A Multiphase Level Set Framework for Motion Segmentation

Daniel Cremers

Department of Computer Science
University of California at Los Angeles
http://www.cs.ucla.edu/~cremers

Abstract. We present a novel variational approach for segmenting the image plane into a set of regions of piecewise constant motion on the basis of only two consecutive frames from an image sequence.
To this end, we formulate the problem of estimating a motion field in the framework of Bayesian inference. Our model is based on a conditional probability for the spatio-temporal image gradient, given a particular velocity vector, and on a prior on the estimated motion field favoring motion boundaries of minimal length. The corresponding negative log likelihood is a functional which depends on motion vectors for a set of regions and on the boundary separating these regions. It can be considered an extension of the Mumford-Shah functional from intensity segmentation to motion segmentation.
We propose an implementation of this functional by a multiphase level set framework. Minimizing the functional with respect to its dynamic variables results in an evolution equation for a vector-valued level set function and in an eigenvalue problem for the motion vectors. Compared to most alternative approaches, we jointly solve the problems of segmentation and motion estimation by minimizing a *single* functional. Numerical results both for simulated ground truth experiments and for real-world sequences demonstrate the capacity of our approach to segment several – possibly multiply connected – objects based on their relative motion.

1 Introduction

Motion estimation from image sequences has a long tradition in computer vision. Two seminal variational methods were proposed by Horn and Schunck [11] and by Lucas and Kanade [15]. 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 flow field.

In practice, flow fields are usually not smooth. The boundaries of moving objects will correspond to discontinuities in the motion field. Such motion discontinuities have been modeled implicitly by non-quadratic robust estimators [2,17,14,28]. Other approaches tackled the problem of segmenting the motion field by treating the problems of motion estimation in disjoint sets and optimization of the motion boundaries separately [25,3,21,9]. Some approaches are based

L.D. Griffin and M. Lillholm (Eds.): Scale-Space 2003, LNCS 2695, pp. 599–614, 2003.

on Markov Random Field formulations and the EM algorithm (cf. [12,1,29]). Yet, as pointed out in [29], exact solutions to the EM algorithm are computationally expensive and therefore suboptimal approximations are employed. For certain tracking applications, it may also be sufficient to perform segmentation on the basis of temporal change detection [23]. Yet this approach does not extend to the cases of moving background and multiple motion considered here.

In [8], we presented a variational approach to motion segmentation with an explicit contour where both the motion estimation and the boundary optimization are derived from minimizing a *single* energy functional. Yet, 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. And secondly, due to the explicit representation of this motion boundary, the segmentation of multiple moving objects is not straight-forward.

In [6], we addressed these drawbacks by proposing a novel geometric interpretation of the optic flow constraint and by reverting to a two-phase level set representation of the motion boundary. Due to this geometric interpretation of the optic flow constraint, all normalizations of the boundary evolution forces are derived in a consistent manner by minimizing the proposed cost functional. And the level set formulation permits a segmentation of the image plane into multiple regions, each of which is associated with one of two motion models.

The present paper extends this work in several ways: Firstly, we formulate the problem of motion estimation in the framework of Bayesian inference to derive an energy functional which extends the Mumford-Shah functional [19] from gray value segmentation to motion segmentation. Secondly, we detail a multiphase level set implementation of this functional, which is based on the corresponding gray value model of Chan and Vese [5]. The multiphase formulation permits to segment an arbitrary number of differently moving regions. We show that minimization leads to an eigenvalue problem for the motion parameters, and to a gradient descent evolution for the level set functions embedding the motion discontinuities. Numerical results are demonstrated on simulated ground-truth experiments and on real world problems.

A related approach to motion segmentation was proposed in [16]. Firstly, our cost functional differs from theirs in that it includes normalizations of the residuals which we believe are important for comparing differently moving regions. As a consequence, most of our purely motion-based segmentations are highly accurate. Secondly, we use an efficient multiphase model which does not suffer from the formation of vacuum or overlap regions. Thirdly, the motion models in [16] are estimated in a separate process on the basis of "feature points".

