boston, USA, pp. 810–815.
- Konrad, J. and Dubois, E. 1992. Bayesian estimation of motion vector fields. *IEEE PAMI*, 14(9):910–927.
- Kornprobst, P., Deriche, R., and Aubert, G. 1999. Image sequence analysis via partial differential equations. *J. Math. Im. Vis.*, 11(1):5–26.
- Lie, J., Lysaker, M., and Tai, X.-C. 2003. A variant of the level set method and applications to image segmentation. Technical Report 03-50, Computational Applied Mathematics, UCLA, Los Angeles.
- Lucas, B.D. and Kanade, T. 1981. An iterative image registration technique with an application to stereo vision. In *Proc. 7th International Joint Conference on Artificial Intelligence*, Vancouver, pp. 674–679.
- Malladi, R., Sethian, J.A., and Vemuri, B.C. 1995. Shape modeling with front propagation: A level set approach. *IEEE PAMI*, 17(2):158–175.
- McInerney, T. and Terzopoulos, D. 1995. Topologically adaptable snakes. In *Proc. 5th Int. Conf. on Computer Vision*, IEEE Comp. Soc. Press: Los Alamitos, California, pp. 840–845.
- Memin, E. and Perez, P. 1998. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Trans. on Im. Proc.*, 7(5):703–719.
- Memin, E. and Perez, P. 2002. Hierarchical estimation and segmentation of dense motion fields. *Int. J. of Computer Vision*, 46(2):129–155.
- Morel, J.-M. and Solimini, S. 1995. *Variational Methods in Image Segmentation*. Birkhäuser: Boston.
- Mumford, D. and Shah, J. 1989. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685.
- Nagel, H.H. and Enkelmann, W. 1986. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE PAMI*, 8(5):565–593.
- Nestares, O., Fleet, D.J., and Heeger, D.J. 2000. Likelihood functions and confidence bounds for total-least-squares problems. In *Proc. Conf. Computer Vis. and Pattern Recog.*, vol. 1, Hilton Head Island, SC, pp. 760–767.
- Odobez, J.-M. and Bouthemy, P. 1995. Robust multiresolution estimation of parametric motion models. *J. of Visual Commun. and Image Repr.*, 6(4):348–365.
- Odobez, J.-M. and Bouthemy, P. 1998. Direct incremental model-based image motion segmentation for video analysis. *Signal Proc.*, 66:143–155.
- Osher, S.J. and Sethian, J.A. 1988. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. *J. of Comp. Phys.*, 79:12–49.
- Paragios, N. and Deriche, R. 2000. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE PAMI*, 22(3):266–280.
- Samson, C., Blanc-Féraud, L., Aubert, G., and Zerubia, J. 2000. A level set model for image classification. *Int. J. of Computer Vision*, 40(3):187–197.
- Schnörr, C. 1992. Computation of discontinuous optical flow by domain decomposition and shape optimization. *Int. J. of Computer Vision*, 8(2):153–165.
- Shi, J. and Malik, J. 1998. Motion segmentation and tracking using normalized cuts. In *Intl. Conf. on Comp. Vision*, Bombay, India.
- Sussman, M. and Fatemi, E. 1999. An efficient, interface-preserving level set redistancing algorithm and its application to interfacial

- incompressible fluid flow. *SIAM J. Sci. Comput.*, 20(4):1165–1191.
- Sussman, M., Smereka P., and Osher, S.J. 1994. A level set approach for computing solutions to incompressible twophase flow. *J. of Comp. Phys.*, 94:146–159.
- Unal, G., Krim, H., and Yezzi, A. (to appear). Fast incorporation of optic flow into active polygons. *IEEE Trans. on Image Processing*.
- Wang, J.Y.A. and Adelson, E.H. 1994. Representating moving images with layers. *IEEE Trans. on Image Processing*, 3(5):625–638.
- Weickert, J. and Schnörr, C. 2001. A theoretical framework for convex regularizers in PDE-based computation of image motion. *Int. J. of Computer Vision*, 45(3):245–264.
- Weiss, Y. 1997. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, Puerto Rico, pp. 520–527.
- Weiss, Y. and Fleet, D.J. 2001. Velocity likelihoods in biological and machine vision. In *Probabilistic Models of the Brain: Perception and Neural Function*, M.S. Lewicki, R.P.N. Rao, and B.A. Olshausen (Eds.), MIT Press, pp. 81–100.
- Yezzi, A. and Soatto, S. 2003. Stereoscopic segmentation. *Int. J. of Computer Vision*, 53(1):31–43.
- Zhao, H.-K., Chan, T., Merriman, B., and Osher, S. 1996. A variational level set approach to multiphase motion. *J. of Comp. Phys.*, 127:179–195.
- Zheng, H. and Blostein, S.D. 1995. Motion-based object segmentation and estimation. *IEEE Trans. on Image Processing*, 4(9):1225–1235.
- Zhu, S.C. and Yuille, A. 1996. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE PAMI*, 18(9):884–900.