

A variational approach to joint denoising, edge detection and motion estimation

Alexandru Telea¹, Tobias Preusser², Christoph Garbe³, Marc Droske⁴, and Martin Rumpf⁵

¹ Eindhoven University of Technology, alex@win.tue.nl

² CeVis, University of Bremen, tp@mevis.de

³ IWR, University of Heidelberg, Christoph.Garbe@iwr.uni-heidelberg.de

⁴ UCLA, Los Angeles, Marc.Droske@math.ucla.edu

⁵ INS, University of Bonn, martin.rumpf@ins.uni-bonn.de

Abstract. The estimation of optical flow fields from image sequences is incorporated in a Mumford–Shah approach for image denoising and edge detection. Possibly noisy image sequences are considered as input and a piecewise smooth image intensity, a piecewise smooth motion field, and a joint discontinuity set are obtained as minimizers of the functional. The method simultaneously detects image edges and motion field discontinuities in a rigorous and robust way. It comes along with a natural multi-scale approximation that is closely related to the phase field approximation for edge detection by Ambrosio and Tortorelli. We present an implementation for 2D image sequences with finite elements in space and time. It leads to three linear systems of equations, which have to be iteratively in the minimization procedure. Numerical results underline the robustness of the presented approach and different applications are shown.

1 Introduction

The task of motion estimation is a fundamental problem in computer vision. In low-level image processing, the accurate computation of object motion in scenes is a long standing problem, which has been addressed extensively. In particular, global variational approaches initiated by the work of Horn and Schunck [1] are increasingly popular. Initial problems such as the smoothing over discontinuities or the high computational cost have been resolved successfully [2,3,4]. Motion also poses an important cue for object detection and recognition. While a number of techniques first estimate the motion field and segment objects later in a second phase [5], an approach of both computing motion as well as segmenting objects at the same time is much more appealing. First advances in this direction were investigated in [6,7,8,9,10,11].

The idea of combining different image processing tasks into a single model in order to cope with interdependencies has drawn attention in several different fields. In image registration, for instance, a joint discontinuity approach for simultaneous registration, segmentation and image restoration has been proposed

by Droske & Ring [12] and extended in [13] incorporating phase field approximations. Yezzi, Zöllei and Kapur [14] and Unal *et al.* [15] have combined segmentation and registration applying geodesic active contours described by level sets in both images. Vemuri *et al.* have also used a level set technique to exploit a reference segmentation in an atlas [16]. We refer to [17] for further references.

Cremers and Soatto [18,19] presented an approach for joint motion estimation and motion segmentation with one functional. Incorporating results from Bayesian inference, they derived an energy functional, which can be seen as an extension to the well-known Mumford–Shah [20] approach. Their functional involves the length of boundaries separating regions of different motion as well as a “fidelity-term” for the optical-flow assumption. Our approach is in particular motivated by their investigations, resolving the drawback of detecting edges in a parametric model, by a non-parametric approach.

Recently, highly accurate motion estimation [21] has been extended to contour-based segmentation [22] following a well known segmentation scheme [23]. The authors demonstrate that extending the motion estimator to edge detection in a variational framework leads to an increase in accuracy. However, as opposed to our framework, the authors do not include image denoising in their framework. Including a denoising functional together with motion estimation in a variational framework has been achieved by [24]. They report significant increases the accuracy of motion estimation, particularly with respect to noisy image sequences. However, edges are not detected, but errors of smoothing over discontinuities are lessened by formulating the smoothness constraint in a L_1 metric.

We present the first approach of combining motion estimation, image denoising and edge detection in the same variational framework. This step will allow us to produce more accurate estimations of motion while detecting edges at the same time and preventing any smoothing across them.