This 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 4, we propose a multiphase level set formulation of the motion segmentation functional. In particular, we detail the special cases of a two-phase and a four-phase model. In Sections 5 and 6, we show numerical results obtained on simulated ground truth data and on real world image sequences. In Section 7, we end with a conclusion.

2 Motion Estimation as Bayesian Inference

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

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

Let

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

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

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

with respect to the motion field v.
To this end, we make the following assumptions:

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

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

  Except for locations where the spatio-temporal gradient vanishes, this constraint 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 β be the angle between the two vectors, then:

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

  By construction, this probability is independent of the length of the two vectors and monotonically increases the more orthogonal the two vectors are. We regularize this expression by replacing

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

  in the denominator. This guarantees that the probability is maximal if the gradient vanishes, while not affecting the result for gradients much larger than ϵ. As long as ϵ is chosen sufficiently small, we did not find a noticeable influence of its precise value in numerical implementations.

[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.

- We discretize the velocity field v by a set of disjoint regions $R_i \subset \Omega$ with constant velocity v_i:
$$v(x) = \{v_i, \text{ if } x \in R_i\} \tag{7}$$
Note that such a discretization in itself does not restrict the class of permissible motion fields, since each image pixel could be considered a separate region. We now assume the prior probability on the velocity field to only depend on the length of the boundary C separating these regions:
$$\mathcal{P}(v) \propto \exp\left(-\nu\,|C|\right) \tag{8}$$
In particular, this means that we do not make any prior assumptions on the velocity vectors v_i. Such a term would necessarily introduce a bias favoring certain velocities.

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:

$$\mathcal{P}(\nabla_3 f \,|\, v) = \prod_{x\in\Omega} \mathcal{P}(\nabla_3 f(x) \,|\, v(x))^h = \prod_{i=1}^{n} \prod_{x\in R_i} \mathcal{P}(\nabla_3 f(x) \,|\, v_i)^h\,, \tag{9}$$

where h denotes the grid size of the discretization of Ω. The first step is based on the assumption that the velocity affects the spatio-temporal gradient only locally. And the second step is based on the discretization of the velocity field given in (7).

With the formulas (5), (8) and (9), maximizing the conditional probability (3) is equivalent to minimizing its negative logarithm, which is given (up to a constant) by the energy functional:

$$\boxed{E(C,\{v_i\}) = \sum_{i=1}^{n} \int_{R_i} \frac{(v_i^t \nabla_3 f(x))^2}{|v_i|^2\,|\nabla_3 f(x)|^2}\,dx \;+\nu\,|C|.} \tag{10}$$

Let us make the following remarks about this functional:

- The functional (10) can be considered an extension of the piecewise constant Mumford-Shah functional [19] from the case of gray value segmentation to the case of motion segmentation. Rather than having a mean gray values f_i for each region R_i, we now have a homogeneous velocity vector v_i for each region R_i.
- Minimizing the functional (10) 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 we believe that

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 sensible 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 [11], our measure (5) of orthogonality introduces an additional normalization with respect to the length of the two vectors. We found this to be essential in the case of motion *segmentation*, where one needs to compare differently moving regions.
- The functional (10) contains only 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 [18], such a scale parameter is fundamental in all segmentation approaches.

4 A Multiphase Level Set Implementation

In order to minimize the functional (10), we need to specify an appropriate representation for the boundary C. In this paper, we choose an implicit level set representation of the boundary [22]. Level set based contour representations have become a popular framework in image segmentation (cf. [4,13,5]), 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 the work of Chan and Vese [5], we will first present a two-phase level set model for the functional (10) with a single level set function ϕ. This model is subsequently extended to a multi-phase model with a vector-valued level set function.

4.1 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 velocities v_1 and v_2. The general case of several velocities $\{v_i\}_{i=1,\dots,n}$ will be treated in the next subsection.

