Statistical Shape Knowledge in Variational Motion Segmentation

Daniel Cremers and Christoph Schnörr

Computer Vision, Graphics, and Pattern Recognition Group
Department of Mathematics and Computer Science
University of Mannheim, D–68131 Mannheim, Germany
http://www.cvgpr.uni–mannheim.de/{cremers,schnoerr}

Abstract— **We present a generative approach to model–based motion segmentation by incorporating a statistical shape prior into a novel variational segmentation method. The shape prior statistically encodes a training set of object outlines presented in advance during a training phase.**

In a region competition manner the proposed variational approach maximizes the homogeneity of the motion vector field estimated on a set of regions, thus evolving the separating discontinuity set. Due to the shape prior, this discontinuity set is not only sensitive to motion boundaries but also favors shapes according to the statistical shape knowledge.

In numerical examples we verify several properties of the proposed approach: for objects which cannot be easily discriminated from the background by their appearance, the desired motion segmentation is obtained, although the corresponding segmentation based on image intensities fails. The region–based formulation facilitates convergence of the contour from its initialization over fairly large distances, and the estimated flow field is progressively improved during the gradient descent minimization. Due to the shape prior, partial occlusions of the moving object by "unfamiliar" objects are ignored, and the evolution of the motion boundary is effectively restricted to the subspace of familiar shapes.

Keywords: statistical learning, variational methods, motion segmentation, Mumford–Shah functional, region competition, shape recognition, diffusion snake

I. Introduction

The variational approach to image segmentation proposed by Mumford and Shah [17] consists in minimizing the following energy functional:

$$E(u,C)=\frac{1}{2}\int_{\Omega}(f-u)^2dx+\lambda^2\frac{1}{2}\int_{\Omega-C}|\nabla u|^2dx+\nu||C||. \quad (1)$$

The input image f is approximated on the image plane Ω by a piecewise smooth function u, which may be discontinuous across a boundary C. The length of this boundary is commonly implemented as

$$||C|| = \int_0^1 \left|\frac{dC(s)}{ds}\right| ds. \quad (2)$$

By further enforcing the constraint of smoothness ($\lambda \to \infty$), one approaches the so–called cartoon limit, in which the input image f is approximated by a piecewise constant function $u(x)$ = constant u_i if $x \in R_i \subset \Omega$:

$$E(u,C) = \sum_i \int_{R_i} (f-u_i)^2\, dx \;+\; \nu\, ||C||\,. \quad (3)$$

The contour then partitions the image plane into a set of pairwise disjoint regions R_i, with u_i taking on the mean grey value of f over the region R_i.

In [7], [6] we proposed modifications of the two functionals (1) and (3) which permitted to implement the segmenting contour as a closed quadratic B–spline curve. These so–called *diffusion snakes* are a hybrid model with the external energy of the Mumford–Shah functional and the internal energy of the snakes [13]. We then proposed to extend this functional by introducing a statistical prior on the shape of the segmenting contour in a combined variational approach. For this purpose we estimated the distribution of spline control points associated with a set of binary training objects. Numerical results showed that the additional shape energy can improve segmentation results in case of noise, clutter or occlusion.

In the present paper we propose to modify the functional (3) so that it measures the inhomogeneity *not* with respect to the image intensities, but rather with respect to a motion hypothesis. As we will show, this provides an elegant way to jointly solve the problems of motion estimation and image segmentation. Just as in the case of the corresponding grey value segmentation, an additional shape energy effectively restricts the contour evolution to the subspace of familiar contours induced by the statistical shape model. This can improve segmentation of a known object in cases of missing or misleading motion information.

Related Work

Discontinuity–preserving motion estimation by variational models and related partial differential equations have a long tradition in computer vision [18], [23], [2], [16], [15], [25]. These approaches are non–generative in that they purely work in a data–driven way. Moreover, they generally model the motion discontinuities implicitly in terms of appropriate (non–quadratic) regularizers.

There exist some variational approaches with explicit discontinuities for grey value segmentation [17] and extensions to color and texture segmentation [26]. The functional (1) has been adapted to the problem of motion segmentation in [18], however there the author prefers an implicit model of the discontinuity by reverting to approximations in terms of Γ–convergence as studied in [1].

Though not derived by minimizing a single energy functional, the approaches in [22], [24], [10] provide similar dynamics by combining variational motion estimation on disjoint sets with a shape optimization procedure. However, like the above approaches the latter approaches do not exploit prior knowledge.

Prior knowledge in terms of motion models was incorporated in motion estimation and motion segmentation by [19], [20]. In contrast to this approach, we focus on prior knowledge with respect to shape and thus directly address the problem of determining accurate motion boundaries in a generative way. Deformable shape models were combined with motion segmentation in [14]. However, there the authors do not pursue a variational integration of motion segmentation and shape prior, they work in the subspace of a small number of model parameters which are optimized by simulated annealing. This restriction to a few model parameters does not permit a direct comparison to segmentation results without the statistical prior.

Concerning the representation of shape, we are well aware of the advantages of level–set based approaches [21]. From the statistical learning perspective, however, the additional dimension introduced by the implicit representation of shape as level sets is a serious drawback. Furthermore, in many applications it is known that topological changes of shapes do not occur.