The combination of denoising and edge detection with the estimation of motion results in an energy functional incorporating fidelity- and smoothness-terms for both the image and the flow field. Moreover, we incorporate an anisotropic enhancement of the flow along the edges of the image in the sense of Nagel and Enkelmann [2]. The model is implemented using the phase-field approximation in the spirit of Ambrosio’s and Tortorelli’s [25] approach for the original Mumford–Shah functional. The identification of edges is phrased in terms of a phase field function, no *a-priori* knowledge of objects is required, as opposed to formulations of explicit contours. In contrast to a level set approach, the built-in multi-scale enables a robust and efficient computation and no initial guess for the edge set is required. We present here a truly $d + 1$ dimensional algorithm, considering time as an additional dimension to the d -dimensional image data. This fully demonstrates the conceptual advantages of the joint approach. The characteristics of our approach are:

- The distinction of smooth motion fields and optical flow discontinuities is directly linked to edge detection, improving the reliability of estimates.
- Image denoising and segmentation profits from explicit coupling of the sequence via the brightness constancy assumption.

- The phase field approximation converges to a limit problem for a vanishing scale parameter, with a representation of edges and motion discontinuities without any additional filtering.
- The algorithm is an iterative approach. In each step, a set of three simple linear systems are solved, requiring only a small number of iterations.

2 Generalized optical flow equation

In image sequences, we observe different types of motion fields: locally smooth motion visible via variations of object shading and texture in time, or jumps in the motion velocity apparent at edges of objects moving in front of a background. We aim for an identification of corresponding piecewise smooth optical flow fields in piecewise smooth image sequences

$$u : [0, T] \times \Omega \mapsto \mathbb{R}; \quad (t, x) \mapsto u(t, x)$$

for a finite time interval $[0, T]$ and a spatial domain $\Omega \subset \mathbb{R}^d$ with $d = 1, 2, 3$. The flow fields are allowed to jump on edges in the image sequence. Hence, the derivative Du splits into a singular and a regular part. The regular part is a classical gradient $\nabla_{(t,x)}u$ in space and time, whereas the singular part lives on the singularity set S - the set of edge surfaces in space-time. Time slices of S are the actual image edges S with respect to space-time. The singular part represents the jump of the image intensity on S , i. e., one observes that $D^s u = (u^+ - u^-)n_s$. Here, u^+ and u^- are the upper and lower intensity values on both sides of S , respectively. We now suppose that the image sequence u reflects an underlying motion with a piecewise smooth motion velocity v , which is allowed to jump only on S . Thus, S represents object boundaries moving in front of a possibly moving background. In this general setting, without any smoothness assumption on u and v , we ask for a generalized optical flow equation. Apart from moving object edges, we derive from the brightness constancy assumption $u(t+s, x+sv) = \text{const}$ on motion trajectories $\{(t+s, x+sv) : s \in [0, T]\}$, that

$$\nabla_{(t,x)}u \cdot w = 0 \quad \text{and on the edge} \quad n_s \cdot (w^+ + w^-) = 0. \quad (1)$$

where $w = (1, v)$ is the space-time motion velocity. This in particular includes the case of a sliding motion without any modification of the object overlap, where $n_s \cdot w^+ = n_s \cdot w^- = 0$.

3 Mumford–Shah approach to optical flow

In their pioneering paper, Mumford and Shah [20] proposed the minimization of the following energy functional:

$$E_{MS}[u, S] = \lambda \int_{\Omega} (u - u_0)^2 d\mathcal{L} + \frac{\mu}{2} \int_{\Omega \setminus S} \|\nabla u\|^2 d\mathcal{L} + \nu \mathcal{H}^{d-1}(S), \quad (2)$$

where u_0 is the initial image defined on an image domain $\Omega \subset \mathbb{R}^d$ and λ, μ, ν are positive weights. Here, one asks for a piecewise smooth representation u of u_0 and an edge set S , such that u approximates u_0 in the least-squares sense. u should be smooth apart from the free discontinuity S . In addition, S should be smooth and thus small with respect to the $(d-1)$ -dimensional Hausdorff-measure \mathcal{H}^{d-1} . Mathematically, this problem has been treated in the space of functions of bounded variation BV , more precisely in the specific subset SBV [26]. In this paper, we will pick up a phase field approximation for the Mumford–Shah functional (2) proposed by Ambrosio and Tortorelli [25]. They describe the edge set S by a phase field ϕ which is supposed to be small on S and close to 1 apart from edges, i. e., one asks for minimizers of the energy functional

