Variational Space-Time Motion Segmentation

Daniel Cremers and Stefano Soatto

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

Abstract

We propose a variational method for segmenting image sequences into spatio-temporal domains of homogeneous motion. To this end, we formulate the problem of motion estimation in the framework of Bayesian inference, using a prior which favors domain boundaries of minimal surface area. We derive a cost functional which depends on a surface in space-time separating a set of motion regions, as well as a set of vectors modeling the motion in each region.

We propose a multiphase level set formulation of this functional, in which the surface and the motion regions are represented implicitly by a vector-valued level set function. Joint minimization of the proposed functional results in an eigenvalue problem for the motion model of each region and in a gradient descent evolution for the separating interface.

Numerical results on real-world sequences demonstrate that minimization of a single cost functional generates a segmentation of space-time into multiple motion regions.

1. Introduction and Related Work

Segmenting images into semantically significant components has been a focus of computer vision research in the last decades. An important requirement of a segmentation approach is that it needs to deal with noise present in most real-world image data. However, a denoising process should conserve the boundaries of objects of interest and therefore implicitly assumes the segmentation to be known. Mumford and Shah [12] suggested that the problems of denoising and segmentation are closely interlaced and should therefore be solved simultaneously. In their approach, the denoising amounts to approximating the intensities in the segmented regions by a piecewise smooth or piecewise constant function. By minimizing a single functional, one simultaneously generates a segmentation of the image and an approximation of the gray values in each region. This idea was extended to color and texture images in [23].

In the present paper, we demonstrate that a similar reasoning can be applied to the case of segmenting moving objects in video sequences.

Many researchers have investigated the problem of motion estimation. Two seminal methods were proposed by Horn and Schunck [7] and by Lucas and Kanade [9]. These methods are based on a least-squares criterion for the optic flow constraint and global or local smoothness assumptions on the estimated flow field. In general, flow fields are not smooth. The boundaries of moving objects correspond to discontinuities in the motion field. Such motion discontinuities have been modeled implicitly by non-quadratic robust estimators [1, 10, 8, 21]. Alternatively, it was proposed to separately estimate the motion in disjoint sets and optimize the motion boundaries [14, 19, 22, 2, 13, 5].

In [4], we presented a variational approach called *motion competition* which jointly solves the problems of motion estimation and segmentation for two consecutive frames from a sequence in a similar way as the Mumford-Shah approach does for denoising and segmenting gray value images. Coupling the two problems of segmentation and approximation is even more important in the case of motion segmentation: Beyond being corrupted by noise, local motion information is not available at all in directions of constant intensity, a limitation which is referred to as the *aperture problem*.

In this work, we generalize this approach to segmenting the motion in an entire sequence. Rather than segmenting the image plane into a set of disjoint regions, we propose to segment the spatio-temporal volume into a set of disjoint phases of homogeneous motion. Compared to the iteration of the two-frame model, this introduces an additional temporal regularity of the estimated motion boundaries.

In Sections 2-4, we formulate the problem of image sequence segmentation as one of Bayesian inference. We derive a cost functional which depends on parameter vectors modeling the motion in a set of disjoint regions and on a surface separating these regions in the spatio-temporal domain. In Sections 5 and 6, we propose a multiphase level set formulation [3, 18] of this functional. The implicit representation of the separating interface by a vector-valued level set function permits to segment several (possibly multiply connected) motion phases. In Section 7, we present numerical results which highlight different properties of our approach. In Section 8, we discuss limitations and possible extensions.

2. Bayesian Formulation

Let $\Omega \subset \mathbb{R}^2$ denote the image plane and let

$$f : \Omega \times [0,T] \longrightarrow \mathbb{R}^+ \qquad (1)$$

be a gray value image sequence, which is assumed to be differentiable. We denote the spatio-temporal image gradient by

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

Let

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

be the velocity vector in homogeneous coordinates defined on the domain $D = \Omega \times [0,T]$. Given the spatio-temporal image gradient at each point and time, the Bayesian estimate for the motion in the sequence is given by the velocity field v which maximizes the posterior probability

$$\mathcal{P}(v \,|\, \nabla_3 f) = \frac{\mathcal{P}(\nabla_3 f \,|\, v) \ \mathcal{P}(v)}{\mathcal{P}(\nabla_3 f)}. \qquad (4)$$