II. The Generative Model

A. *Variational Integration of Motion Information and Shape Prior*

The proposed variational combination of shape statistics and motion information consists in minimizing an energy which is the weighted sum of a motion energy E_m and a shape energy E_c:

$$E\left(C(z)\right) = E_m\left(w, C(z)\right) + \alpha\, E_c(z), \tag{4}$$

with respect to the parameters z defining a closed spline curve C and the motion vectors $w = \{w_i\}$ which estimate the motion in the regions R_i separated by the contour. The motion energy measures the inhomogeneity of motion in the regions separated by the contour, and the shape energy measures the dissimilarity of a given contour with respect to a set of training shapes. Both terms will be detailed in the next two subsections.

Our approach differs from most shape–model based segmentation approaches. We do not enforce a *hard* restriction of the evolving contour to the subspace of familiar shape deformations, for example by working only with the parameters associated with the principal modes of a principal component analysis. Instead, we embed the shape prior $E_c(z)$ as a *soft* constraint defined (and differentiable) on the full space of possible spline curves, as will be detailed below. This approach has a number of favorable properties:

Firstly, it allows a direct comparison of contour evolutions *with* and *without* the prior, because the prior can be switched off by a single parameter α.

Secondly, the prior can be switched on during a contour evolution without further modifications. This enables an encoding of very different object classes in a single shape prior [6]: The contour is first evolved without the prior until stationarity and after switching on the prior, the evolution favors the most appropriate object class.

Thirdly, we believe that the soft constraint is more faithful from a probabilistic point of view: given only a small number of training samples in the learning set, a shape deformation orthogonal to the observed ones should not be assigned a probability of zero (or, equivalently, an infinite energy), as done in the case of the hard constraint.

B. *Variational Motion Segmentation*

The Mumford–Shah approach (3) aims at maximizing a homogeneity criterion with respect to the grey value on a set of regions R_i which are separated by a contour C. We now propose to maximize the homogeneity *not* with respect to the *grey value* but rather with respect to the *estimated motion.*

Let $f(x,t)$ be an image sequence which is assumed to be differentiable. Assuming that the intensity of a moving point is constant throughout time, we obtain a continuity equation given by the classical optic flow constraint:

$$\frac{d}{dt} f(x,t) = \frac{\partial}{\partial t} f + w^t \nabla f = 0,$$

where $w = \frac{dx}{dt}$ denotes the local velocity. Given two consecutive images f_1 and f_2 from this sequence, we approximate $\frac{\partial}{\partial t} f \approx (f_2 - f_1)$ and $\nabla f \approx \frac{1}{2}\nabla(f_1 + f_2)$.

We propose to segment the image into areas R_i of piecewise constant motion w_i by minimizing the energy functional

$$E_m(w,C) = \sum_i \int_{R_i} \left(f_2 - f_1 + \frac{w_i^t}{2} \nabla(f_1 + f_2) \right)^2 dx + \nu\, ||C|| \tag{5}$$

simultaneously with respect to both the contour C and the motion vectors w_i. The proposed motion energy (5) can be interpreted as an extension of the Mumford–Shah cartoon limit (3) to the problem of motion segmentation.

With the extended velocity vector $v = \binom{w}{1}$ and the spatio–temporal structure tensor

$$S = (\nabla_3 f)(\nabla_3 f)^t, \quad \text{with} \quad \nabla_3 f = \begin{pmatrix} \nabla f \\ \frac{\partial}{\partial t} f \end{pmatrix}, \tag{6}$$

the energy (5) can be rewritten as

$$E_m(w,C) = \sum_i \int_{R_i} \left(v_i^t\, S\, v_i \right) dx \; + \; \nu\, ||C||. \tag{7}$$

In practice, the proposed homogeneity term shows a bias towards velocity vectors of large magnitude. As proposed in [9], we therefore perform an isotropy compensation of the structure tensor by replacing

$$S \;\longrightarrow\; S - \lambda_3\, I, \tag{8}$$

where λ_3 is the smallest eigenvalue of S and I is the 3×3 unit matrix. This modification removes a term proportional to $|v_i|^2$ from the cost functional (7) and thereby decreases the bias towards areas of large magnitude of the velocity vector. For a more detailed exposition we refer to [9].

C. Motion Competition

Minimization of (7) with respect to the velocities $w = \{w_i\}$ results in the equations

$$\frac{dE_m}{dw_i} = 2\,\bar{S}_i\,w_i \,+\, 2\,b_i = 0, \tag{9}$$

with

$$\bar{S}_i = \int\limits_{R_i} \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} dx \quad \text{and} \quad b_i = \int\limits_{R_i} \begin{pmatrix} S_{13} \\ S_{23} \end{pmatrix} dx. \tag{10}$$

The solution of (9) is given by

$$w_i = -\bar{S}_i^{\,-1}\,b_i. \tag{11}$$

If the matrix $\bar{S}_i$ is not invertible, we revert to the pseudo–inverse in order to choose the solution of smallest magnitude.

Using Green's theorem, minimization of (7) with respect to the contour C results in the evolution equation

$$\frac{dC}{dt} = \left(e^- - e^+\right) n - \nu\,\kappa_c\,n. \tag{12}$$

The superscripts $j = +/-$ denote the two regions to the left and to the right of the respective contour point (in the sense of the contour parameterization). They compete for the contour in terms of the associated energy densities

$$e^j = v_j^t\,S\,v_j. \tag{13}$$

The curvature of the contour is denoted by κ_c, and n is the normal on the contour pointing out of the region R_+.