$$E_\epsilon[u, \phi] = \int_{\Omega} \lambda(u - u_0)^2 + \frac{\mu}{2}(\phi^2 + k_\epsilon)\|\nabla u\|^2 + \nu\epsilon\|\nabla\phi\|^2 + \frac{\nu}{4\epsilon}(1 - \phi)^2 \, d\mathcal{L}, \quad (3)$$

where ϵ is a scale parameter and $k_\epsilon = o(\epsilon)$ a small positive regularizing parameter, which mathematically ensures strict coercivity with respect to u . Hence, the second term measures smoothness of u but only apart from edges. On edges, the weight ϕ^2 is expected to vanish. The last two terms in the integral encode the approximation of the $d - 1$ dimensional edge set area and strongly favours a phase field value 1 away from edges, respectively. For larger ϵ , one obtains coarse, blurred representations of the edge set and corresponding smoother images u . With decreasing ϵ we successively refine the representation of the edges and include more image details.

Now, we ask for a simultaneous denoising, segmentation and flow extraction on image sequences. Hence, we will incorporate the motion field generating an image sequence into a variational method. We first formulate a corresponding minimization problem in the spirit of the Mumford–Shah model:

Mumford–Shah type optical flow approach. *Given a noisy initial image sequence $u_0 : D \mapsto \mathbb{R}$ on the space time domain $D = [0, T] \times \Omega$, we ask for a piecewise smooth image sequence u , which jumps on a set S , and a piecewise smooth motion field $w = (1, v)$, which is allowed to jump on the same set S , with the constraint $n_s \cdot (w^+ + w^-) = 0$, such that (u, w, S) minimize the energy*

$$\begin{aligned} E[u, w, S] = & \int_D \frac{\lambda_u}{2}(u - u_0)^2 + \frac{\lambda_w}{2}(w \cdot \nabla_{(t,x)} u)^2 \, d\mathcal{L} + \int_D \frac{\mu_u}{2}(\phi^2 + k_\epsilon)\|\nabla_{(t,x)} u\|^2 \, d\mathcal{L} \\ & + \int_D \frac{\mu_w}{2}\|P[\phi]\nabla_{(t,x)} w\|^q \, d\mathcal{L} + \int_D \left(\nu\epsilon\|\nabla\phi\|^2 + \frac{\nu}{4\epsilon}(1 - \phi)^2 \right) \, d\mathcal{L}. \quad (4) \end{aligned}$$

The first and second term of the energy are fidelity terms with respect to the image intensity and the regular part of the optical–flow–constraint, respectively. The third and fourth term encode the smoothness requirement of u and w . Finally, the last terms represents the area of the edge surfaces S , parameterized by the phase file π . The projection operator $P[\phi]$ couples the smoothness of the

motion field w to the image geometry:

$$P[\phi] = \alpha(\phi^2) \left(\mathbb{I} - \beta(\phi^2) \frac{\nabla_{(t,x)}\phi}{\|\nabla_{(t,x)}\phi\|} \otimes \frac{\nabla_{(t,x)}\phi}{\|\nabla_{(t,x)}\phi\|} \right).$$

Here, $k_\epsilon = o(\epsilon)$ is a "safety" coefficient, which will ensure existence of solutions of our approximate problem. $\alpha : \mathbb{R} \rightarrow \mathbb{R}_0^+$ and $\beta : \mathbb{R} \rightarrow \mathbb{R}_0^+$ are continuous blending functions. For vanishing ϵ and a corresponding steepening of the slope of u , this operator basically leads to a 'one sided diffusion' in the energy relaxation. The fidelity weights λ_u, λ_w , the regularity weights μ_u, μ_w and the weight ν controlling the phase field are supposed to be positive and $q \geq 2$. We emphasize that, without any guidance from the local time-modulation of shading or texture on both sides of an edge, there is still a undecidable ambiguity with respect to foreground and background.

