In Computer Vision - ECCV 2006, Springer LNCS 3951, A. Leonardis, H. Bischof, and A. Pinz (Eds.), pp. 471-483, Graz, Austria, May. 2006

Variational Motion Segmentation with Level Sets

Thomas Brox[1], Andrés Bruhn[2], and Joachim Weickert[2]

[1] CVPR Group, Department of Computer Science, University of Bonn
Römerstr. 164, 53113 Bonn, Germany
brox@cs.uni-bonn.de

[2] Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science,
Saarland University, Building 27, 66041 Saarbrücken, Germany
{bruhn,weickert}@mia.uni-saarland.de

Abstract. We suggest a variational method for the joint estimation of optic flow and the segmentation of the image into regions of similar motion. It makes use of the level set framework following the idea of motion competition, which is extended to non-parametric motion. Moreover, we automatically determine an appropriate initialization and the number of regions by means of recursive two-phase splits with higher order region models. The method is further extended to the spatiotemporal setting and the use of additional cues like the gray value or color for the segmentation. It need not fear a quantitative comparison to pure optic flow estimation techniques: For the popular Yosemite sequence with clouds we obtain the currently most accurate result. We further uncover a mistake in the ground truth. Coarsely correcting this, we get an average angular error below 1 degree.

1 Introduction

Motion estimation and segmentation are strongly related topics that can benefit from each other. While motion information gives important hints on how to partition an image, the separation of regions releases motion estimation from the problem of ambiguities near motion boundaries.

Both tasks have a long tradition in computer vision. In motion estimation, especially variational techniques based on modifications of the method of Horn and Schunck [15] have yielded very convincing results. Important milestones on the way to today's state-of-the-art have been presented in [21, 5, 18, 1, 7].

Also in the scope of segmentation, variational methods perform fairly well. Pioneering works in this field include [20, 16, 29, 17, 11, 12, 24]. In recent years, variational segmentation techniques have often been based on level sets [14, 22], which offer many advantages, among others the convenient implicit representation of regions and their separating contours. For the same reason, also the work presented in this paper will make use of the level set framework.

In most cases, segmentation relies on the image gray value or color, sometimes extended by texture representations. In case of image sequences, however, also motion information has been a popular cue for segmentation over the past decades, e.g. in [26, 28, 6]. Most motion segmentation techniques thereby handle the optic flow, or just the image difference, as a precomputed feature that is fed into a standard segmentation method.

In contrast to those methods, some more recent approaches embark on the strategy to solve the problems of optic flow estimation and segmentation simultaneously [27, 19, 13, 25, 2]. Cremers and Soatto introduced in [13] the level set based motion competition technique. The optic flow is estimated separately for each region by a parametric model, and the region contour is evolved directly by means of the fitting error of the optic flow. This idea has been adopted in [2], where the parametric model has been replaced by the better performing non-parametric optic flow model from [7].

Also the method proposed in the present paper follows the concept of motion competition where the fitting error of the optic flow drives the contour represented by level sets. Further on, the underlying optic flow model is also based on the technique from [7]. In this respect, our method is close to the approach in [2].

However, the method presented here is not restricted to two regions. The energy functional is inspired by the energy from [29] where also the number of regions is an unknown variable. Optimization of this energy is performed by means of a methodology suggested for texture segmentation in [8]: Starting with one region, regions are recursively split as long as this splitting decreases the total energy. For dealing with non-translational flow fields, we have to extend this idea by higher order region models. The recursive splitting yields the number of regions and appropriate initializations for the level set functions. These can then be evolved while the optic flow is simultaneously estimated within the regions. Moreover, our technique is not restricted to two frames but can also take more frames into account. In general, increasing the number of frames yields more accurate results.

The motion competition framework suffers from the fact that the optic flow is non-unique in those parts of the image with little or no structure. Although the smoothness term in variational techniques provides a dense flow field, it does not support the localization of the contour. Therefore, we also present a modification that allows the integration of additional cues like the gray value, color, or possibly even texture without the need to manually weight these different kinds of information.

Furthermore, we modify the underlying optic flow functional from [7]. Instead of matching only the gray value and gradient of a single pixel, we match a small Gaussian neighborhood around this pixel. It turns out that this matching of neighborhoods provides the variational model for the nonlinear version of the so-called CLG method from [10].

Apart from all these modelling aspects our paper also offers an experimental evaluation with excellent results. Thanks to the level set framework the precision at motion boundaries is so high that one can even notice a mistake in the ground truth of the popular Yosemite sequence. In fact, it turns out that the horizon is shifted one pixel towards the bottom. Correcting this, we obtain a further improvement of the results. However, even with the original ground truth, our method provides the most accurate flow fields in the literature.