The two forces in equation (12) which drive the contour evolution have the following intuitive interpretation: the first one enforces regions of homogeneous optic flow, thus separating regions moving at different velocities w_j, whereas the second term enforces a smoothing of the separating contour by minimizing its curvature.

Since the two regions adjacent to the respective contour point compete in terms of their energy densities e^+ and e^-, we refer to this process as *motion competition.*

Note that in the equivalent paradigm of *Bayesian inference*, these energy densities can also be interpreted as the negative log likelihoods associated with the probability that a given location belongs to one or the other motion hypothesis.

For comparing different motion hypotheses, it is suggested in [9] to normalize the cost function in (13) by replacing

$$v_j^t\,S\,v_j \;\longrightarrow\; \frac{v_j^t\,S\,v_j}{||v_j||^2\;trS}. \tag{14}$$

Although this modification is not strictly derived by minimizing energy (7), it tends to slightly improve the contour evolution.

D. Statistical Shape Energy

In the following we will restrict the space of possible motion contours to closed spline–curves of the form:

$$C : [0,1] \to \Omega, \quad C(s) = \sum_{n=1}^{N} p_n\,B_n(s), \tag{15}$$

with spline control points $p_n = (x_n\,,\,y_n)^t$ and periodic quadratic basis functions B_n [3], [8]. This facilitates the incorporation of a statistical shape prior on the control point vector $z = (x_1, y_1, \ldots, x_N, y_N)^t$.

In order to prevent a (numerically undesirable) clustering of spline control points during the contour evolution we propose [7] to replace the usual contour length (2) by

$$||C|| := \int\limits_0^1 \left(\frac{dC}{ds}\right)^2 ds. \tag{16}$$

In fact, the Euler–Lagrange equation corresponding to the functional (16) is equivalent to an equidistant spacing of control points in the case of quadratic B–spline basis functions:

$$\frac{d^2C(s)}{ds^2} = 0 \;\forall s \quad \Longleftrightarrow \quad p_i = \frac{p_{i-1} + p_{i+1}}{2} \;\forall i\,.$$

Due to the modified length measure (16), the curvature term $(\kappa_c n)$ in the evolution equation (12) is replaced by $\frac{d^2C}{ds^2}$.

The explicit parameterization of the contour (15) allows to represent a set of sample shapes in a vector space and to approximate their distribution statistically. To this end, the images of training objects are binarized, a spline contour is fitted to the boundary and the set of training contours is aligned with respect to similarity transformations [11] and cyclic permutation of the control points $p_1, \ldots, p_N$.

The distribution of control point vectors $z \in \mathbb{R}^{2N}$ is assumed to be Gaussian:

$$\mathcal{P}\,(z) \,\propto\, \exp\left(-\frac{1}{2}(z - z_0)^t\,\Sigma^{-1}\,(z - z_0)\right). \tag{17}$$

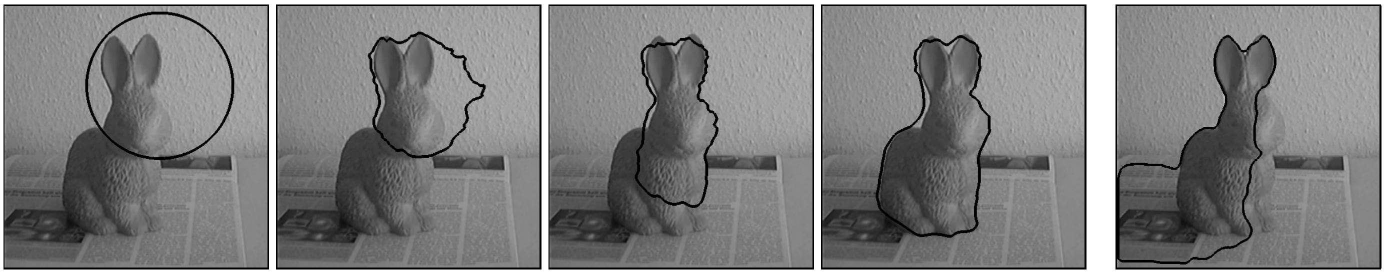
The mean control point vector z_0 and sample covariance matrix Σ are determined for the training set.

If the dimension of the subspace spanned by the training vectors is smaller than the dimension $2N$ of the underlying vector space, the sample covariance matrix Σ will not have full rank. The associated Gaussian probability will vanish for any shape outside the spanned subspace.

