

A Variational Framework for Image Segmentation Combining Motion Estimation and Shape Regularization

Daniel Cremers

Department of Computer Science
University of California, Los Angeles – CA 90095
cremers@ucla.edu

Abstract

Based on a geometric interpretation of the optic flow constraint equation, we propose a conditional probability on the spatio-temporal image gradient. We consistently derive a variational approach for the segmentation of the image domain into regions of homogeneous motion.

The proposed energy functional extends the Mumford-Shah functional from gray value segmentation to motion segmentation. It depends on the spatio-temporal image gradient calculated from only two consecutive images of an image sequence. Moreover, it depends on motion vectors for a set of regions and a boundary separating these regions. In contrast to most alternative approaches, the problems of motion estimation and motion segmentation are jointly solved by minimizing a single functional.

Numerical evaluation with both explicit and implicit (level set based) representations of the boundary shows the strengths and limitations of our approach.

1. Introduction and Related Work

The estimation of motion from image sequences has a long tradition in computer vision. In recent years, many approaches have been proposed to segment the image plane on the basis of this motion information. The fields of image sequence analysis and video compression provide numerous applications. In some approaches, motion discontinuities are modeled implicitly [1, 10, 9, 15]. Other approaches treat the problems of motion estimation in disjoint sets and optimization of the motion boundaries separately [14, 2, 12, 13, 7].

In [5], 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 an important drawback: Satisfactory results were only obtained upon applying two posterior normalizations to the terms driving the evolution of the motion boundary.

In the present paper, we derive a novel variational formulation for segmenting the image plane into regions of homogeneous motion. It is based on a simple probabilistic model for the spatio-temporal image gradient determined from two consecutive images of a sequence. We show that local minimization of an appropriate energy functional leads to an eigenvalue problem for the motion parameters and to a gradient descent evolution for the motion boundaries. In contrast to our previous approach, all normalizations comprised in the evolution equation are derived in a consistent manner by minimizing the proposed functional.

We present numerical results for two implementations of the functional: one with an explicit spline based representation of the contour, and one with an implicit level set based representation. In particular, these results cover the cases of a moving object on a moving background and of multiple moving regions.

2. From the Optic Flow Constraint . . .

Let $f : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be a gray value image sequence. We assume the intensity of a moving point to be constant throughout time. This induces the optic flow constraint equation:

$$\frac{d}{dt}f(x, t) = \frac{\partial f}{\partial x}u + \frac{\partial f}{\partial y}w + \frac{\partial f}{\partial t} = 0, \quad (1)$$

where $(u, w)^t$ is the local velocity vector. Geometrically, this constraint states that the spatio-temporal image gradient

$$\nabla_3 f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial t} \right)^t, \quad (2)$$

must either vanish or be orthogonal to the homogeneous velocity vector $v = (u, w, 1)^t$:

$$\nabla_3 f^t v = 0. \quad (3)$$

This constraint has been employed in many motion estimation approaches. Commonly — for example in the seminal work of Horn and Schunck [8] and many subsequent works

— the square of this constraint is used as a fidelity term. In contrast, we suggest to use the angle α between the two vectors as a measure of the orthogonality.

To this end, let x be a point with spatio-temporal derivative $\nabla_3 f$, and let R be a region of the image with homogeneous velocity v . Then we model the probability that the point x is part of the region R by:

$$P(\nabla_3 f | v) \propto e^{-\cos^2(\alpha)} = \exp\left(-\frac{(v^t \nabla_3 f)^2}{|v|^2 |\nabla_3 f|^2}\right). \quad (4)$$

This probability has the following favorable properties:

- It is maximal if the vectors v and $\nabla_3 f$ are orthogonal.
- It is minimal if the two vectors are parallel.
- It only depends on the *angle* between the spatio-temporal image gradient and the homogeneous velocity vector, and *not on the magnitude* of these vectors.

