A Generative Model Based Approach to Motion Segmentation*

Daniel Cremers[1] and Alan Yuille[2]

[1] Department of Computer Science
University of California at Los Angeles

[2] Department of Statistics and Psychology
University of California at Los Angeles

Abstract. We address the question of how to choose between different likelihood functions for motion estimation. To this end, we formulate motion estimation as a problem of Bayesian inference and compare the likelihood functions generated by various models for image formation. In contrast to alternative approaches which focus on noise in the measurement process, we propose to introduce noise on the level of the velocity, thus allowing it to vary around a given model. We show that this approach generates additional normalizations not present in previous likelihood functions. We numerically evaluate the proposed likelihood in a variational framework for segmenting the image plane into domains of piecewise constant motion. The evolution of the motion discontinuity set is implemented using the level set framework.

1 Introduction

The problem of estimating motion from an image sequence has been addressed by minimizing appropriate cost functionals, which depend on the gray values of the image sequence and its spatial and temporal derivatives. Since the seminal work of Horn and Schunck [5], a wealth such variational methods have been proposed (cf. [9,14,7,15]). Commonly these functionals consist of a fidelity term which measures how well the local gray values are in accordance with a specific motion, and of a prior which enforces a certain regularity of the estimated motion field.

Yet, the question remains: Which cost functional is appropriate for the given task? As for the regularity term, it clearly depends on the prior knowledge about what kinds of motion fields can be expected. In particular, one can impose smoothness of the estimated motion fields [5], smoothness with discontinuities [12,15], parametric [1] and piecewise parametric motion models [2], or higher-level regularity constraints derived from fluid mechanics [6].

In this paper, we are concerned with the fidelity term. This likelihood can be derived from a generative model of image formation (cf. [16]). One makes certain assumptions about how the image sequence is generated – for example one may

*Supp: ONR N00014-02-1-0720, NSF SBR-9631682,0240148, NIH R01EY013875-01A1.

B. Michaelis and G. Krell (Eds.): DAGM 2003, LNCS 2781, pp. 313–320, 2003.

assume that a static scene is transformed according to a certain velocity. As we will see, it is of particular importance to address the question at which level noise is allowed to enter this image formation process: Different models of noise will induce different normalizations (i.e. scalings or weightings) of the data term in the resulting cost functional.

Commonly noise is introduced as additive Gaussian noise to the measurements [16,13,10]. In contrast, we argue that one should allow statistical variation of the quantity which is to be estimated and introduce noise directly on the velocity. As a consequence, we derive a novel fidelity term for motion estimation.

Based on this fidelity term, we propose a variational method which permits to segment the image plane into domains of piecewise constant motion. Segmentation and motion estimate are obtained by minimizing the proposed energy functional jointly with respect to the motion vectors for each region and with respect to the boundary separating these regions. We implement this motion boundary by the level set method [11], since this implicit representation facilitates topological changes of the evolving boundary such as splitting and merging. Energy minimization amounts to alternating the two fractional steps of solving an eigenvalue problem for the motion vectors in each region, and of evolving the level set function which encodes the motion boundaries.

Numerical results demonstrate that the proposed likelihood function induces accurate segmentations of the image plane based exclusively on the motion information extracted from two consecutive images of a sequence. Our approach is computationally efficient and tracking applications are conceivable.

2 Motion Estimation as Bayesian Inference

Let $\Omega \subset \mathbb{R}^2$ denote the image plane and let $I : \Omega \times \mathbb{R} \to \mathbb{R}$ be a gray value image sequence. In the following, we will assume that this image sequence represents an unknown scene function $s : \Omega \to \mathbb{R}$ which undergoes a motion $v : \Omega \to \mathbb{R}^2$ at each point $x \in \Omega$.[1] The motion field v can be estimated by maximizing the conditional probability

$$\mathcal{P}(v \mid I) = \frac{\mathcal{P}(I \mid v)\ \mathcal{P}(v)}{\mathcal{P}(I)}, \tag{1}$$

with respect to v. Here $\mathcal{P}(v)$ represents the prior on the velocity field.