Paper organization. The next section introduces the variational energy model that integrates motion estimation and multi-region segmentation. Section 3 then deals with the minimization of this energy. This includes the iterative scheme and the way how the level sets can be initialized. It is further described how the motion segmentation model can be extended by additional cues. Section 4 presents experimental results and a comparison to methods from the literature. The paper is concluded by a brief summary.

2 Model

The variational model is based on the optic flow functional in [7] and the segmentation model presented in [29]. It further makes use of the level set framework [14, 22, 12] in order to represent regions and their boundaries.

Given an image sequence $I(x, y, t) : \Omega \to \mathbb{R}$, we seek at each (spatiotemporal) point $\mathbf{x} := (x, y, t)$ the optic flow vector $\mathbf{w}(\mathbf{x}) := (u(\mathbf{x}), v(\mathbf{x}), 1)$ that describes the shift of the pixel at (x, y, t) to its new location $(x + u, y + v, t + 1)$ in the next frame. Additionally, we seek the set of level set functions $\Phi_i(x, y, t) : \Omega \to \mathbb{R}$, $i = 1, ..., N$, that represent the partitioning of the image domain Ω into disjoint regions Ω_i. The regions are represented such that $\mathbf{x} \in \Omega_i$ if and only if $\Phi_i(\mathbf{x}) > 0$, and region contours are represented by the zero-level lines of Φ_i. The number of regions N is also a free variable that is to be optimized.

In order to allow the motion estimation to benefit from the segmentation, we estimate a separate flow field $\mathbf{w}_i$ for each region. The final flow field $\mathbf{w}$ can then be assembled from $\mathbf{w}_i$ and the level set functions Φ_i.

Our model can be described by the spatiotemporal energy functional

$$\begin{aligned} E(\mathbf{w}, \mathbf{\Phi}, N) = & \\ & \sum_{i=1}^{N} \int_{\Omega} H(\Phi_i) \Big(\Psi\big(|\underbrace{I(\mathbf{x}+\mathbf{w}_i) - I(\mathbf{x})}_{\text{gray value constancy}}|^2 \big) + \gamma\, \Psi\big(|\underbrace{\nabla_2 I(\mathbf{x}+\mathbf{w}_i) - \nabla_2 I(\mathbf{x})}_{\text{gradient constancy}}|^2 \big) \Big)\, \mathbf{dx} \\ + & \sum_{i=1}^{N} \int_{\Omega} \Big(\alpha\, \Psi\big(\underbrace{|\nabla_3 u_i|^2 + |\nabla_3 v_i|^2}_{\text{spatiotemporal smoothness}} \big) + \nu\, \underbrace{|\nabla_3 H(\Phi_i)|}_{\text{contour length}} + \lambda \Big)\, \mathbf{dx} \end{aligned} \tag{1}$$

that is sought to be minimized under the constraint of disjoint regions. Thereby, ∇_3 denotes the spatiotemporal gradient $(\partial_x, \partial_y, \partial_t)^\top$, while ∇_2 stands for its spatial counterpart. Moreover, $H(s)$ denotes a regularized Heaviside function, which is in our case the error function. Its derivative $H'(\Phi)$ is a Gaussian with standard deviation 1.

The robust function $\Psi(s^2) := \sqrt{s^2 + 0.001^2}$ is applied in order to deal with outliers. In contrast to [7] we apply a separate robust function to both the gray value and the gradient constancy assumption, as suggested in [9]. This has the advantage that the relative importance of both terms is locally adjusted to the reliability of each term. The robust function applied to the smoothness term yields a model that allows for discontinuities. This is important, since although the level set framework captures the main motion discontinuies, there may be further smaller discontinuities within the regions. The parameter $\gamma \geq 0$ globally weights the influence of the gradient constancy assumption, whereas $\alpha \geq 0$ determines the penalty for non-smooth flow fields.

The energy in (1) follows the basic concept of motion competition [13, 2]. In comparison to the classic Chan-Vese model [12], the distance between the local value and the mean of the region is replaced by the local energy evoked by the data term of the optic flow model, i.e., it is tested how well the estimated optic flow fits the constancy assumptions. This energy drives the contour. Simultaneously, the model separates the estimation of the optic flow within the different regions. The parameter $\nu \geq 0$ weights the penalty for the length of the region contours.

In order to allow for more than two regions, the classic two-phase model is replaced by a level set based version of the segmentation functional from [29]. It handles not only

more than two regions, but also optimizes the number of regions. The fixed penalty $\lambda = 0.1$ is added for each region in order to keep the number of regions small. The additional optimization variable increases the complexity of the segmentation task and, consequently, makes it more dependent on the initialization. How to find a good initialization will be an issue in Section 3.