4 Variations of the energy and an algorithm

In what follows, we will consider the Euler–Lagrange equations of the above energies. Thus, we need to compute the variations of the energy contributions with respect to the involved unknowns u, w, ϕ . Using straightforward differentiation for sufficiently smooth u, w, ϕ and initial data u_0 and summing up the resulting terms, we can integrate by parts and end up with the following system of PDEs

$$-\operatorname{div}_{(t,x)} \left(\frac{\mu_u}{\lambda_u} (\phi^2 + k_\epsilon) \nabla_{(t,x)} u + \frac{\lambda_w}{\lambda_u} w (\nabla_{(t,x)} u \cdot w) \right) + u = u_0 \quad (5)$$

$$-\epsilon \Delta_{(t,x)} \phi + \left(\frac{1}{4\epsilon} + \frac{\mu_u}{2\nu} \|\nabla_{(t,x)} u\|^2 \right) \phi = \frac{1}{4\epsilon} \quad (6)$$

$$-\frac{\mu_w}{\lambda_w} \operatorname{div}_{(t,x)} (P[\phi] \nabla_{(t,x)} w) + (\nabla_{(t,x)} u \cdot w) \nabla_{(t,x)} u = 0 \quad (7)$$

as the Euler–Lagrange equations characterizing the necessary conditions for a solution (u, w, ϕ) of the above stated phase field approach. Let us emphasize that the full Euler–Lagrange equations, characterizing a global minimizer of the energy, would in addition involve variations of $E_{\text{reg},w}$ with respect to ϕ .

Following again Ambrosio and Tortorelli, our resulting algorithm involves an iteration solving three linear partial differential equations:

- Step 0.** Initialize $u = u_0$, $\phi \equiv 1$, and $w \equiv (1, 0)$.
- Step 1.** Solve (5) for fixed w, ϕ .
- Step 2.** Solve (6) for fixed u, w .
- Step 3.** Solve (7) for fixed u, ϕ , return to **Step 1** if not converged.

5 Finite Element Discretization

We proceed similarly to the Finite Element method proposed by Bourdin and Chambolle [27,28] for the phase field approximation of the Mumford–Shah functional. To solve the above system of PDEs, we discretize $[0, T] \times \Omega$ by a regular

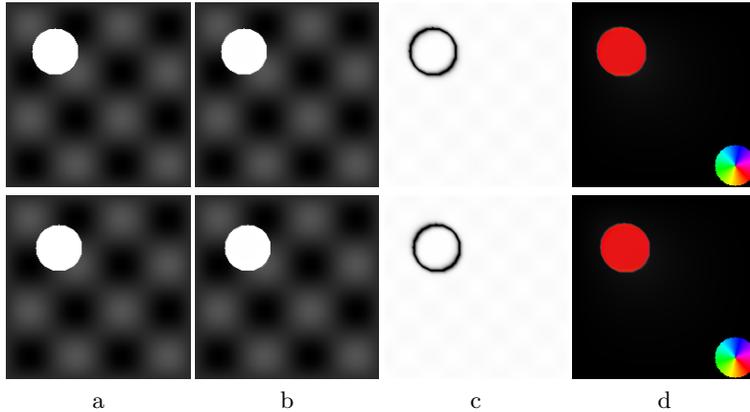


Fig. 1. Top to bottom two frames of the test sequence (a) and corresponding smoothed image (b), phase field (c) and optical flow (color coded) (d).