As discussed in [7], we therefore define a probability density which is continuous and differentiable in the full $2N$–dimensional space by regularizing the covariance matrix:

$$\Sigma_\perp \;=\; \Sigma \,+\, \sigma_\perp\,\left(I \,-\, VV^t\right),$$

where V is the matrix of eigenvectors of Σ corresponding to nonzero eigenvalues, and $\sigma_\perp \in (0, \sigma_r]$ is a regularizing constant, σ_r being the smallest non–zero eigenvalue of Σ.



initial intermediate steps (motion) final grey value only

Fig. 1. Motion segmentation versus grey value segmentation: Contour evolution for model (7) of piecewise homogeneous motion and final contour for model (3) of piecewise homogeneous intensity (right). Due to the lighting conditions, the image intensity is not a reliable cue for segmentation. The discrepancy of the final motion contour at the left side of the head is partly due to the fact that the tapestry is not sufficiently structured (compared to the newspaper on the bottom right) to transmit the correct motion information. This fundamental limitation of motion estimation is commonly referred to as the *aperture problem.* Moreover, in the first frame of the sequence (not shown here) the figure of the rabbit is further to the left.

This regularization corresponds to replacing the zero eigenvalues by a constant value which we fixed to $\sigma_\perp = 0.5\, \sigma_r$ in all numerical studies.

The negative logarithm of the probability (17) can be interpreted as a shape energy of Mahalanobis type:

$$E_c(z) = \frac{1}{2}\,(z - z_0)^t\, \Sigma_\perp^{-1}\,(z - z_0). \tag{18}$$

III. Parameter Estimation

A. Gradient Descent Evolution

To obtain a variational approach which maximizes both the homogeneity of motion inside and outside the contour and the similarity of the contour to a set of training shapes, we propose to minimize the total energy (4) with respect to the spline control point vector z and the motion vectors $w = \{w_i\}$.

Equation (12) can be converted to an evolution equation for the spline control points by inserting the definition (15) of the contour as a spline curve. Then the equation is discretized with a set of nodes s_i along the contour, where s_i is chosen as the point where the spline basis function B_i attains its maximum. Including the contribution of the shape energy (18), we obtain for the x–coordinates of control point m:

$$\begin{aligned} \frac{dx_m(t)}{dt} \;=\; & \sum_k \left(\mathbf{B}^{-1}\right)_{mk} \Big[\left(e^+_{s_k} - e^-_{s_k}\right) n_x(s_k) \\ & \qquad + \nu\,(x_{k-1} - 2x_k + x_{k+1}) \Big] \\ & - \alpha \left[\Sigma_\perp^{-1}\,(z - z_0)\right]_{2m-1}, \end{aligned} \tag{19}$$

where n_x denotes the x–coordinate of the normal vector and the index $2m-1$ refers to the component of the given vector which is associated with the x–coordinate of control point m. The cyclic tridiagonal matrix $\mathbf{B}$ contains the spline basis functions evaluated at the nodes: $B_{ij} = B_i(s_j)$. A similar expression holds for the y–coordinates.

The three terms in (19) can be interpreted as follows:

The first term forces the contour towards the boundaries of the two motion fields by minimizing the inhomogeneity which is measured by the energy density (13) with the normalizations (8) and (14). The second term enforces an equidistant spacing of control points, thus minimizing the length measure (16). And the third term causes a relaxation towards the most probable shape which is induced by the shape energy (18). Note that the relaxation towards the most probable shape is weighted by the inverse of the regularized covariance matrix, such that less familiar shape deformations will decay faster.

B. Similarity Invariant Shape Prior

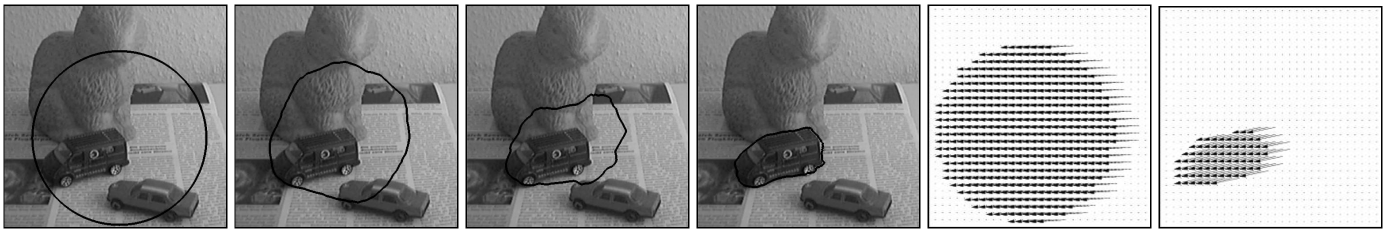
In most applications it is desirable to measure shape dissimilarity only up to certain group transformations. In [6] we presented a closed–form solution for incorporating such invariances into the variational approach (18).

To incorporate similarity invariance, the training shapes are simultaneously aligned with respect to translation, scaling, rotation [4], [11] and cyclic permutation of the control points. The remaining shape variability is approximately encoded by mean and covariance matrix. The final shape energy is then obtained by applying the function (18) to the argument after alignment of the respective contour with respect to the mean shape z_0:

$$E_{shape}(z) = E_c\left(\frac{R(z - z_c)}{|R(z - z_c)|}\right), \tag{20}$$

where z_c denotes the center and R is the optimal (Procrustes) rotation of the centered shape with respect to the mean z_0. As shown in [6], the resulting expression (20) can be differentiated with respect to the control point vector z. The last term in the gradient descent equation in (19) is then replaced by the gradient of the total energy (20). For details, we refer to [6].