Let the boundary C in the functional (10) be represented as the zero level set of a function $\phi : \Omega \to \mathbb{R}$:

$$C = \{x \in \Omega \mid \phi(x) = 0\}. \tag{11}$$

Using the Heaviside step function

$$H(\phi) = \begin{cases} 1 \text{ if } \phi \geq 0 \\ 0 \text{ if } \phi < 0 \end{cases}, \tag{12}$$

and, for notational simplification, the matrix

$$T(x) = \frac{\nabla_3 f \; \nabla_3 f^t}{|\nabla_3 f|^2}, \tag{13}$$

again with the regularization (6) in numerical implementations, we can embed the motion energy (10) by the following *two-phase functional*:

$$E(v_1, v_2, \phi) = \int_\Omega \frac{v_1^t T v_1}{|v_1|^2} H(\phi) dx + \int_\Omega \frac{v_2^t T v_2}{|v_2|^2} \big(1 - H(\phi)\big) dx + \nu \int_\Omega \big|\nabla H(\phi)\big| dx. \tag{14}$$

This functional is now simultaneously minimized with respect to the velocity vectors v_1 and v_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) An Eigenvalue Problem for the Motion Vectors.

For fixed ϕ, minimization of the functional (14) with respect to the motion vectors v_1 and v_2 results in the eigenvalue problem:

$$v_i = \arg\min_v \frac{v^t \, M_i \, v}{v^t v}, \tag{15}$$

for the 3×3-matrices

$$M_1 = \int_\Omega T(x)\, H(\phi)\, dx \quad \text{and} \quad M_2 = \int_\Omega T(x)\, \big(1 - H(\phi)\big)\, dx. \tag{16}$$

The solution of (15) is given by the eigenvectors corresponding to the smallest eigenvalues of M_1 and M_2.

(b) Evolution of the Level Set Function.

Conversely, for fixed motion vectors, the gradient descent on the functional (14) for the level set function ϕ is given by:

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[\nu \, \text{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + e_2 - e_1 \right], \tag{17}$$

with the energy densities e_i given by

$$e_i(x) = \frac{v_i^t \, T(x) \, v_i}{v_i^t v_i} \tag{18}$$

As suggested in [5], we implement the Delta function $\delta(\phi) = \frac{d}{d\phi} H(\phi)$ by a smooth approximation of finite width τ:

$$\delta_\tau(s) = \frac{1}{\pi} \frac{\tau}{\tau^2 + s^2}. \tag{19}$$

Depending on the size of τ, this permits to detect interior motion boundaries.

4.2 The General Multiphase Model

The above approach to represent 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 vectors). 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 [30,24,16], is to represent each phase $R_i \subset \Omega$ by a different level set function ϕ_i: $R_i = \{x \in \Omega \,|\, \phi_i(x) \geq 0\}$. 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 alternative more elegant approach to model multiple phases was proposed by Chan and Vese in [5]. They introduce a more compact representation of up to n phases which needs only $m = \log_2(n)$ level set functions. 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.

Let $\Phi = (\phi_1, \ldots, \phi_m)$ be a vector level set function, with $\phi_i : \Omega \to \mathbb{R}$. Let $H(\Phi(x)) = (H(\phi_1(x)), \ldots, H(\phi_m(x)))$ 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 R_i defined by:

$$R = \{x \in \Omega \mid H\big(\Phi(x)\big) = \text{constant}\}. \tag{20}$$

In analogy to the case of the Mumford-Shah functional treated in [5], we propose to replace the two-phase functional (14) by the *multiphase functional*:

$$\boxed{E(\{v_i\}, \Phi) = \sum_{i=1}^{n} \int_{\Omega} \frac{v_i^t \, T \, v_i}{|v_i|^2} \, \chi_i(\Phi) \, dx \; + \; \nu \sum_{i=1}^{n} \int_{\Omega} \big|\nabla H(\phi_i)\big| \, dx,} \tag{21}$$

where χ_i denotes the indicator function for the region R_i. Note, that for $n=2$, this is equivalent to the two-phase model introduced in (14).