hexahedral grid. In the following, the spatial and temporal grid cell sizes are denoted by h and τ respectively, i.e. image frames are at a distance of τ and pixels of each frame are sampled on a regular mesh with cell size h . To avoid tri-linear interpolation problems, we subdivide each hexahedral cell into 6 tetrahedra. On this tetrahedral grid, we consider the space of piecewise affine, continuous functions \mathcal{V} and ask for discrete functions $U, \Phi \in \mathcal{V}$ and $V \in \mathcal{V}^2$, such that the discrete and weak counterparts of the Euler Lagrange equations (5), (6) and (7) are fulfilled. This leads to solving systems of linear equations for the vectors of the nodal values of the unknowns U, Φ, V . Using an efficient custom-designed compressed row sparse matrix storage, we can treat datasets of up to $K = 10$ frames of $N = 500, M = 320$ pixels in less than 1GB memory. The linear systems of equations are solved applying a classical conjugate gradient method. For the pedestrian sequence (Fig. 5), one such iteration takes 47 seconds on a Pentium IV PC at 1.8 GHz running Linux. The complete method converges after 2 or 3 such iterations. Large video sequences are computed by shifting a window of $K = 6$ frames successively in time. Thus temporal boundary effects are avoided.

6 Results and Discussion

We present here several results of the proposed method for two dimensional image sequences. In the considered examples, the parameter setting $\epsilon = h/4$, $\mu_u = h^{-2}$, $\mu_w = \lambda_u = 1$, $\lambda_w = 10^5 h^{-2}$ and $C(\epsilon) = \epsilon$, $\delta = \epsilon$ has proven to give good results.

We first consider a simple example of a white disk moving with constant speed $v = (1, 1)$ on a black background (Fig. 1). A small amount of smoothing results from the regularization energy $E_{\text{reg},u}^\epsilon$ (Fig. 1(b)), which is desirable to ensure robustness in the resulting optical flow term $\nabla_{(t,x)} u \cdot w$ and removes noisy artifacts in real-world videos, e.g. Fig. 4 and Fig. 5. The phase field clearly captures the moving object's contour. The optical flow is depicted in Fig. 1(c) by color coding

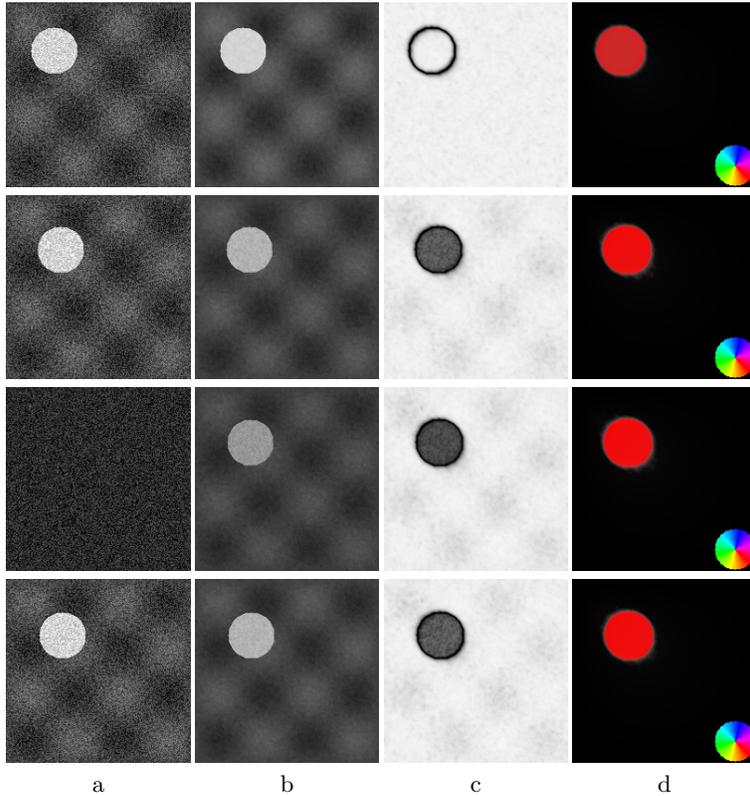


Fig. 2. Noisy circle sequence: From top to bottom, frames 3 and 9 – 11 are shown. (a) original image sequence, (b) smoothed images, (c) phase field, (d) estimated motion (color coded)

the vector directions as shown by the lower-right color wheel. Clearly, the method is able to extract the uniform motion of the disc. The optical flow information, available only on the motion edges (black in Fig. 1(c)), is propagated into the information-less area inside the moving disk, yielding the final result.