3. . . . to Motion Segmentation

Based on the probability measure in equation (4), we will now define a variational approach for motion segmentation. We suggest to segment the image plane Ω into a set of pairwise disjoint regions R_i of homogeneous velocity v_i by minimizing the functional

$$E(\{v_i\}, C) = \sum_i \int_{R_i} \left(-\log P(\nabla_3 f | v_i)\right) dx + \nu \mathcal{L}(C) \quad (5)$$

simultaneously with respect to the motion vectors v_i of each region R_i and with respect to the motion boundary C separating these regions. The term $\mathcal{L}(C)$ denotes the length of this boundary.

Inserting the model (4), we get (up to a constant):

$$E(\{v_i\}, C) = \sum_i \int_{R_i} \frac{|v_i^t \nabla_3 f|^2}{|v_i|^2 |\nabla_3 f|^2} dx + \nu \mathcal{L}(C). \quad (6)$$

This functional (6) has the simple form

$$E(\{v_i\}, C) = \sum_i \frac{v_i^t M_i v_i}{|v_i|^2} + \nu \mathcal{L}(C), \quad (7)$$

where

$$M_i = \int_{R_i} \frac{\nabla_3 f \nabla_3 f^t}{|\nabla_3 f|^2} dx \quad (8)$$

In numerical implementations, we replace the term $|\nabla_3 f|^2$ in the denominator by $|\nabla_3 f|^2 + \epsilon^2$ with a small constant ϵ . This guarantees that the matrix M_i is always well defined. Moreover, points with very small spatio-temporal gradient (in the order of ϵ), i. e. weak motion information, will contribute less to the segmentation process.

The functional (6) can be considered a generalization of the Mumford–Shah functional [11, 16] from gray value segmentation to motion segmentation.

4. Energy Minimization

Given two consecutive images f_1 and f_2 from an image sequence, we approximate the spatio-temporal gradient by:

$$\nabla_3 f \approx \begin{pmatrix} \nabla \frac{f_1 + f_2}{2} \\ f_2 - f_1 \end{pmatrix}. \quad (9)$$

From this, we determine the matrix

$$M = \frac{\nabla_3 f \nabla_3 f^t}{|\nabla_3 f|^2} \quad (10)$$

for each point in the image plane.

Given an initial contour C , we minimize the energy (6) by alternating the two fractional steps of updating the motion parameters v_i for the regions R_i , and of iterating a gradient descent evolution for the boundary C separating these regions. This will be detailed in the following.

4.1. An Eigenvalue Problem

For a fixed boundary C , minimization of the functional (6) results in the eigenvalue problem

$$v_i = \arg \min_v \frac{v^t M_i v}{v^t v}. \quad (11)$$

The homogeneous velocity vector v_i for each region R_i is therefore given by the eigenvector corresponding to the smallest eigenvalue of the 3×3 -matrix M_i defined in (8). It is normalized, such that the third component is 1.

4.2. Motion Competition

For fixed velocity vectors v_i , a gradient descent on the energy (6) for the boundary C results in the evolution equation:

$$\frac{dC}{dt} = -\frac{dE}{dC} = (e_j - e_k) \cdot n - \nu \frac{d\mathcal{L}(C)}{dC}, \quad (12)$$

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

$$e_i = \frac{v_i^t M_i v_i}{v_i^t v_i} \quad (13)$$

is an energy density.

The two terms in the contour evolution (12) have the following intuitive interpretation:

- The first term is proportional to the difference of the energy densities e_i in the regions neighboring the boundary. The neighboring regions compete for the boundary in terms of their motion energy, 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. A Spline Based Implementation

In this section, we propose an implementation of the contour evolution (12) with an explicit closed spline curve:

$$C : [0, 1] \times \mathbb{R}^+ \rightarrow \Omega, \quad C(s, t) = \sum_{n=1}^N p_n(t) B_n(s), \quad (14)$$

with quadratic periodic B-spline basis functions B_n and control points $p_n = (x_n, y_n)^t$.