2.1 Adding Color and Neighborhood Constraints

To further improve the quality of the estimated optic flow, the model can be extended by making use of color. For this purpose, the term from (1) that is responsible for the gray value constancy

$$E_1(\mathbf{x}) = \Psi\left(|I(\mathbf{x}+\mathbf{w}_i(\mathbf{x})) - I(\mathbf{x})|^2\right) \tag{2}$$

is replaced by a term that supposes a multi-channel image $\mathbf{I} = (I_1, I_2, I_3)$:

$$E_2(\mathbf{x}) = \Psi\left(\sum_{k=1}^{3} |I_k(\mathbf{x}+\mathbf{w}_i(\mathbf{x})) - I_k(\mathbf{x})|^2\right). \tag{3}$$

The gradient constancy assumption in (1) is changed in the same way. This simple extension motivates another idea. Instead of assuming only the gray value and gradient of the point itself to stay constant, one can suppose also the neighborhood around this point not to change during motion. This is the standard assumption for block matching approaches. Since it is well known that block matching methods suffer seriously under affine transformations and run into problems at motion boundaries, we consider only a very small Gaussian neighborhood $K_\rho(x,y) = \frac{1}{2\pi\rho^2}\exp\left(-\frac{x^2+y^2}{2\rho^2}\right)$ with $\rho = \sqrt{2}$. Consequently, we obtain additional constraints in the new term

$$E_3(\mathbf{x}) = \Psi\left(\int_{\mathbb{R}^2} K_\rho(\xi)\left(|I(\mathbf{x}-\xi+\mathbf{w}_i(\mathbf{x})) - I(\mathbf{x}-\xi)|^2\right)\mathbf{d}\xi\right) \tag{4}$$

that can improve the robustness in the case of corrupted data. The gradient constancy assumption is extended in the same way. Minimization of the resulting energy functional in the next section will show that the latter extension comes down to the so-called *combined local-global (CLG)* method suggested in [10], which has been demonstrated to yield good results also in the presence of noise.

3 Optimization

In the optimization problem stated in (1), the optic flow field, the regions, and the number of regions are all unknown. Since variational approaches perform a local optimization, one has to take care of local optima. The initialization decides which optimum is hit by the method. For both optic flow estimation and segmentation techniques, coarse-to-fine strategies have proven their value in this respect. Coarse-to-fine strategies shift the problem of initialization to successively coarser scales. Starting with a scale where multiple optima are rare, one can use the coarse result as initialization for the next finer scale. In the iteration scheme described in Section 3.3 we also make use of this concept.

3.1 Initialization of the Regions at the Coarsest Scale

Before we minimize (1) by deriving the Euler-Lagrange equations, this section focuses on the initialization of the regions. Note that the number of regions N is an integer variable that cannot be optimized with a variational approach. For this purpose, we adopt a technique presented in the scope of texture segmentation in [8] that recursively splits the image domain for determining both N and good initializations for Φ_i. To this end, one needs a preliminary estimate of the optic flow, which is obtained by computing the optic flow without any partitioning.

The splitting works on the coarsest level, i.e., the flow is downsampled to this scale. The coarsest scale is chosen such that the image comprises at least 30 pixels in x and y-direction. At this scale, dominant regions are still visible and most disturbing structures have vanished. The scale of the temporal axis remains unchanged.

One starts with the whole image domain as one region, in which the level set function Φ is initialized by 8 horizontal stripes/boxes. The region is then split into two regions by minimizing the energy

$$E(\Phi) = \int_\Omega \big(- H(\Phi) \log p_1 - (1 - H(\Phi)) \log p_2 + \nu \, |\nabla_3 H(\Phi)| \big) \, \mathbf{dx} \tag{5}$$

which leads to the gradient descent

$$\partial_\tau \Phi = H'(\Phi) \left(\log \frac{p_1}{p_2} + \nu \, \mathrm{div} \left(\frac{\nabla_3 \Phi}{|\nabla_3 \Phi|} \right) \right). \tag{6}$$

At a first step, the two regions are modelled by the Gaussian probability densities

$$p_j(u,v) \propto \frac{1}{\sqrt{2\pi}(\sigma_u)_j} \exp\left(- \frac{(u - (\mu_u)_j)^2}{2(\sigma_u)_j^2} \right) \cdot \frac{1}{\sqrt{2\pi}(\sigma_v)_j} \exp\left(- \frac{(v - (\mu_v)_j)^2}{2(\sigma_v)_j^2} \right) \tag{7}$$