In the next example, we revisit the simple moving circle sequence, but add noise to it. We also completely destroy the information of frame 10 in the sequence (Fig. 2). Figure 2 shows the results for frames 3 and 9 – 11. We see that the phase field detects the missing circle in the destroyed frame as a temporal edge surface in the sequence, i.e. ϕ drops to zero in the temporal vicinity of the destroyed frame. This is still visible in the previous and next frames, shown in the second and third row. However, this does not hamper the restoration of the correct optical flow field, shown in the fourth column. This result is due to the anisotropic smoothing of information from the frames close to the destroyed frame. For this example, we used $\epsilon = 0.4h$.

A second synthetic example is shown in Fig. 3, using data from the publicly available collection at [29]. Here, a textured sphere spins on a textured background (Fig. 3(a)). Again, our method is able to clearly segment the moving

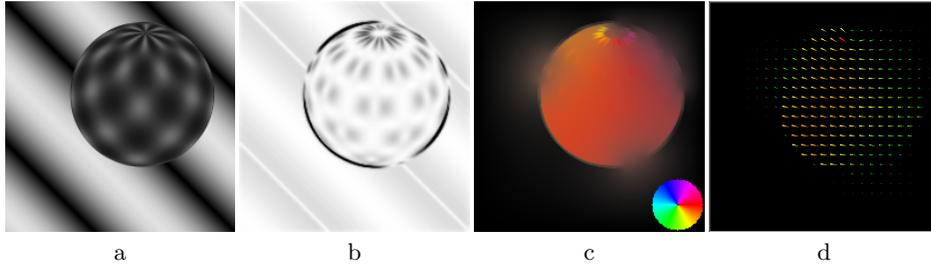


Fig. 3. Rotating sphere: smoothed image (a), phase field (b), optical flow (color coded) (c), optical flow (vector plot, color coded magnitude) (d)

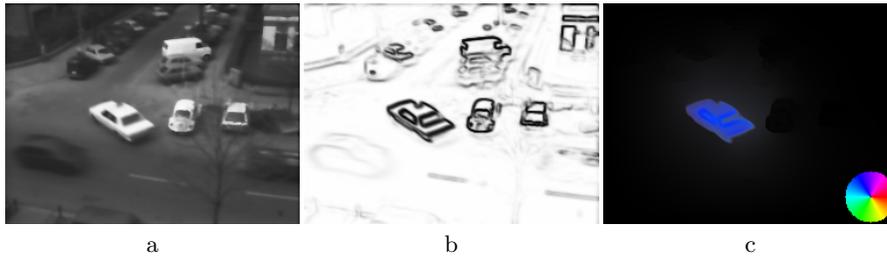


Fig. 4. Taxi sequence: smoothed image (a), phase field (b), and flow field (c)

object from the background, even though the object doesn't change position. We used a phase field parameter $\epsilon = 0.15h$. The extracted optical flow clearly shows the spinning motion (Fig. 3(d)) and the discontinuous motion field.

We next consider a known real video sequence, the so-called Hamburg taxi sequence. Figure 4 shows the smoothed image (u), phase field ϕ and color-coded optical flow field (w). Our method detects well the image edges (Fig. 4 b). Also, the upper-left rotating motion of the central car is extracted accurately (Fig. 4 c). As it should be, the edges of the stationary objects, clearly visible in the phase field, do not contribute to the optical flow. Moreover, the moving car is segmented as one single object in the optical flow field, i.e. the motion information is extended from the moving edges, i.e. car and car windscreen contours, to the whole moving shape.