Assuming the intensity of a moving point to be constant as it moves, we obtain the optic flow constraint equation:

$$\frac{d}{dt} f(x,t) = \frac{\partial f}{\partial x} u + \frac{\partial f}{\partial y} w + \frac{\partial f}{\partial t} = \nabla_3 f^t v = 0, \qquad (5)$$

Geometrically, this constraint states that the spatio-temporal image gradient $\nabla_3 f$ must either vanish or be orthogonal to the homogeneous velocity vector $v = (u,\, w,\, 1)^t$. This constraint has been employed in many motion estimation approaches. Commonly – for example in the seminal work of Horn and Schunck [7] and many subsequent works[1] – 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 $(x,t) \in D$ be a point with spatio-temporal derivative $\nabla_3 f$, and let $D_i \subset D$ be a domain of the image sequence with homogeneous velocity v_i. We model the probability that (x,t) is part of the domain D_i by:

$$P(\nabla_3 f \,|\, v_i) \ \propto\ e^{-\cos^2(\alpha)} = \exp\left(-\frac{(v_i^t \nabla_3 f)^2}{|v_i|^2 \, |\nabla_3 f|^2} \right). \quad (6)$$

This probability has the following properties:

- It is maximal if the vectors v_i and $\nabla_3 f$ are orthogonal.

- It is minimal if the two vectors are parallel.

---

[1] For a discussion of some alternative approaches, we refer to [16].

- 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.

In particular, this last property is crucial for the case of motion *segmentation* considered here. In contrast, the probability associated with the classical least squares formulation in the approach of Horn and Schunck *depends* on the length of the two vectors. Firstly, this induces a bias toward velocities of larger magnitude when segmenting several differently moving regions. And secondly, this implies that a spatio-temporal gradient which is not orthogonal to v_i is less likely the larger its magnitude.

To account for the case where the spatio-temporal gradient vanishes, we introduce the regularization

$$|\nabla_3 f| \ \rightarrow\ |\nabla_3 f| + \epsilon \qquad (7)$$

in the denominator of (13). The small constant $\epsilon > 0$ can be interpreted as a noise scale of the image gradient. We did not observe a noticeable influence of the precise choice of ϵ in numerical experiments.

We discretize the velocity field on a set of disjoint regions $D_i \subset D$, separated by a surface S:

$$v(x,t) = \{v_i, \text{ if } (x,t) \in D_i\}. \qquad (8)$$

Moreover, we assume a prior $\mathcal{P}(v)$ on the velocity field which only depends on the separating interface S and favors interfaces of minimal surface area $|S|$:

$$\mathcal{P}(v) \propto \exp(-\nu |S|). \qquad (9)$$

3. Piecewise Parametric Motion

In the above formulation, we restricted the motion model (8) to spatio-temporal domains of piecewise constant motion. However, the same geometric reasoning can be generalized to piecewise parametric motion models. The velocity on the domain D_i is then permitted to vary in space and time according to a predefined model of the form:

$$v_i(x,t) = M(x,t)\, p_i, \qquad (10)$$

where M is a matrix depending only on space and time and p_i is the parameter vector associated with each region. Particular examples are the case of *spatially affine motion* with the matrix

$$M(x,t) = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad (11)$$

and a parameter vector $p_i = (a_i, b_i, c_i, d_i, e_i, f_i, 1)$ for each domain D_i, and the case of *accelerated motion* with

$$M(x,t) = \begin{pmatrix} 1 & 0 & t & 0 & 0 \\ 0 & 1 & 0 & t & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad (12)$$

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

Inserting model (10) into the optic flow constraint (5) gives a relation which – again interpreted geometrically – states that the the vector $M^t \nabla_3 f$ must either vanish or be orthogonal to the parameter vector p_i. We model the conditional probability that the point $(x, t) \in D$ belongs to the domain D_i by:

$$P(\nabla_3 f \,|\, p_i) \propto \exp\left(-\frac{(p_i^t M^t \nabla_3 f)^2}{|p_i|^2 \, |M^t \nabla_3 f|^2}\right). \qquad (13)$$

4. Variational Sequence Segmentation

Maximizing the conditional probability (4) is equivalent to minimizing its negative log likelihood. With formulas (9) and (13), we obtain an energy of the form:

$$E(S, \{p_i\}) = \sum_i \int_{D_i} \frac{|p_i^t M \nabla_3 f|^2}{|p_i|^2 |M \nabla_3 f|^2} \, dx \, dt \,+\, \nu \, |S|. \quad (14)$$

This functional couples region-based motion information and boundary information in a similar way as does the Mumford-Shah functional [12] for the case of gray value segmentation.