where $(\mu_{u/v})_j$ and $(\sigma_{u/v})_j$ are the means and the variances of the precomputed optic flow components u and v in the two regions. They are updated iteratively with the evolving contour. Since this model assumes constant flow fields in the regions, which is often unrealistic, the model is, after 500 iterations, extended to a linear approximation model $u(x,y,t) \approx a_j + b_j x + c_j y + d_j t$. Its parameters are estimated within the regions by least squares. With this model one can replace $(u - (\mu_u)_j)^2$ in (7) by $(u - a_j - b_j x - c_j y - d_j t)^2$ and $(\sigma_u)_j^2$ by $\int_{\Omega_j} (u - a_j - b_j x - c_j y - d_j t)^2 \, \mathbf{dx} / |\Omega_j|$. The values for v can be handled the same way. After again 500 iterations, we finally switch to a quadratic model, which can coarsely capture most smooth motion fields. The model parameters are again obtained by least squares, and the counterparts to the expressions in (7) are the deviation $(u - a_j - b_j x - c_j y - d_j t - e_j xx - f_j yy - g_j tt - h_j xy - r_j xt - s_j yt)^2$, the corresponding standard deviation, and the same for v. The successive increase of the model complexity avoids possible local optima that might disturb the partitioning when starting directly with the quadratic model.

The quadratic model is finally used to measure the energy according to (5), both in the original region and the separated regions. If the energy decrease is larger than $\lambda |\Omega|$, the region is split and the same process is repeated for the two new regions. When all further splits do not decrease the energy anymore, one has determined N. Moreover, a good initialization for Φ_i is available.

3.2 Euler-Lagrange Equations

With the abbreviations from [7]

$$\begin{array}{ll} (I_x)_i := \partial_x I(\mathbf{x}+\mathbf{w}_i), & (I_{xy})_i := \partial_{xy} I(\mathbf{x}+\mathbf{w}_i), \\ (I_y)_i := \partial_y I(\mathbf{x}+\mathbf{w}_i), & (I_{yy})_i := \partial_{yy} I(\mathbf{x}+\mathbf{w}_i), \\ (I_z)_i := I(\mathbf{x}+\mathbf{w}_i) - I(\mathbf{x}), & (I_{xz})_i := \partial_x I(\mathbf{x}+\mathbf{w}_i) - \partial_x I(\mathbf{x}), \\ (I_{xx})_i := \partial_{xx} I(\mathbf{x}+\mathbf{w}_i), & (I_{yz})_i := \partial_y I(\mathbf{x}+\mathbf{w}_i) - \partial_y I(\mathbf{x}) \end{array} \tag{8}$$

one can derive the following Euler-Lagrange equations from (1):

$$\begin{aligned} & H(\Phi_i)\Big(\Psi'\left((I_z)_i^2\right)\right)(I_x)_i(I_z)_i + \gamma\,\Psi'\left((I_{xz})_i^2 + (I_{yz})_i^2\right)\left((I_{xx})_i(I_{xz})_i + (I_{xy})_i(I_{yz})_i\right)\Big) \\ & \quad -\alpha\,\mathrm{div}\left(\Psi'\left(|\nabla_3 u_i|^2 + |\nabla_3 v_i|^2\right)\nabla_3 u_i\right) = 0, \\ & H(\Phi_i)\Big(\Psi'\left((I_z)_i^2\right)\right)(I_y)_i(I_z)_i + \gamma\,\Psi'\left((I_{xz})_i^2 + (I_{yz})_i^2\right)\left((I_{yy})_i(I_{yz})_i + (I_{xy})_i(I_{xz})_i\right)\Big) \\ & \quad -\alpha\,\mathrm{div}\left(\Psi'\left(|\nabla_3 u_i|^2 + |\nabla_3 v_i|^2\right)\nabla_3 v_i\right) = 0, \\ & H'(\Phi_i)\Big(-\Psi((I_z)_i^2) - \gamma\,\Psi\left((I_{xz})_i^2 + (I_{yz})_i^2\right) + \nu\,\mathrm{div}\left(\tfrac{\nabla_3\Phi_i}{|\nabla_3\Phi_i|}\right)\Big) = 0. \end{aligned} \tag{9}$$

Obviously, the flow estimates $\mathbf{w}_i$ are only influenced by the image data in areas where $H(\Phi_i) > 0$, i.e., $\Phi_i > 0$. Thus they cannot be disturbed by data outside the region.
The contour evolution is driven by the fitting energy of the optic flow. Note that the corresponding Euler-Lagrange equation does not respect the additional constraint of disjoint regions yet. This has to be ensured in the gradient descent equation by establishing a competition between neighboring regions [8]:

$$\begin{aligned} \partial_\tau \Phi_i &= H'(\Phi_i)\left(e_i - \max_{\substack{H'(\Phi_j)>0.3 \\ j\neq i}} (e_j, e_i - 1)\right), \\ e_k &:= -\Psi((I_z)_k^2) - \gamma\,\Psi\left((I_{xz})_k^2 + (I_{yz})_k^2\right) + \nu\,\mathrm{div}\left(\tfrac{\nabla_3\Phi_k}{|\nabla_3\Phi_k|}\right). \end{aligned} \tag{10}$$