Finally, we consider a complex video sequence, taken under outdoor conditions by a monochrome video camera. The sequence shows a group of walking pedestrians (Fig. 5 (top)). The human silhouettes are well extracted and captured by the phase field (Fig. 5(middle)). We do not display a vector plot of the optical flow, as it is hard to interpret it visually at the video sequence resolution of 640 by 480 pixels. However, the color-coded optical flow plot (Fig. 5(bottom)) shows how the method is able to extract the moving limbs of the pedestrians. The overall red and blue color corresponds to the walking directions of the pedestrians. The estimated motion is smooth inside the areas of the individual pedestrians and not smeared across the motion boundaries. In addition, the algorithm nicely segments the different moving persons. The cluttered background poses no big problem to the segmentation, nor are the edges of occluding and overlapping pedestrians, who are moving at almost the same speed.

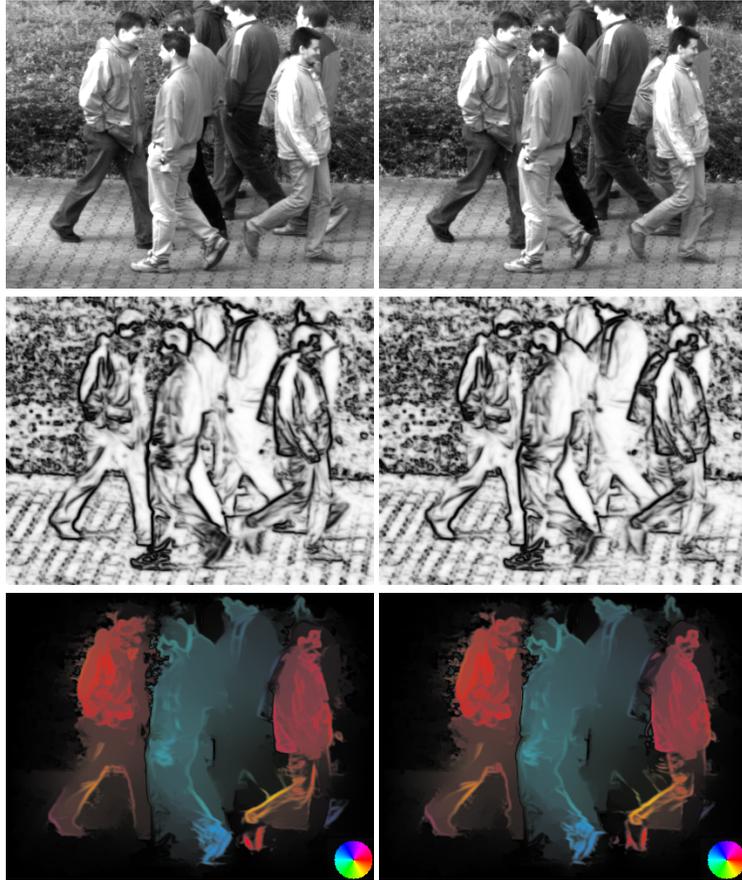


Fig. 5. Pedestrian video: frames from original sequence (top); phase field (middle); optical flow, color coded (bottom)

References

1. Horn, B.K.P., Schunk, B.: Determining optical flow. *Artificial Intelligence* **17** (1981) 185–204
2. Nagel, H.H., Enkelmann, W.: An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. on PAMI* **8**(5) (1986) 565–593
3. Weickert, J., Schnörr, C.: A theoretical framework for convex regularizers in pde-based computation of image motion. *Int. J. of Comp. Vision* **45**(3) (2001) 245–264
4. Bruhn, A., Weickert, J., Feddern, C., Kohlberger, T., Schnörr, C.: Real-time optical flow computation with variational methods. In Petkov, N., Westenberg, M.A., eds.: *CAIP 2003*. Volume 2756 of LNCS., Springer (2003) 222–229
5. Wang, J.Y.A., Adelson, E.H.: Representating moving images with layers. *IEEE Trans. on Im. Proc.* **3**(5) (1994) 625–638
6. Schnörr, C.: Segmentation of visual motion by minimizing convex non-quadratic functionals. In: *12th ICPR*. (1994)