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

$$\begin{aligned} E(\{v_i\}, \Phi) = & \int_{\Omega} \frac{v_{11}^t T v_{11}}{|v_{11}|^2} \, H(\phi_1) H(\phi_2) \, dx \; + \int_{\Omega} \frac{v_{10}^t T v_{10}}{|v_{10}|^2} \, H(\phi_1)\big(1 - H(\phi_2)\big) \, dx \\ & + \int_{\Omega} \frac{v_{01}^t T v_{01}}{|v_{01}|^2} \, \big(1 - H(\phi_1)\big) H(\phi_2) dx + \int_{\Omega} \frac{v_{00}^t T v_{00}}{|v_{00}|^2} \, \big(1 - H(\phi_1)\big)\big(1 - H(\phi_2)\big) dx \\ & + \nu \int_{\Omega} \big|\nabla H(\phi_1)\big| \, dx \; + \nu \int_{\Omega} \big|\nabla H(\phi_2)\big| \, dx. \end{aligned} \tag{22}$$

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

$$v_i = \arg\min_v \frac{v^t M_i v}{v^t v}, \tag{23}$$

with four 3×3-matrices M_i given by

$$\begin{cases} M_{11} = \text{mean}(T) \text{ in } \{\phi_1 \geq 0, \phi_2 \geq 0\} \\ M_{10} = \text{mean}(T) \text{ in } \{\phi_1 \geq 0, \phi_2 < 0\} \\ M_{01} = \text{mean}(T) \text{ in } \{\phi_1 < 0, \phi_2 \geq 0\} \\ M_{00} = \text{mean}(T) \text{ in } \{\phi_1 < 0, \phi_2 < 0\} \end{cases} \tag{24}$$

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

$$\begin{aligned} \frac{\partial \phi_1}{\partial t} &= \delta(\phi_1)\left[\nu \,\text{div}\left(\frac{\nabla \phi_1}{|\nabla \phi_1|}\right) + (e_{01} - e_{11})\, H(\phi_2) + (e_{00} - e_{10})\, \big(1 - H(\phi_2)\big)\right], \\ \frac{\partial \phi_2}{\partial t} &= \delta(\phi_2)\left[\nu \,\text{div}\left(\frac{\nabla \phi_2}{|\nabla \phi_2|}\right) + (e_{10} - e_{11})\, H(\phi_1) + (e_{00} - e_{01})\, \big(1 - H(\phi_1)\big)\right], \end{aligned} \tag{25}$$

with the energy densities e_i defined in (18).

4.3 Redistancing

During their evolution according to equations (17) or (25), 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. In numerical implementations, we found that a very steep slope of the level set functions eventually inhibits the flexibility of the boundary to displace.

Many people have advocated the use of a redistancing procedure to constrain the slope of ϕ to $|\nabla \phi| = 1$, c.f. [10]. In order to reproject the evolving level set function to the space of distance functions, we intermittently iterate several steps of the redistancing equation [27]:

$$\frac{\partial \phi}{\partial t} = \text{sign}(\hat{\phi})\ (1 - |\nabla \phi|)\,, \tag{26}$$

where $\hat{\phi}$ denotes the level set function before redistancing. Although this regularization is optional in the proposed level set model — see also [5] — it improves the convergence of the boundary evolution. Since the data term given by the image motion information dominates the evolution of the boundary, we found this simple redistancing process to be sufficiently accurate for our application. Therefore we did not revert to more elaborate iterative redistancing schemes such as the one presented in [26].

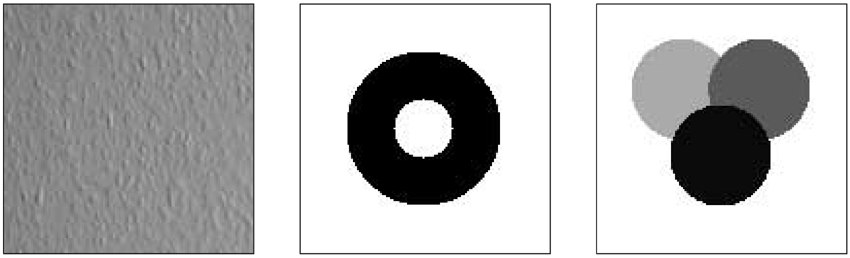


Fig. 1. Data for ground truth experiments. Left: Specific image regions of the wallpaper shot are artificially translated to generate input data. **Middle and Right:** Chosen image regions indicated by different gray values.