Here, each region competes with the best performing neighboring region. This ensures that each pixel is part of exactly one region: the one where it fits best.
The extensions to color images and the matching of neighborhoods lead to simple adaptations in (9). Using (4) instead of (2) leads to replacing $(I_x)_i^2$ by $K_\rho * (I_x)_i^2$, $(I_y)_i^2$ by $K_\rho * (I_y)_i^2$, $(I_x)_i(I_y)_i$ by $K_\rho * ((I_x)_i(I_y)_i)$, and so on, where $K_\rho * (\cdot)$ denotes a convolution with K_ρ. One realizes that the resulting Euler-Lagrange equations coincide with those from the CLG method in [10]. The same way, one obtains the color case by replacing $(I_x)_i^2$ by $\sum_{k=1}^3((I_k)_x)_i)^2$ and so on.

3.3 Iteration scheme

The iteration scheme for solving for $\mathbf{w}_i$ and Φ_i is similar to [7, 2] and consists of three nested iteration loops. Starting with the level set functions and the preliminary optic flow from Section 3.1 at the coarsest scale, the most outer iteration loop transfers the current flow estimates $\mathbf{w}_i$ and the level set functions Φ_i to the next finer scale before warping the second frame according to $\mathbf{w}_i$ towards the first one. Thus in each iteration,

only an update $(du, dv)_i$ on $\mathbf{w}_i$ has to be computed; see [7]. The scaling factor between two successive levels is $\eta = 0.95$ like in [7]. Depending on the size of the image, it determines the number of outer iterations. The parameter ν is scaled at each level by $0.0002 \cdot A^{0.7}$ where A denotes the size of the image at the respective level. This scaling has been determined empirically in many segmentation experiments not restricted to the motion segmentation examples in this paper.
The central iteration loop is a fixed point iteration loop that removes the remaining nonlinearity in the optic flow equations. Furthermore it alternates the solution of the resulting linear equations and the update on Φ_i. We perform 10 of these central iterations, as suggested in [7].
There are two inner iteration loops, one for solving the linear equations on $(du, dv)_i$ via SOR with over-relaxation parameter $\omega = 1.95$, and one for evolving the level sets via (10). We perform 15 SOR iterations and 50 iterations on the level sets. The curvature term in (10) is implemented by mean curvature motion restricted to the narrow band given by $H'(\Phi_i) > 0.3$.

3.4 Integrating Additional Cues for Segmentation

The location of the contour can only be determined reliably by the data fitting error of the optic flow, if there is distinctive data available. In areas with little structure, the optic flow is not uniquely determined and hence cannot drive the contour. For such cases it is helpful to integrate cues besides motion for the contour evolution. This can be achieved by adding the term [29, 24, 8]

$$E_{\text{Image}} = -\sum_{i=1}^{N} \int_{\Omega} H(\Phi_i) \log p_i \, \mathbf{dx} \tag{11}$$

to (1). It includes a statistical region model by means of the probability densities p_i. For making use of a color image $\mathbf{I} = (I_1, I_2, I_3)$, we model p_i as

$$p_i(\mathbf{I}) \propto \prod_{k=1}^{3} \frac{1}{\sqrt{2\pi}\sigma_{ik}} \exp\left(-\frac{(I_k - \mu_{ik})^2}{2\sigma_{ik}^2}\right) \tag{12}$$

with the local means μ_{ik} and standard deviations σ_{ik} of channel k in region i.
Unfortunately, the contributions of the color channels and the optic flow can be different by some orders of magnitude and depend severely on the choice of the parameters in the optic flow model. A further weighting parameter seems necessary to balance the contributions. However, a manual choice of this parameter can be avoided.
One can verify that using a Gaussian model including the standard deviations like in (12) makes the model independent from a different scaling of each feature channel. The same independence from scaling can also be achieved for the motion channel by normalizing the fitting error in (10) by the average error in the whole image domain. Together with the contribution due to (11) one obtains:

$$e_k := -\frac{\Psi((I_z)_k^2) + \gamma \Psi\left((I_{xz})_k^2 + (I_{yz})_k^2\right)}{\frac{1}{|\Omega|}\int_{\Omega}\left(\Psi(I_z^2) + \gamma \Psi\left(I_{xz}^2 + I_{yz}^2\right)\right)\mathbf{dx}} + \log p_k + \nu \operatorname{div}\left(\frac{\nabla_3 \Phi_k}{|\nabla_3 \Phi_k|}\right). \tag{13}$$

The normalization factor, like the standard deviation in (12), can be regarded as an adaptive weight for the motion term that avoids a manual choice.