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 behaviour, we use the length measure $\mathcal{L}(C) = \int_0^1 \left(\frac{\partial C}{\partial s} \right)^2 ds$. As discussed in [6], minimizing this constraint enforces an equidistant spacing of control points which strongly improves the numerical stability. The contour evolution then reads:

$$\frac{\partial C}{\partial t} = (e_j - e_k) \cdot n - \nu \frac{\partial^2 C}{\partial s^2}. \quad (15)$$

By inserting the spline definition (14), this equation is easily converted to an evolution for the control points $p_n(t)$.

In practice, we iterate this gradient descent for the control points $p_n(t)$, and update in alternation the motion energy densities e_i according to (13) and (11).

6. A Level Set Implementation

The explicit contour implementation presented in the previous section is quite fast, because the contour evolution only involves the update of a small number of control point coordinates. Yet, explicit contour representations have certain disadvantages. Firstly, one needs to take care of a regriding 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.

We therefore propose an implicit level set based implementation of the functional (6). It is based on the analogous gray value approach proposed in [3]. The idea of all level set contour descriptions is to define the contour C as the zero level set of a function $\phi : \Omega \rightarrow \mathbb{R}$:

$$C = \{x \in \Omega \mid \phi(x) = 0\}. \quad (16)$$

Using the Heaviside step function

$$H(\phi) = \begin{cases} 1 & \text{if } \phi \geq 0 \\ 0 & \text{if } \phi < 0 \end{cases}, \quad (17)$$

we can embed the motion energy (6) by the two-phase functional:

$$\begin{aligned} E(v_1, v_2, \phi) &= \int_{\Omega} \frac{|v_1^t \nabla_3 f|^2}{|v_1|^2 |\nabla_3 f|^2} H(\phi) dx \\ &+ \int_{\Omega} \frac{|v_2^t \nabla_3 f|^2}{|v_2|^2 |\nabla_3 f|^2} (1 - H(\phi)) dx + \nu \int_{\Omega} |\nabla H(\phi)| dx. \end{aligned} \quad (18)$$

The contour evolution (12) then corresponds to the gradient descent evolution on the embedding function ϕ :

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[\nu \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + e_2 - e_1 \right], \quad (19)$$

with e_i as defined in (13). In numerical implementations, we use for the delta function $\delta(\phi) = \frac{d}{d\phi} H(\phi)$ a smoothed approximation of finite width τ , as suggested in [3]. Depending on the size of τ , this permits to detect interior contours, as shown in the results of Figure 5.

Minimization with respect to the motion parameters v_1 and v_2 will again lead to an eigenvalue problem of the form (11) for the 3×3 -matrices

$$M_1 = \int_{\Omega} \frac{\nabla_3 f \nabla_3 f^t}{|\nabla_3 f|^2} H(\phi) dx, \quad M_2 = \int_{\Omega} \frac{\nabla_3 f \nabla_3 f^t}{|\nabla_3 f|^2} \bar{H}(\phi) dx,$$

where $\bar{H} = 1 - H$. As in the explicit scheme, we minimize the functional (18) by alternating the update of the motion vectors v_i with the iteration of the contour evolution (19).

7. Numerical Results

7.1. Segmenting Multiple Motion

To evaluate the explicit scheme introduced in Section 5, we used the Avengers sequence.¹ A moving car is captured by a moving camera.

Figure 1 shows segmentation results obtained on frames 18 through 34. We fixed an initial contour (left), determined the spatio-temporal derivative for the given pair of consecutive frames and iterated the minimization of energy (6), alternating the motion estimation (11) and the contour evolution (15). 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 image in Figure 1.

Minimizing energy (6) 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 Figure 3. Both

¹We thank P. Bouthemy and his group for providing us with the image data from the Avengers sequence.



Figure 1: Segmentation of multiple motion for the frames 18–34 from the Avengers sequence: Contour evolution for the functional (6) with an explicit contour initialized as shown in the top left image. The first two images show the evolution of the contour for the first pair of frames, the following images show the segmentation 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 and only slowly degrades once the car starts moving perpendicular to the viewing plane (right).