Simultaneous minimization with respect to the motion parameters p_i and the separating surface S jointly solves the problems of motion estimation and segmentation of an image sequence. The motion in the sequence is approximated by a piecewise parametric motion over a set of spatio-temporal domains D_i. Note that the proposed functional contains only one free parameter ν, representing a scale parameter which is fundamental to all segmentation methods (cf. [11]).

5. A Multiphase Level Set Framework

In this section, we propose to implement the separating hypersurface $S \subset \mathbb{R}^3$ in a multiphase level set framework which is based on the corresponding gray value model of Chan and Vese [3]. Compared to alternative multiphase models, this one permits to compactly represent up to n motion phases by only $\log_2(n)$ level set functions. Moreover, it does not suffer from overlap or vacuum formation, since all points are by definition ascribed to one of n phases.

Let Φ denote a vector-valued level set function:

$$\Phi = (\phi_1, \ldots, \phi_m), \quad \text{where } \phi_i : D \subset \mathbb{R}^3 \to \mathbb{R}, \quad (15)$$

and let $H(\Phi) = (H(\phi_1), \ldots, H(\phi_m))$ be the associated vector Heaviside function, with:

$$H(\phi_i) = \begin{cases} 1 & \text{if } \phi_i \geq 0 \\ 0 & \text{if } \phi_i < 0 \end{cases} . \qquad (16)$$

The function $H\Phi$ maps each point (x, t) in space-time to a binary vector. It therefore permits to encode a set of $n = 2^m$ phases defined by:

$$D_i = \{(x, t) \in D \mid H(\Phi(x, t)) = \text{constant}\}, \qquad (17)$$

with the separating hypersurface given by:

$$S = \{(x, t) \in D \;\mid\; \exists\, i \text{ with } \phi_i(x, t) = 0\}. \qquad (18)$$

With this representation, the cost functional (14) can be replaced by the *multiphase functional*:

$$\begin{aligned} E(\Phi, \{p_i\}) \;=\; & \sum_{i=1}^{n} \int_{D} \frac{p_i^t T p_i}{|p_i|^2} \, \chi_i(\Phi) \, dx \, dt \\ & + \; \nu \sum_{j=1}^{m} \int_{D} \big|\nabla H(\phi_j)\big| \, dx \, dt \qquad (19) \end{aligned}$$

where χ_i denotes the indicator function for the region D_i and the matrix T is given by:

$$T(x, t) = \frac{M^t \nabla_3 f \, \nabla_3 f^t M}{|M^t \nabla_3 f|^2}. \qquad (20)$$

For illustrative purposes, we will detail this for the *two-phase model* which is given by:

$$\begin{aligned} E(\phi, p_1, p_2) \;=\; & \int_{D} \frac{p_1^t T p_1}{|p_1|^2} \, H(\phi) \, dx \, dt \\ & + \int_{D} \frac{p_2^t T p_2}{|p_2|^2} \, (1 - H(\phi)) \, dx \, dt \\ & + \; \nu \int_{D} \big|\nabla H(\phi)\big| \, dx \, dt, \qquad (21) \end{aligned}$$

with a single level set function ϕ separating two spatio-temporal motion phases

$$D_1 = \{(x, t) \,|\, \phi \geq 0\}, \quad D_2 = \{(x, t) \,|\, \phi < 0\}, \quad (22)$$

with motion parameters p_1 and p_2.

6. Motion Competition in Space-Time

The cost functional (19) must be simultaneously minimized with respect to the level set functions ϕ_i and with respect to the motion parameters p_i. We implement this minimization by alternating the two fractional steps of:

- Motion estimation

  For fixed Φ, minimization with respect to the motion parameters p_i results in a generalized eigenvalue problem of the form:

$$p_i = \arg\min_p \frac{p^t T_i \, p}{p^t p}, \qquad (23)$$

where T_i denotes the matrix T in (20) integrated over region D_i:

$$T_i = \int_D T(x,t)\,\chi_i\,dx\,dt\,. \qquad (24)$$

Therefore, p_i is given by the eigenvector associated with the smallest eigenvalue of T_i, normalized so that its third component is 1. Note that the motion is estimated over the largest supporting region, namely over the entire region D_i. This is an important property of our approach, because it permits to circumvent the problem of estimating motion locally.