5 Ground Truth Experiments

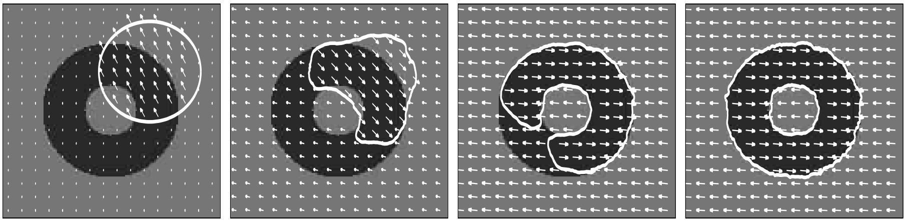
In order to verify the precision of our segmentation approach, we performed a number of ground truth experiments in the following way. We took a snapshot of homogeneously structured wallpaper, which is shown in Figure 1, left side. Then we artificially translated certain image regions according to a particular motion. The respective image regions are highlighted in various shades of gray in Figure 1, middle and right side.

We determined the spatio-temporal image gradient from two consecutive images and specified a particular initialization of the boundary. Then we minimized the functional (21) by alternating the three fractional steps of:

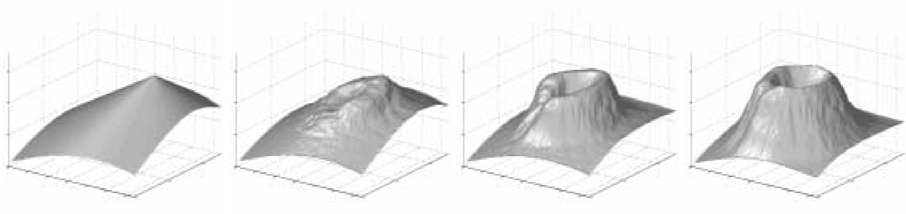
- iterating the gradient descent (17) or (25) for the level set functions,
- iterating the redistancing procedure (26) for the level set functions,
- and updating the motion vectors for all phases by solving the corresponding eigenvalue problem (23).

For all experiments, we show the evolving motion boundaries and corresponding motion estimates superimposed onto the ground truth region information. Yet, it should be noted that in these experiments the objects cannot be distinguished from the background based on their appearance, as they correspond to homogeneously textured parts of the wallpaper. Therefore, all results are obtained exclusively on the basis of the motion information. For the purpose of illustration, we also show in the first example the evolution of the level set function, the zero level set of which represents the motion boundary.

The three ground truth experiments are chosen so as to highlight different properties of the proposed approach: The first example shows a result obtained with the two-phase model, in which a multiply connected moving object is segmented on a differently moving background. The second and third experiment show an application of the four-phase model in which three differently moving regions are segmented, once on a static and once on a moving background.



Evolution of boundary and motion field superimposed on true region.



Corresponding evolution of the embedding level set function.

Fig. 2. Segmenting a multiply connected moving object. The two input images show the wallpaper of Figure 1, left side, with a ring-shaped region translating to the right and the remaining region translating to the left. Both the motion estimates and the evolution of the boundary between the two phases are obtained by minimizing the two-phase motion functional (14) simultaneously with respect to both the level set function ϕ and the motion vectors v_1 and v_2. Thus, the minimization of the *single* energy functional generates both the precise object location and the motion information for object and background. Due to the level set representation, the evolving motion discontinuity set can change topology. Note also that in the input data the region of interest is not perceivable based on its appearance.

5.1 Segmenting Multiply Connected Moving Objects

In this experiment, we demonstrate the capacity of the proposed approach to segment multiply connected moving objects. The two input images consisted of the wallpaper shown in Figure 1, left side, and the same image with the ring-shaped area indicated in Figure 1, center, translated to the right, and the remaining image area translated to the left.

Figure 2, bottom row, shows four steps in the evolution of the level set function ϕ, generated by minimizing the two-phase model (14). The top row shows the evolution of the corresponding motion boundary, given by the zero level set of ϕ, and the estimated motion field, superimposed on the ground truth information about the ring region. The results show that one can obtain both a precise information about the location of a multiply connected object and about the motion of object and background by minimizing a *single* energy functional.