This incorporates similarity invariance on the basis of the control point polygons without any additional parameters to encode rotation angle, scale and translation. In contrast, a minimization of *explicit* pose parameters requires the balancing of associated gradient descent equations. In an experimental comparison, we found that this may introduce local minima and can have a negative effect on convergence and stability of the numerical implementation.



initial step 1 step 2 final init. flow estim. final flow estim.

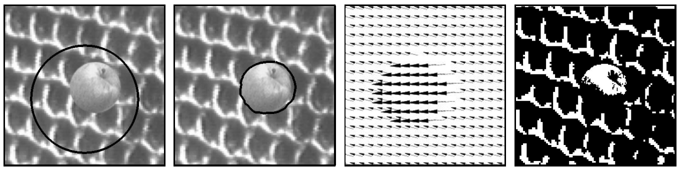
Fig. 2. Convergence over large distances: contour evolution for two images of a moving bus. The estimated flow field corresponding to the initial and final contour are shown in a close–up. The estimated object motion is gradually improved during the contour evolution.

IV. Experimental Results

Given two consecutive images of a motion sequence, we minimize the total energy (4) by iterating the contour evolution (19) in alternation with an update of the motion estimation (11) in the neighboring regions.

A. Intensity–based versus Motion–based Segmentation

The first example shows an artificial sequence of an apple which is translated, with the background translated at a different velocity and in a different direction. This can be considered a simplified analogue with the case of a moving object and a moving camera. Figure 3 shows the initial contour, the final contour and the final flow field estimation. The latter shows the two different motion fields which were estimated — namely two segments of motion with different magnitude and different direction.



initial final estim. flow binarized

Fig. 3. From left to right: initial and final contour for a motion segmentation of a moving apple and differently moving background, zoom of the estimated motion field corresponding to the final contour, and a binarized version of the (first) input image.

Although derived from a grey value segmentation approach, the proposed motion segmentation is substantially different from grey value segmentation in that it segments into regions of constant motion rather than constant grey value. Segmenting the previous example of the apple sequence based on grey value constancy would entirely fail as can be seen from the corresponding binarized image in Figure 3, right side: about half of the apple has disappeared although the background structure is still quite prominent.

The completely different properties of grey value and motion segmentation are demonstrated in Figure 1. A rabbit is moving with respect to the background. Due to the difficult lighting conditions in this example, the image grey value is not a good cue for segmentation and therefore segmentation based on grey value constancy fails — see Figure 1, right image. The segmentation based on motion constancy (with the same initialization) gives a better result — see Figure 1, second to last image.

B. Convergence over Large Distances

The example of the rabbit in Figure 1 shows a central property of our approach: since it is a region–based approach, it tends to converge over fairly large distances. This aspect is shown in the example of a moving bus in an otherwise static scene in Figure 2.

During the gradient descent minimization both the contour and the estimate of the flow field are improved simultaneously. The flow fields estimated for the initial and the final contour are shown in the last two images of Figure 2. Note that the final estimate of the motion of the bus is strongly improved compared to the initial one.

C. Moving Background

A central difficulty in motion estimation is the separation of differently moving regions. In many real–world applications, this problem arises for example if a moving object is filmed by a differently moving camera. Commonly the camera motion is eliminated by determining the dominant motion in a robust estimation framework first and then subtracting the latter [19], [2]. However, the assumption that the moving background fills the dominant part of the image plane may not always be valid.[1]

The variational approach (7) does not rely on any assumptions about the size of the segmented motion fields. In fact, examples such as the moving bus sequence in Figure 2 show that the object motion does not even have to fill the dominant part of the initially enclosed area for the minimization to converge correctly. This property arises due to the fact that both the contour evolution equation (19) and the motion estimation (11) were derived by minimizing a *single* energy functional. It is in fact the *non–robust* estimation of the motion inside and outside the contour which defines the driving force for the contour via the energy densities e^+ and e^- in the evolution equation (19). In the example in Figure 2, a robust estimation of the motion inside the initial contour would, on the contrary, produce a zero velocity and the contour would not evolve towards the bus.

The results in Figure 4 shows that similar convergence properties of our method can be observed if both the object

[1] 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.

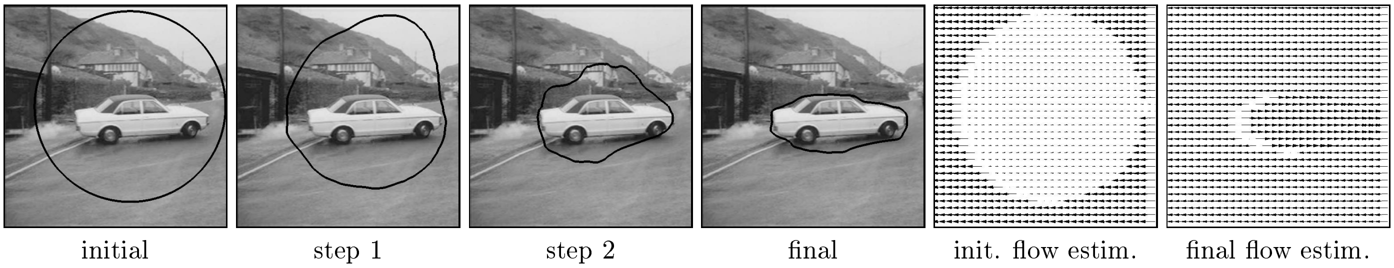


Fig. 4. Example from the Avengers sequence[2] showing a moving car captured by a moving camera. Despite large (about 4-5 pixels in many areas) and not purely translatory motion of both car and background, and despite the little grey value structure of car and street, the final segmentation is rather good. The discrepancy between car and street is partly due to the shadow moving with the car. During the contour evolution, the estimated flow fields are progressively separated, such that the final estimate (right) clearly shows the motion of the car.