- Surface evolution

Conversely, for fixed motion parameters $\{p_i\}$, minimization of (19) with respect to the separating hypersurface S can be implemented by iterating a gradient descent for the level set functions ϕ_i:

$$\frac{\partial \phi_i}{\partial t} = -\frac{\partial E}{\partial \phi_i}. \qquad (25)$$

We will illustrate this for the case of the two-phase model (21), the extension to multiple phases is straightforward. For $n = 2$ phases, the embedding function evolves according to:

$$\frac{\partial \phi}{\partial t} = \delta(\phi)\left[\nu\,\mathrm{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + e_2 - e_1\right]. \qquad (26)$$

While the first term aims at minimizing the area of the separating hypersurface, the last two terms can be interpreted in that neighboring domains compete for the separating surface in terms of their motion energy densities e_i given by

$$e_i(x,t) = \frac{p_i^t\,T(x,t)\,p_i}{p_i^t p_i}. \qquad (27)$$

This competition process is what gave rise to the term *motion competition*.

As suggested in [3], we implement the delta function $\delta(\phi) = \frac{d}{d\phi}H(\phi)$ by a smooth approximation of finite width τ:

$$\delta_\tau(s) = \frac{1}{\pi}\frac{\tau}{\tau^2 + s^2}\,. \qquad (28)$$

7. Numerical Results

In the following, we will present a number of results which are chosen to highlight different properties of our approach. These results are based on a model of piecewise constant motion. We found no improvement in the results by reverting to more general classes of motion fields. On the contrary, preliminary results indicate that increasing the number of parameters to model the motion of each region tends to decrease the robustness of boundary evolutions.

7.1. Intensity versus Motion Segmentation

Although conceptually similar to the Mumford-Shah functional for gray value segmentation, the motion segmentation functional (14) induces fundamentally different segmentation results. Figure 1 shows one frame from a sequence showing a rabbit moving on a table.

Superimposed on the image is the evolution of the separating interface (with a particular initialization) and the corresponding estimated piecewise constant motion field, obtained with the two-phase model. Note that during the optimization of the domain boundary the motion estimates are gradually improved.

The segmentation process obtained for one of the frames from the sequence using the corresponding piecewise constant gray value functional [3] is shown in Figure 2. Due to the difficult lighting conditions, intensity-based segmentation fails to capture the object of interest.

7.2. Dependence on Initialization

The cost functional (14) is generally not convex. Therefore, the final segmentation will depend on the initialization. Figure 3 shows the evolution of motion estimates and boundary obtained for the rabbit sequence by minimizing the two-phase model with an initialization which is shifted to the right. Note that the final segmentation and motion estimate are fairly similar. The comparison of Figures 1 and 3 demonstrates how the fundamental limitations of motion estimation imposed by the aperture problem are compensated by the regularization of the separating interface: Regions of weak gray value structure are associated with one or the other motion region according to the area constraint on the motion boundary.

7.3. Multiphase Motion Segmentation

The above examples showed segmentation results obtained for the two-phase model. Moreover, only one object was moving in an otherwise static scene. In the following, we will apply our approach to the flower garden sequence which shows a complex scene filmed by a moving camera. Different parts of the image plane are undergoing different two-dimensional motion, depending on their relative position to the camera.

Figure 4 shows the evolving boundary and motion estimate obtained by minimizing the two-phase model. Note that the final segmentation clearly separates the tree in the foreground from the remainder of the image which is associated with an average background velocity.