Technique	AAE $\pm$ STD
Alvarez et al. [1]	$5.53^\circ \pm 7.40^\circ$
Mémin–Pérez [18]	$4.69^\circ \pm 6.89^\circ$
Bruhn et al. [10]	$4.17^\circ \pm 7.72^\circ$
Papenberg et al. (frames 8,9) [23]	$2.44^\circ \pm 6.90^\circ$
Amiaz–Kiryati [2]	$2.04^\circ \pm 7.83^\circ$
Papenberg et al. (all frames) [23]	$1.78^\circ \pm 7.00^\circ$
Amiaz–Kiryati [3]	$1.73^\circ \pm 5.85^\circ$
Bruhn–Weickert [9]	$1.72^\circ \pm 6.88^\circ$
Our method (frames 8,9)	$1.67^\circ \pm 6.30^\circ$
Our method (frames 7,8,9)	$1.39^\circ \pm 6.32^\circ$
Our method (all frames)	$1.22^\circ \pm 6.37^\circ$

Table 1. Comparison between our results and those from the literature with 100% density for the Yosemite sequence with cloudy sky. AAE = average angular error. STD = standard deviation.

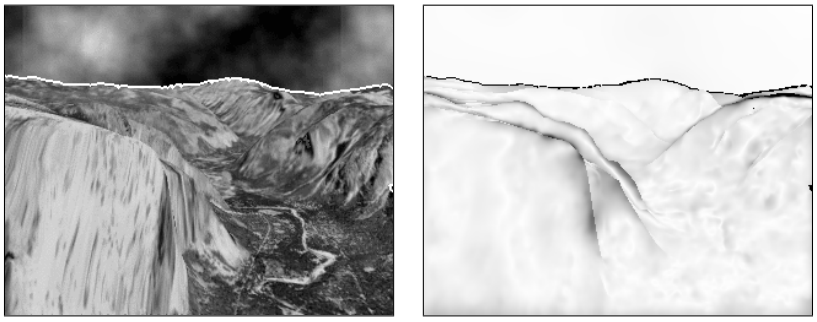

Fig. 1. Left: Frame 8 from the Yosemite sequence together with the estimated contour. **Right:** Angular error between the estimated flow field and the ground truth.

4 Experiments

4.1 Quantitative Evaluation

The Yosemite sequence with clouds, created by Lynn Quam, is currently still the most interesting test sequence for comparing motion estimation techniques, as it contains many typical challenges: One large discontinuity at the horizon and some smaller ones in the canyon, divergent and translational motion, relatively large displacements in the lower left, brightness changes in the sky, and small occlusions at the image boundaries. Furthermore, the ground truth and many published results from other methods are available.

Table 1 shows a comparison of our results to those from the literature using the angular error measure introduced in [4]. Obviously, modeling the sky and the canyon by separate regions yields a significant improvement in comparison to the method in [7, 23]. Also the changes in comparison to the method in [3] still have a large impact, especially the spatiotemporal motion estimation in combination with 3-D level sets. Already taking one additional frame into account improves the result considerably. The parameters

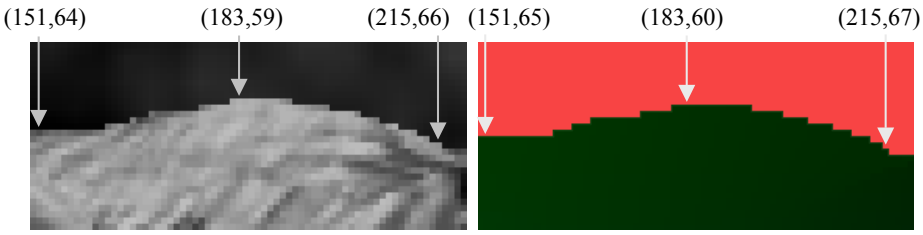

Fig. 2. Discrepancy between frame 8 and the ground truth. In the ground truth, the horizon is consistently 1 pixel too low, which cannot be explained by interpolation artifacts.

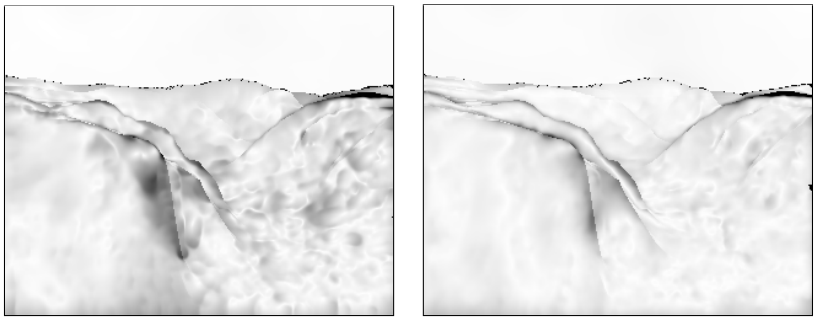

Fig. 3. Angular error with the corrected ground truth for the spatial (left) and the spatiotemporal method (right).

have been set to $\alpha = 5500$, $\gamma = 550$, and $\nu = 0.15$ for the variant with two frames, and $\alpha = 4000$, $\gamma = 550$, and $\nu = 0.15$ for the spatiotemporal version. The images have been presmoothed with a Gaussian kernel of size $\sigma = 1$ like in [7].