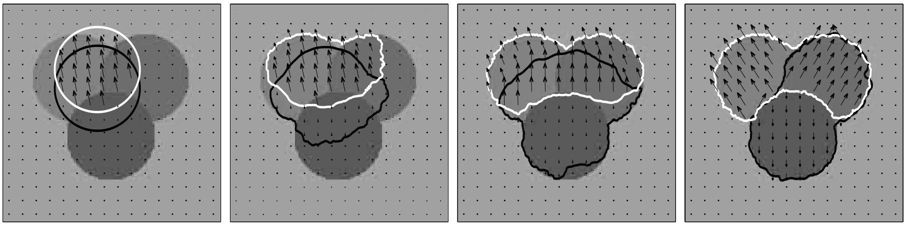


Fig. 3. Segmenting multiple moving regions. The two input images show the wallpaper of Figure 1, left side, with three circular regions moving away from the center. The magnitude of the velocity of the upper two regions is 1.4 times larger than that of the bottom region. Superimposed on the true region 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 four-phase model (22) 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.

5.2 Segmenting Several Differently Moving Regions

In this experiment, we demonstrate an application of the four-phase model (22) 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 (22). 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:

$$R_1 = \{x \in \Omega \mid \phi_1 \geq 0, \phi_2 \geq 0\}, \quad R_2 = \{x \in \Omega \mid \phi_1 \geq 0, \phi_2 < 0\},$$

$$R_3 = \{x \in \Omega \mid \phi_1 < 0, \phi_2 \geq 0\}, \quad R_4 = \{x \in \Omega \mid \phi_1 < 0, \phi_2 < 0\}.$$

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. Again, we stress 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*.

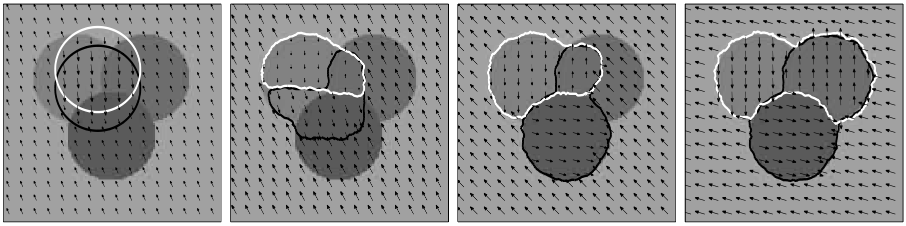


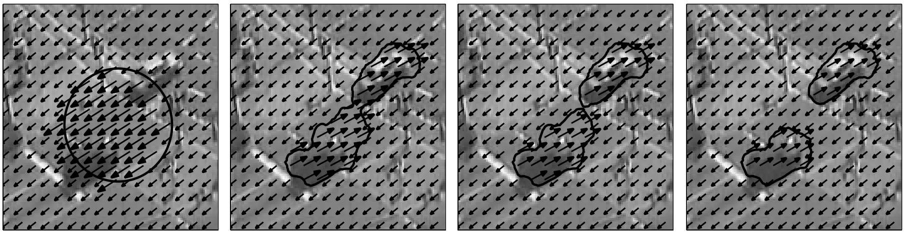
Fig. 4. Segmenting multiple moving regions on moving background. The two input images show the wallpaper of Figure 1, left side, with the three circular regions and the background moving in different directions. The true motion directions for all regions are down (top left), up (top right), right (bottom) and left (background). Superimposed on the true region information are the evolving zero level sets of ϕ_1 (black contour) and ϕ_2 (white contour), defining four different phases. The simultaneously evolving piecewise constant motion field is represented by the black arrows. Both the phase boundaries and the piecewise constant motion field are obtained by minimizing the four-phase model (22). Upon convergence, the two boundaries clearly separate the four motion phases corresponding to the three regions and the background.