A more detailed segmentation of the motion in this sequence is obtained by minimizing the four-phase model. The corresponding evolution of motion estimates and region
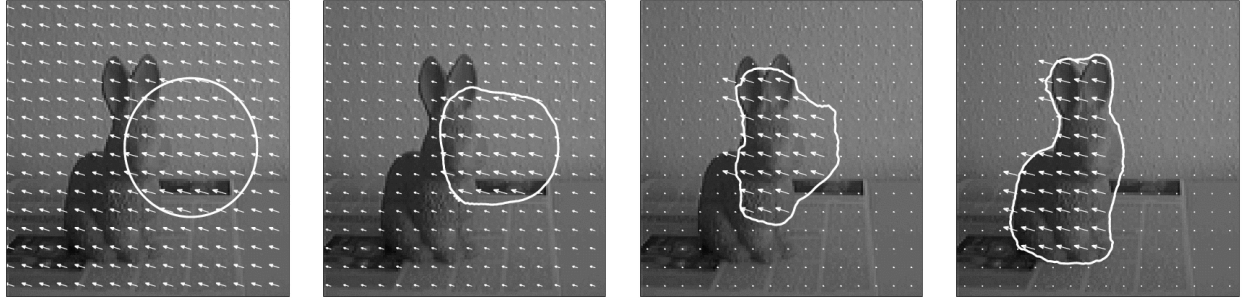
Figure 1. Motion segmentation. Evolution of the estimated motion field and motion boundary, superimposed on one frame of a sequence showing a rabbit moving to the left. By minimizing a single cost functional, both the domain boundary and the estimated motion are progressively improved. The final segmentation captures the shape of the rabbit and the motion of object and background.
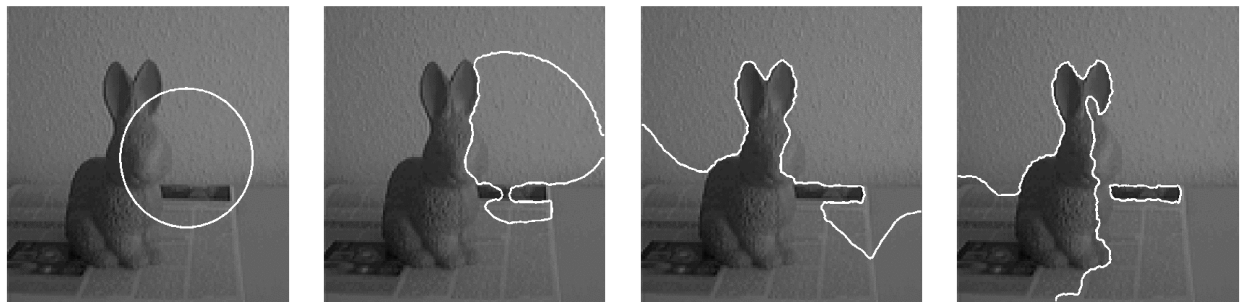

Figure 2. Intensity segmentation. Boundary evolution for one frame of the rabbit sequence, obtained with the corresponding two-phase gray value model [3]. Due to the difficult lighting conditions, the object is not well-defined in terms of constant intensity. The respective segmentation process therefore fails to capture the object of interest.
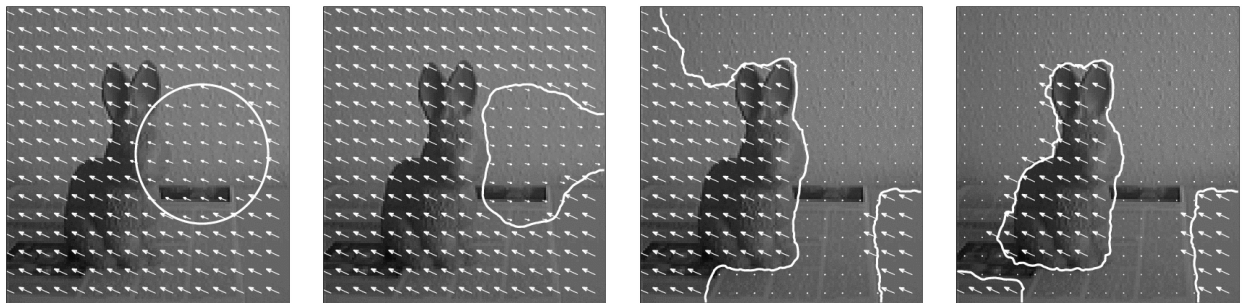

Figure 3. Dependence on initial contour. Evolution of estimated motion field and motion boundary for the same sequence used in Figure 1, but with an initialization which is shifted to the right. The final segmentation and motion estimate obviously depend on the initialization: Regions which are not sufficiently structured to permit reliable motion estimation are ascribed to one or the other domain, according to the area constraint on the separating interface.