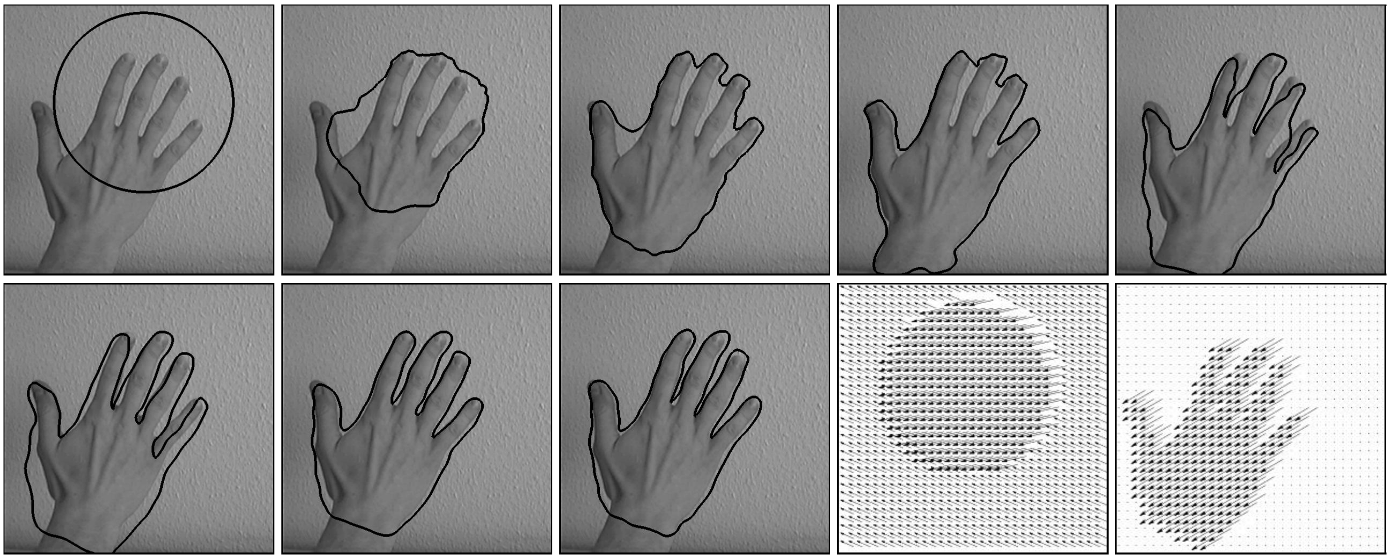


Fig. 5. Effect of the statistical shape prior: Hand moving to the bottom left. The statistical shape prior is introduced upon stationarity after the fourth frame. Initial and final estimates of the flow field show the improved separation of the two motion fields. The final segmentation is cut at the wrist, because the training shapes were all cut for simplicity.

and the background are moving. In this example from the "Avengers" sequence both the car and the background are moving.[2] The evolving contour and the initial and final estimated piecewise homogeneous motion fields show several properties of our method:

- The contour converges to the object of interest over a fairly large distance.
- The final segmentation is fairly good despite large (4-5 pixels in many areas) and non–translatory motion of both object and background, and despite the weak grey value structure of the car and the street.
- The initial contour does not need to be close to the true motion boundary for the evolution to converge.
- The moving object does not need to fill the dominant part of the region inside (or outside) the initial contour.
- The proposed variational method is fairly simple, it does not require presmoothing or elaborate robust estimation. The evolution from initial to final segmentation for the example in Figure 4 took less than 15 seconds on a 300 MHz SUN Ultra 10. Real–time implementations are therefore conceivable.

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

D. Effect of the Statistical Shape Prior

In cases of ambiguous motion information, e.g. due to missing or misleading information, the proposed motion segmentation may fail to converge to the correct result. If the object of interest is known one may introduce some prior knowledge into the segmentation approach.

In our example the object of interest is a moving hand. As explained in sections II-D and III-B, a statistical shape energy was derived from a set of 10 hand shapes, none of which corresponds to the hand in the image sequence.

We will demonstrate the effect of this shape prior on the motion segmentation process by introducing a shape energy in two different ways.

First we minimize the variational approach (4) without any shape prior ($\alpha = 0$) until stationarity — see Figure 5, fourth image. Then we determine the cyclical permutation of spline control points which — given the optimal similarity transformation — best aligns the present contour with the mean of the training contours. Finally we switch on the shape prior ($\alpha > 0$) and minimize the total energy (4) until convergence — see Figure 5, eighth image.