Figure 2: Motion segmentation with an explicit contour for the frames 35–45 from the Avengers sequence. The contour is initialized as shown on the left, then the minimization of (6) is iterated a fixed number of steps on each pair of consecutive frames (the first two images showing frame 35). The comoving shadow is initially associated with the car, but attributed to the background later on. Indeed, due to its semi-transparency, it is unclear whether the shadow is part of the car or not. There is no hypothesis of motion continuity, therefore our approach can also be used for estimating temporally discontinuous motion and for tracking.

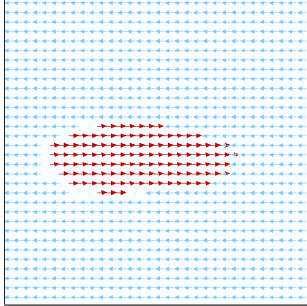


Figure 3: Motion estimate generated by minimizing energy (6) for the first two frames from Figure 1. Both car and background move at different velocities – cf. Fig. 1, 2nd image.

the car and the background are moving, with velocities of different direction and magnitude.

Figure 2 shows similar results for the frames 35–45 of the Avengers sequence. The proposed method always uses only two consecutive frames. Although using more than two frames has been shown to stabilize the problem of motion estimation, we believe that using only two frames has several advantages, in particular:

- No hypothesis is made on temporal continuity of the motion. Therefore temporally discontinuous motion can be estimated and segmented as well.

- Only two consecutive frames are used, motion estimation reduces to a simple eigenvalue problem and the contour evolution to an update of a few control points. Therefore the proposed method is amenable to real-time implementations for online tracking.

7.2. Segmenting Multiply Connected Regions

To evaluate the implicit scheme introduced in Section 6, we used two consecutive frames from a sequence showing a moving object which is not simply connected: A roll of scotch tape is moving on a newspaper.

Figure 4 shows the initial contour and the contour evolution obtained by minimizing energy (18) which amounts to alternating the gradient descent (19) and the motion parameter update (11). The images in the top row show one of the two consecutive frames with the evolving contour and the estimated motion superimposed.

The figures in the bottom row show the corresponding evolution of the embedding surface ϕ , underlying the contour evolution. It explains the change of contour topology from the fourth to the fifth image.

Figure 5 shows the same segmentation process for a different initialization. These images demonstrate that the contour converges over fairly large distances. Moreover, our numerical scheme is capable of detecting interior motion boundaries.

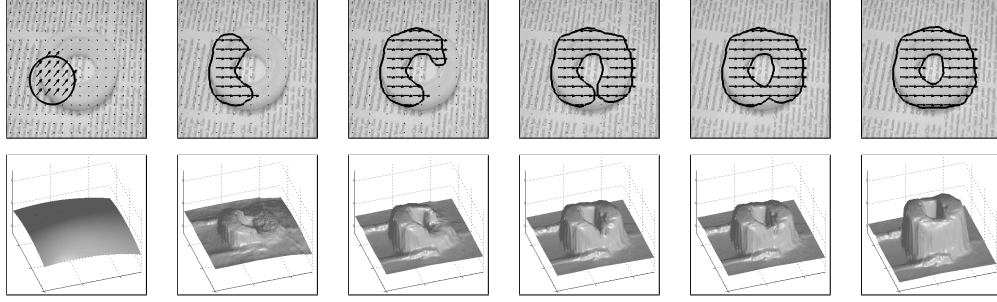


Figure 4: Level set motion segmentation for the energy (18). **Top row:** One of the two input images (showing a scotch tape moving on a newspaper) with the evolving contour and the estimated motion superimposed. Note that the object of interest is hardly distinguishable from the background on the basis of its intensity. Yet, the minimization of the energy (18) generates both a segmentation of the image plane and an estimate of the motion in each region. **Bottom row:** Corresponding evolution of the embedding surface ϕ . The evolving surface induces a change of the contour topology from the fourth to the fifth image. Moreover, the embedding surface is less negative in regions of weak gray value structure because these are less easily ascribed to one or the other motion hypothesis.