5.3 Multiple Moving Objects and Moving Background

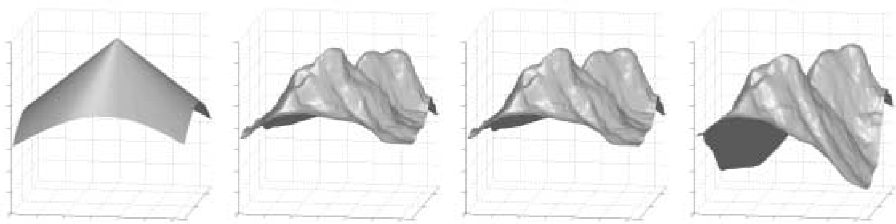
In the previous example, the three regions were moving, while the background was static. In many real-world applications of motion estimation and motion segmentation, the background may also undergo a certain motion — for example in a motion sequence filmed by a moving camera. This problem has been addressed by a number of researchers. In particular, it has been proposed to estimate the *dominant* motion in a robust estimator framework (cf. [20,2]). Although this may permit to compensate for the background motion, it strongly relies on the assumption that the background forms the dominant part of the image plane[2].

Our approach does not rely on any assumptions about the relative size of the different moving regions. Figure 4 shows the segmentation of a sequence containing three moving regions on a moving background, obtained by minimizing the four-phase model (22). During energy minimization, both the motion estimates and the motion boundary are progressively improved. The boundaries of the four motion phases converge over a fairly large distance, yet the region boundaries are precisely reconstructed in the final segmentation. As in the previous example, the zero level sets of ϕ_1 and ϕ_2 define a segmentation which is essentially identical with the ground truth. The directions of the motion estimated for the lower region and the background deviate slightly from the ground truth. It is unclear where this small discrepancy stems from.

[2] In [2], for example, it is stated that the robust estimation of the background motion works well on an artificial sequence (involving translatory motion only) if the background motion takes up at least 60% of the image plane.



Evolution of boundary and motion field superimposed on the first frame.



Corresponding evolution of embedding level set function.

Fig. 5. Segmenting moving cars captured by a moving camera.
The image sequence shows two cars moving to the top right, and the background moving to the bottom left. Due to the level set representation, the topology of the motion boundary is not constrained, such that splitting and merging is possible. By minimizing the two-phase model (14), one obtains a fairly accurate segmentation of the two cars and an estimate of the motion of cars and background.

6 Real-World Application: Segmentation of Moving Cars

As a final demonstration of the proposed level set based motion segmentation approach, we now present an application to 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[3]. 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 shifting one of the two frames, thereby simulating the case of a moving camera.

The images in figure 5, top row, show the contour evolution with the corresponding motion estimates superimposed on one of the two frames. These were generated by minimizing the two-phase model (14). The bottom row shows the evolution of the underlying level set function. Due to the level set representation of the boundary, the zero level set can undergo topological changes such as the splitting and merging from the third to the fourth frame.

[3] KOGS/IAKS, Univ. of Karlsruhe, `http://i21www.ira.uka.de/image_sequences/`

7 Conclusion

We approached the problems of motion estimation and motion segmentation from the viewpoint of Bayesian inference. Based on a few very basic assumptions, we rigorously derived a variational framework for segmenting the image plane into a set of regions of homogeneous motion. Our model is based on a conditional probability for the spatio-temporal image gradient, given a particular velocity vector, and on a prior on the estimated piecewise constant motion field favoring motion boundaries of minimal length.

The proposed functional depends on velocity vectors for a set of disjoint regions and the boundary separating them. It can be considered an extension of the Mumford-Shah functional from the case of gray value segmentation to the case of motion segmentation. The only free parameter is given by a fundamental scale parameter intrinsic to all segmentation approaches.