The shape prior improves segmentation in areas where

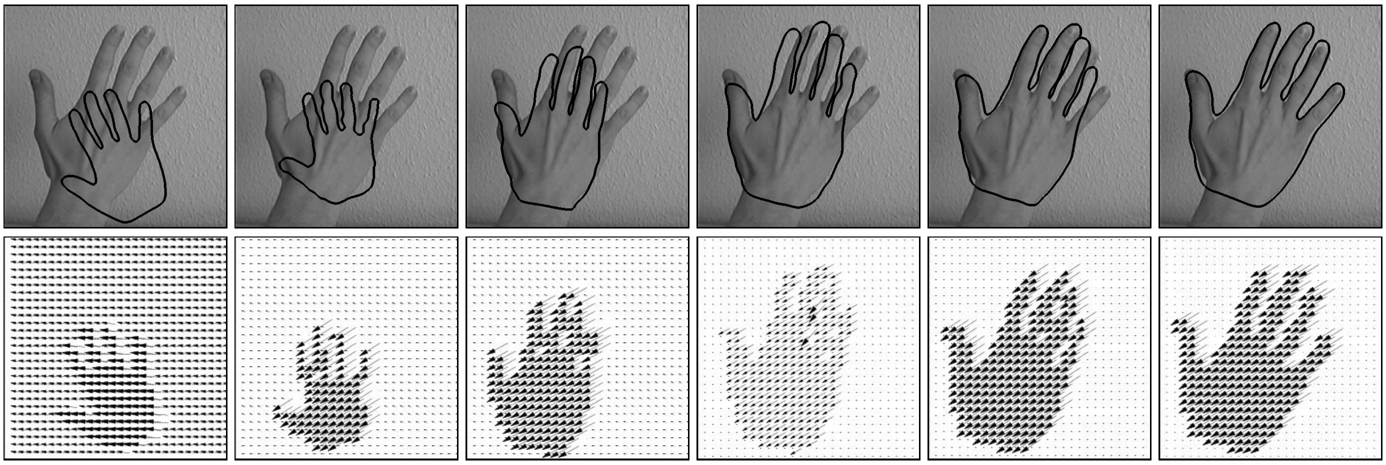


Fig. 6. Motion segmentation with statistical shape prior. During the contour evolution (top row, from left to right) the motion estimates (bottom) are progressively updated.

the motion information is not strong enough to drive the segmentation process — such as in the area between the fingers.

The estimated flow fields corresponding to the initial and the final contour show that the energy minimization separates the regions corresponding to different motion. During the contour evolution the corresponding motion estimation is gradually improved.

The evolution of the estimated flow field associated with the contour evolution is shown in a second example in Figure 6: for the same sequence of the moving hand, we included the statistical shape prior from the very beginning. The estimate of the object motion progressively improves during the contour evolution — see Figure 6, bottom row. Note also that due to the shape prior, the contour is strongly biased towards the subspace of familiar contours throughout the evolution process.

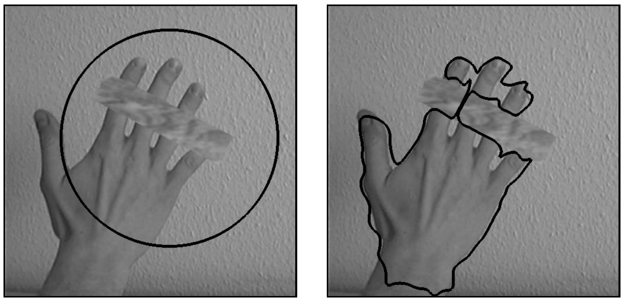
E. Dealing with Occlusion

In the above sequence of a moving hand the motion information by itself is not sufficiently strong to drive the contour to the desired segmentation. In a final example, we go one step further and artificially perturb the motion information by occluding part of the moving hand with a static structured object. Figure 7 shows the initial and the final contour obtained by minimizing the total energy (4) without any shape prior ($\alpha = 0$). Note that the contour separates moving from non–moving regions, given the constraint that no splitting of the contour is permitted.

Figure 8 shows a contour evolution obtained with a statistical shape prior on the same sequence of a moving hand occluded by a static bar. Due to the shape prior, the occlusion is ignored although it is not in accordance with the hand motion.

V. Current Limitations and Ongoing Work

Encouraged by the results presented above, we intend to tackle a number of limitations of the proposed approach for variational motion segmentation. First, the linearization in the optic flow constraint is only valid for small velocities



initial contour   final segmentation

Fig. 7. Motion segmentation without shape prior for a moving hand occluded by a static object. Note that the contour separates moving and non–moving regions. A contour splitting is not permitted.

w_i. If velocities are much larger than one pixel per frame, one needs to refer to multi–scale implementations of the flow estimation. Secondly, piecewise constant motion is assumed. Currently we pursue extensions to piecewise affine motion fields in order to segment objects which are rotating or changing in size upon motion orthogonal to the camera plane. Thirdly, the optic flow calculation only works for structured objects. Areas of constant grey value cannot be assigned to one or the other motion hypothesis. This aperture problem is well–known, the shape prior helps to overcome it. Although the shape prior based on the assumption of a Gaussian distribution works quite well in practice, its applicability to more complex shape variations and shapes of different classes is limited. Possible extensions are shape statistics based on mixtures of Gaussians [12], [5] or Gaussians in feature space [6].

VI. Conclusion

We presented a modification of the Mumford–Shah functional, which segments the image plane into domains of piecewise homogeneous motion. In the manner of region competition, a *single* functional is simultaneously minimized with respect to the motion vectors in these regions and the contour forming boundaries between the regions.