7. Odohez, J.M., Bouthemy, P.: Robust multiresolution estimation of parametric motion models. *J. of Vis. Comm. and Image Rep.* **6**(4) (1995) 348–365
8. Odohez, J.M., Bouthemy, P.: Direct incremental model-based image motion segmentation for video analysis. *Sig. Proc.* **66** (1998) 143–155
9. Caselles, V., Coll, B.: Snakes in movement. *SIAM J. Num. An.* **33** (1996) 2445–2456
10. Memin, E., Perez, P.: A multigrid approach for hierarchical motion estimation. In: *ICCV.* (1998) 933–938
11. Paragios, N., Deriche, R.: Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans. on PAMI* **22**(3) (2000) 266–280
12. Droske, M., Ring, W.: A Mumford-Shah level-set approach for geometric image registration. *SIAM Appl. Math.* (2005) to appear.
13. Authors: Mumford-shah based registration. *Computing and Visualization in Science* (2005) submitted.
14. Kapur, T., Yezzi, L., Zöllei, L.: A variational framework for joint segmentation and registration. *IEEE CVPR* (2001) 44–51
15. Unal, G., Slabaugh, G., Yezzi, A., Tyan, J.: Joint segmentation and non-rigid registration without shape priors. (2004)
16. Vemuri, B., Ye, J., Chen, Y., Leonard, C.: Image registration via level-set motion: Applications to atlas-based segmentation. *Med. Im. Analysis* **7** (2003) 1–20
17. Davatzikos, C.A., Bryan, R.N., Prince, J.L.: Image registration based on boundary mapping. *IEEE Trans. Med. Imaging* **15**(1) (1996) 112–115
18. Cremers, D., Soatto, S.: Motion competition: A variational framework for piecewise parametric motion segmentation. *Int. J. of Comp. Vision* **62**(3) (2005) 249–265
19. Cremers, D., Kohlberger, T., Schnörr, C.: Nonlinear shape statistics in mumford-shah based segmentation. In: *7th ECCV.* Volume 2351 of LNCS. (2002) 93–108
20. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* **42** (1989) 577–685
21. Brox, T., Bruhn, A., Papenber, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In Pajdla, T., Matas, J., eds.: *Proc. of the 8th ECCV.* Volume 3024 of LNCS. (2004) 25–36
22. Amiaz, T., Kiryati, N.: Dense discontinuous optical flow via contour-based segmentation. In: *Proc. ICIP 2005.* Volume III. (2005) 1264–1267
23. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the mumford and shah model. *Int. J. Computer Vision* **50** (2002) 271–293
24. Nir, T., Kimmel, R., Bruckstein, A.: Variational approach for joint optic-flow computation and video restoration. Technical report, Dep. of C. S. - Israel Inst. of Tech., Haifa, Israel (2005)
25. Ambrosio, L., Tortorelli, V.M.: On the approximation of free discontinuity problems. *Boll. Un. Mat. Ital. B* **6**(7) (1992) 105–123
26. Ambrosio, L., Fusco, N., Pallara, D.: *Functions of bounded variation and free discontinuity problems.* Oxford University Press (2000)
27. Bourdin, B.: Image segmentation with a Finite Element method. *ESIAM: Math. Modelling and Num. Analysis* **33**(2) (1999) 229–244
28. Bourdin, B., Chambolle, A.: Implementation of an adaptive Finite-Element approximation of the Mumford-Shah functional. *Numer. Math.* **85**(4) (2000) 609–646
29. Group, C.V.R.: *Optical flow datasets.* Univ. of Otago, New Zealand, www.cs.otago.ac.nz/research/vision (2005)