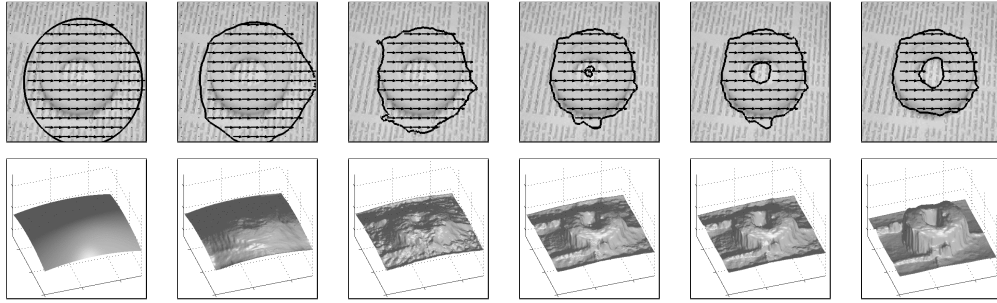


Figure 5: Detecting interior motion boundaries. **Top row:** Evolving motion boundary and estimated motion superimposed on the first of the two input images. **Bottom row:** Corresponding evolution of the embedding level set function ϕ . The transition from the third to the fourth image illustrates the process of detecting interior motion boundaries.

7.3. Segmenting Several Moving Objects

The following example presents an application of the level set framework (18) in a real-world traffic scene showing several moving objects with a differently moving background. We used two consecutive images from a sequence recorded by D. Koller and H. Nagel (KOGS/IAKS, University of Karlsruhe).² The sequence shows several cars moving in the same direction, filmed by a static camera. In order to increase the complexity of the scene, we artificially induced a background motion by shifting one of the two frames, thereby simulating a moving camera.

The images in figure 6, top row, show the contour evolution with the corresponding motion estimates superimposed on one of the two frames. The bottom row shows the evolution of the underlying level set function. Due to the level set representation, the boundary can undergo topological changes such as the split from the third to the fourth frame. Therefore this framework permits to segment multiple moving objects against a differently moving background.

²http://i21www.ira.uka.de/image_sequences/

8. Summary and Conclusions

We presented a probabilistic approach to the problem of segmenting images on the basis of motion information. Starting from a geometric interpretation of the well-known optic flow constraint, we proposed a conditional probability on the spatio-temporal image gradient at a point x , given the velocity v . From this probability model we derived a novel variational framework for segmenting the image plane into regions of homogeneous motion.

We showed that minimizing the proposed energy leads to an eigenvalue problem for the motion parameters and to an evolution for the separating motion boundary. We demonstrated the generality of our approach by detailing two implementations of this functional — one with an explicit spline representation of the motion boundary, and one with an implicit level set based representation.

Numerical results on real-world image sequences demonstrate the capacity of our approach to segment multiply moving regions (moving cars captured by a moving camera), and to segment multiply connected moving