Numerical results show several properties of the pro-

Fig. 8. Motion segmentation with statistical shape prior for a moving hand occluded by a static object. Note that in this example it appears energetically favorable for the contour to decrease in size during the first iteration steps (2nd and 3rd image). Compared to the segmentation without shape prior in Figure 7, right side, the effect of the occlusion is compensated by the statistical prior.

posed approach: due to the region–based formulation the contour converges over fairly large distances in the gradient descent minimization. Due to the interlaced optimization of motion estimates and contour location, the estimates of object and background motion are progressively improved during the contour evolution.

The contour is implemented as a closed spline curve. Although this limits the class of permissible contours, it allows to incorporate a statistical prior on the shape of expected motion boundaries. The spline control point vectors of permissible shapes are assumed to form a Gaussian distribution. Mean and covariance matrix of this Gaussian are determined from a set of training images in a fully unsupervised way. The shape statistics are integrated in a combined variational approach in terms of a statistical shape energy.

Numerical results show that the shape prior improves the motion segmentation in several ways: it effectively reduces the dimension of the search space by restricting the contour motion to the subspace of learnt contours, it improves convergence in cases where the motion information is not sufficiently strong, and it permits reconstruction of occluded motion.

Acknowledgments

We thank P. Bouthemy, E. Mémin, T. Corpetti, C. Kervrann and the members of these groups for hospitality and many fruitful discussions, and J. Keuchel for comments on the manuscript.

References

[1] L. Ambrosio and V. M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ–convergence. *Comm. Pure Appl. Math.*, 43:999–1036, 1990.

[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] A. Blake and M. Isard. *Active Contours*. Springer, London, 1998.

[4] T. F. Cootes, C. J. Taylor, D. M. Cooper, and J. Graham. Active shape models – their training and application. *Comp. Vision Image Underst.*, 61(1):38–59, 1995.

[5] T.F. Cootes and C.J. Taylor. A mixture model for representing shape variation. *Image and Vis. Comp.*, 17(8):567–574, 1999.

[6] D. Cremers, T. Kohlberger, and C. Schnörr. Nonlinear shape statistics in Mumford–Shah based segmentation. In A. Heyden, P. Johansen, M. Nielsen, and G. Sparr, editors, *Proc. of the Europ. Conf. on Comp. Vis.*, LNCS, Copenhagen, May, 28–31 2002. Springer. To appear.

[7] D. Cremers, C. Schnörr, and J. Weickert. Diffusion–snakes: Combining statistical shape knowledge and image information in a variational framework. In *IEEE First Workshop on Variational and Level Set Methods*, pages 137–144, Vancouver, 2001.

[8] G. Farin. *Curves and Surfaces for Computer–Aided Geometric Design*. Academic Press, San Diego, 1997.

[9] G. Farnebäck. *Spatial Domain Methods for Orientation and Velocity Estimation*. PhD thesis, Dept. of Electrical Engineering, Linköpings universitet, 1999.

[10] G. Farnebäck. Very high accuracy velocity estimation using orientation tensors, parametric motion, and segmentation of the motion field. In *Proc. 8th Int. Conf. on Computer Vision*, volume 1, pages 171–177, 2001.

[11] C. Goodall. Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc., Ser. B.*, 53(2):285–339, 1991.

[12] T. Heap and D. Hogg. Improving specificity in pdms using a hierarchical approach. In *Brit. Mach. Vis. Conf.*, Colchester, UK, 1997.

[13] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. of Comp. Vision*, 1(4):321–331, 1988.

[14] C. Kervrann and F. Heitz. Statistical deformable model-based segmentation of image motion. *IEEE Trans. on Image Processing*, 8:583–588, 1999.

[15] P. Kornprobst, R. Dériche, and G. Aubert. Image sequence analysis via partial differential equations. *J. Math. Imag. Vision*, 11(1):5–26, 1999.

[16] E. Mémin and P. Pérez. Dense estimation and object–based segmentation of the optical flow with robust techniques. *IEEE Trans. on Image Processing*, 7(5):703–719, 1998.

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

[18] P. Nesi. Variational approach to optical flow estimation managing discontinuities. *Image and Vis. Comp.*, 11(7):419–439, 1993.

[19] 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.

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

[21] N. Paragios and R. Dériche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 22(3):266–280, 2000.

[22] C. Schnörr. Computation of discontinuous optical flow by domain decomposition and shape optimization. *Int. J. of Comp. Vision*, 8(2):153–165, 1992.

[23] C. Schnörr. Segmentation of visual motion by minimizing convex non-quadratic functionals. In *12th Int. Conf. on Pattern Recognition*, Jerusalem, Israel, Oct 9-13 1994.

[24] C. Schnörr and W. Peckar. Motion-based identification of deformable templates. In R. Šára V. Hlaváč, editor, *Proc. 6th Int. Conf. on Computer Analysis of Images and Patterns (CAIP '95)*, volume 970 of *LNCS*, pages 122–129, Prague, Czech Republic, Sept. 6-8 1995. Springer.

[25] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE–based computation of image motion. *Int. J. of Comp. Vision*, 45(3):245–264, 2001.

[26] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 18(9):884–900, 1996.