boundaries is shown in Figure 5. The zero level sets of the two functions ϕ_1 and ϕ_2 (represented by the black and the white contour) progressively separate foreground, midplane and background, based exclusively on their relative motion. In the final segmentation, the estimated regions model the motion of the tree, the grass, and the background. The sky region does not contain sufficient gray value structure and is therefore ascribed to one of the regions according to the

boundary constraint. Both the segmentation and the estimated motion for foreground, midplane and background are quite accurate, given that we merely minimize a single cost functional defined on the spatio-temporal image gradients. In contrast, most alternative approaches to layer extraction perform extensive preprocessing such as local disparity estimation, camera calibration and prior rectification of the individual frames (cf. [17, 15]).
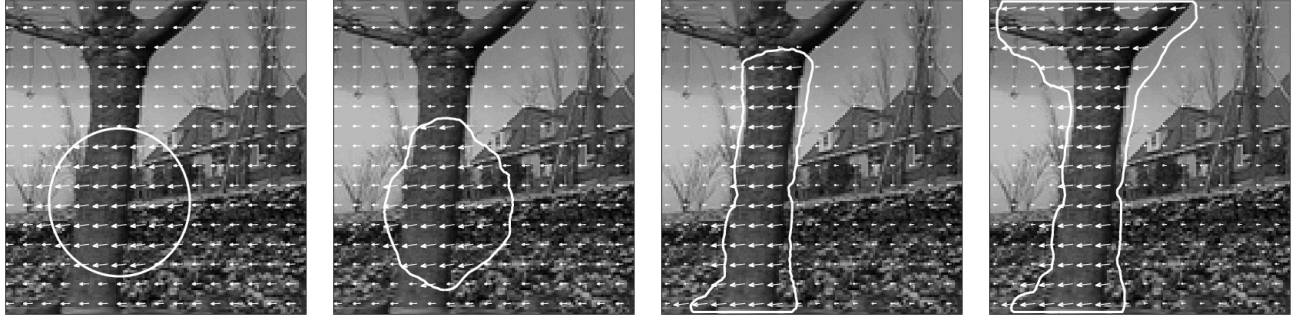
Figure 4. Two motion phases. Motion segmentation of the flower garden sequence generated by minimizing the two-phase model (21). The two motion regions separate the tree in the foreground from the differently moving background.
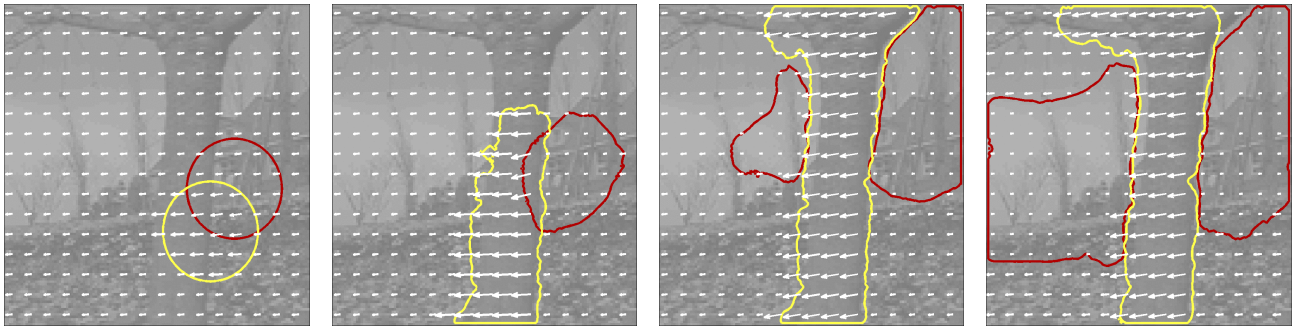

Figure 5. Four motion phases. Segmentation of the flower garden sequence with the four-phase model. Compared to the result with the two-phase model shown in Figure 4, this model generates a more detailed reconstruction of the background layers. The final segmentation separates the tree in the foreground, the grass in the midplane and the houses and smaller trees in the background. Note that both the boundary information and the motion estimates are obtained by simultaneously minimizing an appropriate cost functional which is defined on the spatio-temporal image derivatives. Unlike most alternative approaches to layer extraction, no preprocessing is applied to the image data.

7.4. Spatio-temporal Motion Segmentation

In the previous figures, we have only shown the segmentation and motion estimates for one frame of the sequences. Yet, our approach generates a segmentation of the entire spatio-temporal volume. For the four-phase model, Figure 6 shows the evolution of the surfaces separating the motion regions in space-time (top rows). The lower rows depict the temporal slices of these surfaces associated with the frames 2, 5 and 8 of the sequence. During energy minimization, the surfaces propagate to the final segmentation both in space and in time. The final segmentation clearly separates foreground, midplane and background. For better visibility, the simultaneously estimated motion fields are not shown.