We proposed an implementation of the motion segmentation functional in a multiphase level set framework. The resulting model has the following favorable 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 into a set of disjoint regions of homogeneous velocity.
- The *implicit* representation of the motion discontinuity set does not depend on a particular choice of parameterization. Moreover, it allows for topological changes of the boundary such as splitting or merging.
- The *multiphase* level set formulation permits a segmentation of the image plane into several (possibly multiply-connected) motion phases.
- Minimizing the proposed functional is straight-forward. It results in an eigenvalue problem for the motion vectors, and a gradient descent evolution for the level set functions embedding the motion boundary.
- Due to the region-based homogeneity criterion rather than an edge-based formulation, the functional is robust to noise and the motion boundaries tend to converge over fairly large spatial distances.
- The segmentation and motion estimates are obtained on the basis of the spatio-temporal image gradient calculated from only two consecutive frames of an image sequence. Therefore the presented approach is in principle amenable to real-time implementations and tracking.

We demonstrated these properties by a number of experimental results which were obtained both on simulated ground truth data and on real-world sequence data. Present work focuses on extending the proposed approach to the simultaneous segmentation of multiple frames in a sequence and to estimating piecewise parametric motion fields (cf. [7]).

Acknowledgements

The author would like to thank P. Favaro, S. Soatto, A. Yuille and S.-C. Zhu for fruitful discussions. This research was supported by ONR N00014-02-1-0720 and AFOSR F49620-03-1-0095.

References

1. S. Ayer and H.S. Sawhney. 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.*, pages 777–784, Boston, USA, 1995.
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. V. Caselles and B. Coll. Snakes in movement. *SIAM J. Numer. Anal.*, 33:2445–2456, 1996.
4. V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. IEEE Internat. Conf. on Comp. Vis.*, pages 694–699, Boston, USA, 1995.
5. T. Chan and L. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.
6. D. Cremers. A variational framework for image segmentation combining motion estimation and shape regularization. In C. Dyer and P. Perona, editors, *IEEE Int. Conf. on Comp. Vis. and Patt. Recog.*, Madison, Wisconsin, June 2003. To appear.
7. 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.
8. D. Cremers and C. Schnörr. Statistical shape knowledge in variational motion segmentation. *Image and Vision Computing*, 21(1):77–86, 2003.
9. G. Farnebäck. Very high accuracy velocity estimation using orientation tensors, parametric motion, and segmentation of the motion field. In *Proc. 8th ICCV*, volume 1, pages 171–177, 2001.
10. J. Gomes and O. D. Faugeras. Level sets and distance functions. In D. Vernon, editor, *Proc. of the Europ. Conf. on Comp. Vis.*, volume 1842 of *LNCS*, pages 588–602, Dublin, Ireland, 2000. Springer.
11. B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artif. Intell.*, 17:185–203, 1981.
12. 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.
13. S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi. Gradient flows and geometric active contour models. In *Proc. IEEE Internat. Conf. on Comp. Vis.*, pages 810–815, Boston, USA, 1995.
14. P. Kornprobst, R. Deriche, and G. Aubert. Image sequence analysis via partial differential equations. *J. Math. Im. Vis.*, 11(1):5–26, 1999.
15. B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc.7th International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, 1981.
16. A. Mansouri, B. Sirivong, and J. Konrad. Multiple motion segmentation with level set. In *Proc. SPIE Conf. on Image and Video Communications and Processing*, pages 584–595, Santa Fe, 2000.

17. 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.
18. J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation.* Birkhäuser, Boston, 1995.
19. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
20. J.-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *J. of Visual Commun. and Image Repr.*, 6(4):348–365, 1995.
21. J.-M. Odobez and P. Bouthemy. Direct incremental model-based image motion segmentation for video analysis. *Signal Proc.*, 66:143–155, 1998.
22. 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.
23. N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 22(3):266–280, 2000.
24. C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia. A level set model for image classification. *Int. J. of Comp. Vis.*, 40(3):187–197, 2000.
25. C. Schnörr. Computation of discontinuous optical flow by domain decomposition and shape optimization. *Int. J. of Comp. Vis.*, 8(2):153–165, 1992.
26. M. Sussman and E. Fatemi. An efficient, interface-preserving level set redistancing algorithm and its application to interfacial incompressible fluid flow. *SIAM J. Sci. Comput.*, 20(4):1165–1191, 1999.
27. 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.
28. 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.
29. 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.
30. H.-K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *J. of Comp. Phys.*, 127:179–195, 1996.