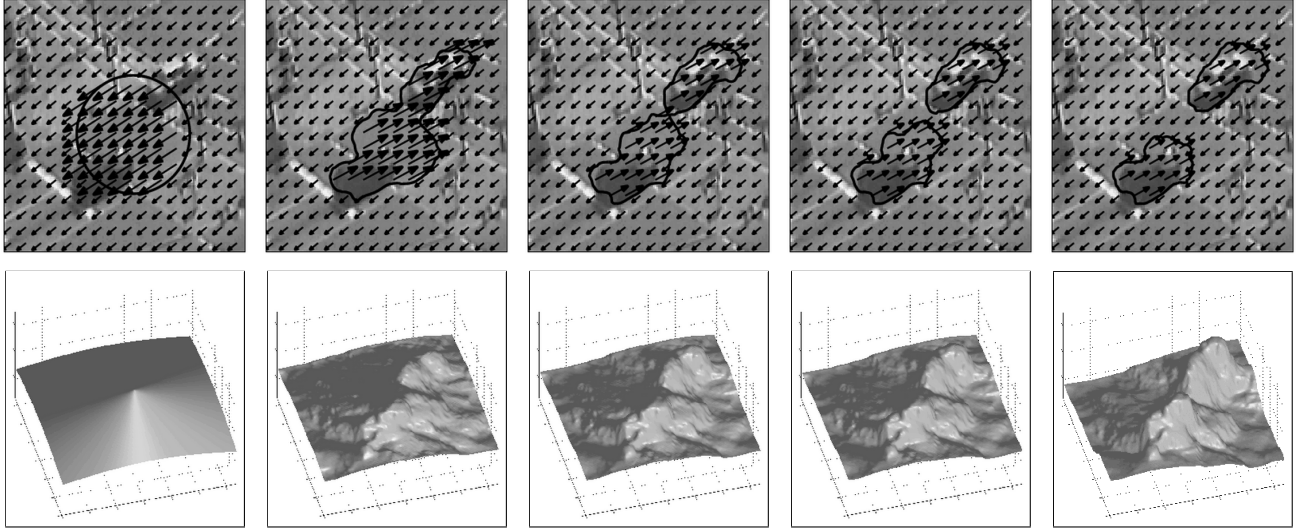


Figure 6: Segmenting multiple moving objects on a moving background. **Top row:** Evolving motion boundary and estimated motion field superimposed on one of the two input images. **Bottom row:** Corresponding evolution of the embedding level set function ϕ . The motion segmentation functional permits to segment differently moving regions. Due to the level set implementation, the contour topology is not fixed such that multiple regions can be segmented on the basis of their motion. Note that both the location of the motion boundary and the motion estimates for cars and background are gradually improved during minimization of the proposed energy. Minor discrepancies between the final segmentation and the car boundaries are probably due to the fact that the gray value of the street is not sufficiently structured to permit a reliable motion estimation.

regions. We showed that the method is robust to non-translatory motion.

We illustrated the implicit scheme by showing how the embedding surface generates topological changes of the motion boundary. In particular, we demonstrated that our method is capable of detecting interior motion boundaries.

Present work focuses on generalizations of the proposed approach to more than two motion phases [4] and to the simultaneous segmentation of multiple frames in a sequence.

Acknowledgments

The author thanks C. Schnörr, J. Weickert, S. Weber, S. Soatto, P. Favaro, A. Yuille and S.-C. Zhu for stimulating discussions. This research was supported by ONR N00014-02-1-0720 and AFOSR F49620-03-1-0095.

References

- [1] 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.
- [2] V. Caselles and B. Coll. Snakes in movement. *SIAM J. Numer. Anal.*, 33:2445–2456, 1996.
- [3] T. Chan and L. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.
- [4] D. Cremers. A multiphase level set framework for variational motion segmentation. In L. Griffith, editor, *Int. Conf. on Scale Space Theories in Computer Vision*, Isle of Skye, 2003. Springer. To appear.
- [5] D. Cremers and C. Schnörr. Statistical shape knowledge in variational motion segmentation. *Im. and Vis. Comp.*, 21(1):77–86, 2003.
- [6] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion Snakes: Introducing statistical shape knowledge into the Mumford–Shah functional. *Int. J. of Comp. Vis.*, 50(3):295–313, 2002.
- [7] G. Farneback. 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.
- [8] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artif. Intell.*, 17:185–203, 1981.
- [9] P. Kornprobst, R. Deriche, and G. Aubert. Image sequence analysis via partial differential equations. *J. Math. Im. Vis.*, 11(1):5–26, 1999.
- [10] 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.
- [11] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
- [12] J.-M. Odobez and P. Bouthemy. Direct incremental model-based image motion segmentation for video analysis. *Signal Proc.*, 66:143–155, 1998.
- [13] 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.
- [14] C. Schnörr. Computation of discontinuous optical flow by domain decomposition and shape optimization. *Int. J. of Comp. Vis.*, 8(2):153–165, 1992.
- [15] 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.
- [16] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(9):884–900, 1996.