8. Limitations and Future Work

Although the results presented here are quite promising, the proposed approach relies on a number of assumptions. Firstly, the optic flow constraint (5) holds for small displacements only. Given larger displacements (i.e. faster moving objects or a slower frame rate of the camera), one may need to revert to multiscale formulations (cf. [6, 20]).

Secondly, the results presented here are approximations of the image motion by piecewise constant motion fields. However, we found that extending the model complexity of motion models tends to reduce the robustness of boundary evolutions. As in all model fitting problems, one therefore has to find a trade-off between model simplicity and sufficient generality.

We want to point out that our approach does not make any assumptions about the relation between the estimated region motion and the estimated boundary motion. This turns out to be important for modeling multiple motion and occlusion in the example of the flower garden sequence. Another example in which region and boundary do not undergo the same motion is that of a rotating cylinder in a 2d projection. On the other hand, explicit modeling of the motion of occluding boundaries may be worth investigating.

9. Summary and Conclusions

We presented a variational method for segmenting a video sequence into spatio-temporal domains of homogeneous motion. To this end, we formulated the problem of motion estimation as one of Bayesian inference. We proposed a geometric interpretation of the optic flow constraint in terms of the angle between the spatio-temporal gradient and the homogeneous velocity vector. We argued that the associated conditional probability on the spatio-temporal gradient is superior to the commonly considered least-squares criterion in the context of motion *segmentation*. Our approach constitutes a generalization of the *motion competition* framework [4] to the space-time domain.

Based on a few explicitly stated assumptions, we derived a cost functional which depends on the motion parameters for a set of disjoint regions and on a hypersurface separating these regions in space-time. We proposed an implementation of this functional by a multiphase level set framework, in which a set of up to n motion regions and the separating hypersurface are modeled implicitly by a set of $\log_2 n$ level set functions. Minimization of the proposed cost functional with respect to its dynamic variables results in an eigenvalue problem for the parameters modeling the motion in each region, and in a gradient descent evolution for the level set functions embedding the separating hypersurface.
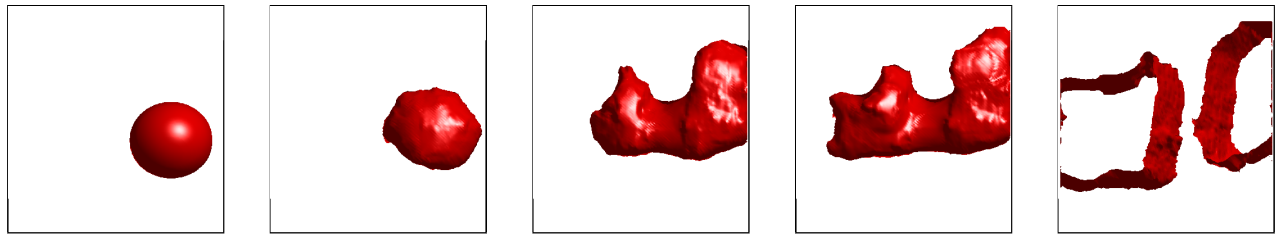
Numerical results on real-world image sequences demonstrate that minimizing a single cost functional can generate a sensible segmentation of the motion in a video sequences. Compared to alternative approaches of layer extraction, we do not perform any preprocessing such as local disparity estimation, camera calibration or prior rectification of individual frames. Instead, all results are obtained by minimizing a cost functional which exclusively depends on the spatio-temporal image gradients of the sequence.
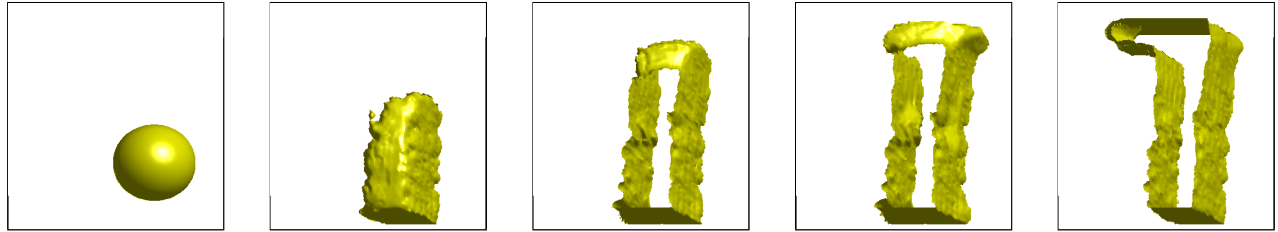
Acknowledgments

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 variational framework for image segmentation combining motion estimation and shape regularization. In C. Dyer and P. Perona, editors, *IEEE CVPR*, volume 1, pages 53–58, Madison, June 2003.