Fig. 1 depicts the resulting contour and the remaining error between the estimated flow and the ground truth. Only few areas with larger errors persist. One of these is still the horizon, which is actually very well estimated by the partitioning. The width of the error is almost exactly one pixel along the horizon, which is a good motivation to take a closer look at the ground truth.

Indeed it turns out that the ground truth delivered with the sequence is erroneous, as demonstrated in Fig. 2. Among other mistakes, the horizon is consistently one pixel too low. While this has not been decisive as long as the techniques had large errors at the horizon, it becomes quite important as soon as one is able to correctly estimate the flow there. We coarsely corrected the mistake by shifting the first 75 lines of the ground truth one pixel towards the top. We then obtained an average angular error of $1.40^\circ \pm 3.69^\circ$ for the method with two frames and $0.92^\circ \pm 3.35^\circ$ for the spatiotemporal version. In particular, the improvement of both statistical measures – the mean error and the standard deviation – confirms that our observation of an erroneous ground truth was right. This is also reflected in the corresponding error plots in Fig. 3. They show no more than a few misclassified pixels at the horizon. In fact, our results for the Yosemite sequence with clouds are so accurate that they even outperform the results obtained by the method in [7] for the much less challenging variant *without* cloudy sky.

The shifted horizon is the most significant mistake in the so-called ground truth, but unfortunately not the only one. The reader is invited to compare a zoomed version of frame 8 and the ground truth in order to realize further discrepancies in the canyon region that cannot be corrected so easily. In the near future, it may hence be necessary to have a good successor of this sequence.

4.2 Motion Segmentation with Multiple Objects

While the Yosemite sequence allows for comparing the quality of the estimated optic flow to that of other methods, it cannot demonstrate one of the main novelties of this paper, namely the possibility to deal with more than two regions. Therefore, we show in Fig. 4 a test scenario with two objects moving in different directions and the camera moving towards them. This yields a divergent flow field for the background and two nearly translational motion fields for the two objects.
In this experiment, we also integrated the CIELAB color channels into the segmentation. They can help to improve the result in areas with little structure where the optic flow is not well-determined. Due to the implicit weighting, it was not necessary to put explicit weights to the optic flow term and the color channels. The free parameters were set to $\alpha = 550$, $\gamma = 50$, and $\nu = 0.5$, and the images were presmoothed with $\sigma = 1$ as above.
Fig. 4 shows the segmentation result and the estimated flow field where the hue represents the direction and the intensity the length of the flow vectors. With the same parameter $\lambda = 0.1$ as for the Yosemite sequence, three regions have been detected by the initialization part. The final contours are very precise and the estimated optic flow is not disturbed by the occlusions that appear near the object boundaries. Note the hull of the boat having basically no gradients in motion direction. Here, integrating color information helps significantly to determine the object boundary.

5 Summary

We have presented an approach for joint motion estimation and segmentation with a non-parametric motion model. It is capable to automatically detect and deal with an arbitrary number of regions, and it can take more than two frames into account. Moreover, it has been shown that it is possible to make use of further cues besides the motion information. To the best of our knowledge, our experiments yielded the currently most accurate results in the literature. The accuracy of the motion boundaries is so high that it is even possible to spot a mistake of a one-pixel shift in the ground truth of the Yosemite sequence. Obviously, today's motion estimation techniques are partially more accurate than certain presumably correct flow fields. To support further research, our next effort hence will be to provide a new, possibly also more challenging, synthetic test sequence with ground truth.

Acknowledgements

We thank Luis Garrido (University Pompeu Fabral, Barcelona, Spain) for raising the question if there is an error in the ground truth of the Yosemite sequence.