[5] G. Farnebäck. Very high accuracy velocity estimation using orientation tensors, parametric motion, and segmentation of the motion field. In *ICCV*, volume 1, pages 171–177, 2001.

[6] F. Heitz, P. Perez, and P. Bouthemy. Multiscale minimization of global energy functions in some visual recovery problems. *Comp. Vis. Graph. Image Proc.: IU*, 59(1):125–134, 1994.

[7] B. Horn and B. Schunck. Determining optical flow. *A.I.*, 17:185–203, 1981.

[8] P. Kornprobst, R. Deriche, and G. Aubert. Image sequence analysis via partial differential equations. *J. Math. Im. Vis.*, 11(1):5–26, 1999.

[9] 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.

[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] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhäuser, Boston, 1995.

[12] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

[13] J.-M. Odobez and P. Bouthemy. Direct incremental model-based image motion segmentation for video analysis. *Signal Proc.*, 66:143–155, 1998.

[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] H.-Y. Shum and R. Szeliski. Construction of panoramic mosaics with global and local alignment. *Int. J. of Comp. Vis.*, 36(2):101–130, 2000.

[16] E. Simoncelli. *Distributed Representation and Analysis of Visual Motion*. PhD thesis, Dept. of Elect. Eng. and Comp. Sci., MIT, Cambridge, 1993.

[17] P. H. S. Torr, R. Szeliski, and P. Anandan. An integrated bayesian approach to layer extraction from image sequences. *IEEE PAMI*, 23(3):297–303, 2002.

[18] A. Tsai, A. Yezzi, and W. A.S. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Trans. on Image Processing*, 10(8):1169–1186, 2001.

[19] J. Wang and E. Adelson. Representing moving images with layers. *IEEE Trans. on Image Processing*, 3(5):625–638, 1994.

[20] J. Weber and J. Malik. Robust computation of optical flow in a multi-scale differential framework. *Int. J. of Comp. Vis.*, 14(1):67–81, 1995.

[21] 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.

[22] A. Yuille and N. Grzywacz. *High-level Motion Processing, T. Watanabe (Ed.)*, chapter "A theoretical framework for visual motion". North Holland, 1996.

[23] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE PAMI*, 18(9):884–900, 1996.

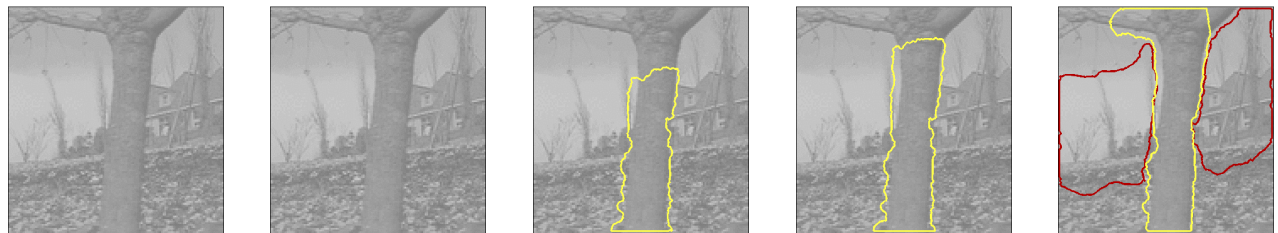**Evolution of the first spatio-temporal motion interface**

Evolution of the second spatio-temporal motion interface

Evolution for frame number 2

Evolution for frame number 5

Evolution for frame number 8

Figure 6. Spatio-temporal sequence segmentation for the four-phase model. The top rows show the evolution of the spatio-temporal surfaces given by the zero level sets of the embedding functions ϕ_1 and ϕ_2. The lower rows show various temporal slices of these surfaces, corresponding to the 2nd, 5th and 8th frame of the sequence. Note that the evolving surfaces propagate both in space and time during energy minimization. Due to the level set representation, topological changes of the evolving surface are facilitated – see the last two frames of the top row. In the final segmentation the phases clearly separate foreground, midplane and background. For better visibility, the simultaneously estimated piecewise constant motion field is not shown here.