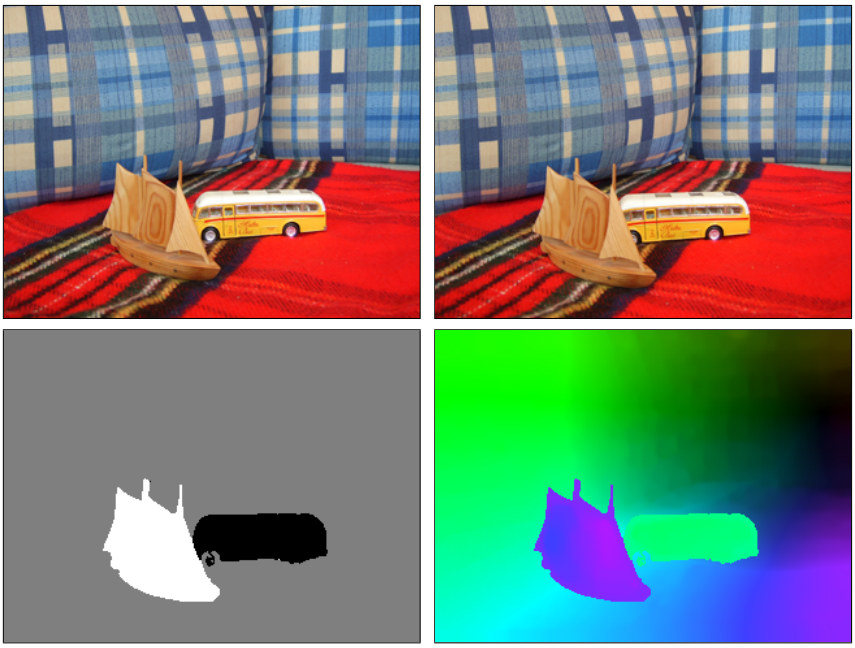

Fig. 4. Top row: Two input images with two moving objects. Both objects move straight forward. The camera moves into the scene. **Bottom left:** Segmentation result. **Bottom right:** Estimated motion. The color distinguishes the direction of the flow vector, the intensity its magnitude.

References

1. L. Alvarez, J. Weickert, and J. Sánchez. Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39(1):41–56, Aug. 2000.
2. T. Amiaz and N. Kiryati. Dense discontinuous optical flow via contour-based segmentation. In *Proc. International Conference on Image Processing*, volume 3, pages 1264–1267, Genoa, Italy, Sept. 2005.
3. T. Amiaz and N. Kiryati. Piecewise-smooth dense optical flow via level sets. Technical Report VIA-2005-6-2, Vision and Image Analysis Laboratory, School of Electrical Engineering, Tel Aviv University, Israel, June 2005.
4. J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, Feb. 1994.
5. M. J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, Jan. 1996.
6. P. Bouthemy and E. François. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *International Journal of Computer Vision*, 10(2):157–182, 1993.
7. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision - Proc. 8th European Conference on Computer Vision*, LNCS 3024, pages 25–36. Springer, May 2004.
8. T. Brox and J. Weickert. Level set based segmentation of multiple objects. In C. Rasmussen, H. Bülthoff, M. Giese, and B. Schölkopf, editors, *Pattern Recognition*, LNCS 3175, pages 415–423. Springer, Aug. 2004.

9. A. Bruhn and J. Weickert. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *Proc. 10th International Conference on Computer Vision*. IEEE Computer Society Press, Beijing, China, pages 749–755, Oct. 2005.
10. A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *Int. Journal of Computer Vision*, 61(3):211–231, 2005.
11. V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22:61–79, 1997.
12. T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb. 2001.
13. D. Cremers and S. Soatto. Motion competition: A variational framework for piecewise parametric motion segmentation. *Internatonal Journal of Computer Vision*, 62(3):249–265, 2005.
14. A. Dervieux and F. Thomasset. A finite element method for the simulation of Rayleigh–Taylor instability. In R. Rautman, editor, *Approximation Methods for Navier–Stokes Problems*, volume 771 of *Lecture Notes in Mathematics*, pages 145–158. Springer, Berlin, 1979.
15. B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
16. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1988.
17. S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Conformal curvature flows: from phase transitions to active vision. *Archive for Rational Mechanics and Analysis*, 134:275–301, 1996.
18. E. Mémin and P. Pérez. A multigrid approach for hierarchical motion estimation. In *Proc. 6th International Conference on Computer Vision*, pages 933–938, Bombay, India, 1998.
19. E. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. *International Journal of Computer Vision*, 46(2):129–155, 2002.
20. D. Mumford and J. Shah. Boundary detection by minimizing functionals, I. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 22–26, San Francisco, CA, June 1985. IEEE Computer Society Press.
21. H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.
22. S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.
23. N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. Technical Report 124, Dept. of Mathematics, Saarland University, Saarbrücken, Germany, Jan. 2005. To appear in *International Journal of Computer Vision*.
24. N. Paragios and R. Deriche. Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13(1/2):249–268, 2002.
25. N. Paragios and R. Deriche. Geodesic active regions and level set methods for motion estimation and tracking. *Computer Vision and Image Understanding*, 97(3):259–282, 2005.
26. J. L. Potter. Velocity as a cue to segmentation. *IEEE Transactions on Systems, Man and Cybernetics*, 5:390–394, 1975.
27. C. Schnörr. Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class. *International Journal of Computer Vision*, 6(1):25–38, Apr. 1991.
28. W. B. Thompson. Combining motion and contrast for segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(6):543–549, 1980.
29. S.-C. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, Sept. 1996.