The focus of the present paper lies on modeling the conditional probability $\mathcal{P}(I \mid v)$ for an image sequence I given a velocity field v. To this end, we will revert to generative models of image formation.

3 Generative Models with Measurement Noise

Let us assume that the image sequence is obtained from a scene function $s : \Omega \to \mathbb{R}$ by applying a certain velocity field v:

$$I(x + vt, t) = s(x), \tag{2}$$

[1] Since we are only concerned with the velocity at a fixed time instance, we will ignore the temporal variation of the velocity field.

where $s(x) = I(x, 0)$. Given a particular model of how noise enters this image formation process, one can derive the conditional distribution $\mathcal{P}(I \,|\, v, s)$ for the intensity function I given the velocity field v and the (unknown) scene function s. The desired likelihood $\mathcal{P}(I \,|\, v)$ can then be obtained by marginalization with respect to the scene function:

$$\mathcal{P}(I \,|\, v) = \int \mathcal{P}(I \,|\, v, s)\, \mathcal{P}(s)\, ds. \tag{3}$$

In the following, we will review three different approaches to introduce noise into the above image formation process. In all three cases, the noise enters on the level of the measurements. Each noise model entails a different likelihood function $\mathcal{P}(I \,|\, v)$.

3.1 Additive Gaussian Noise in the Measurement Equation

Weiss and Fleet [16] suggested to introduce additive Gaussian noise to the measurement equation: $I(x+vt, t) = s(x) + \sigma\eta$, where η represents zero mean Gaussian noise with variance 1. It follows that the conditional distribution is given by

$$\mathcal{P}(I \,|\, v, s) \propto \exp\left(-\frac{1}{2\sigma^2} \int \left(I(x+vt,t) - s(x)\right)^2 \, dx\, dt\right). \tag{4}$$

Assuming a uniform prior over the scene functions s and given a fixed time interval for observation, the marginalization in (3) can be carried out analytically, one obtains [16]:

$$\mathcal{P}(I \,|\, v) \propto \exp\left(-\frac{1}{2\sigma^2} \int \left(I(x,t) - \hat{s}(x-vt)\right)^2 \, dx\, dt\right), \tag{5}$$

where the function $\hat{s}(x)$ turns out to be the mean of $I(x+vt, t)$ over the time window of observation.

3.2 Additive Gaussian Noise on the Temporal Derivative

If the velocity in (2) is sufficiently small and the intensity function sufficiently smooth, then one can perform a first-order Taylor expansion which yields the well-known optic flow constraint equation:

$$v^T \nabla I \,+\, I_t = 0. \tag{6}$$

Due to this approximation, the resulting conditional density no longer depends on the unknown scene function such that the marginalization in (3) becomes trivial. Simoncelli [13] suggested to introduce additive Gaussian noise to the temporal derivative in (6). This yields the conditional probability

$$\mathcal{P}(I \,|\, v) \propto \exp\left(-\frac{1}{2\sigma^2} \int (v^T \nabla I \,+\, I_t)^2 \, dx\, dt\right), \tag{7}$$

which has become a popular likelihood function for motion estimation since the work of Horn and Schunck [5].

3.3 Additive Gaussian Noise on Spatial and Temporal Derivatives

Extending the above model, one can assume that both the spatial and the temporal derivative in (6) are corrupted by zero-mean additive Gaussian noise. As shown in [10], this assumption results in a conditional probability of the form

$$\mathcal{P}\big(I \,|\, v\big) \propto \exp\left(-\frac{1}{2\sigma^2}\int \frac{(v^T\,\nabla I \,+\, I_t)^2}{1+|v|^2}\,dx\,dt\right). \tag{8}$$

Compared to the previous likelihood, this introduces an additional normalization with respect to the length of the homogeneous velocity vector. However, it was pointed out in [8] that the noise on the spatio-temporal derivatives may not be independent, since the latter are calculated from digitized images.

4 A Generative Model with Noise on the Velocity Field

The above generative models incorporate noise on different levels of the measurement process. In the following, we will argue that it may be favorable to consider noise models for the quantity which is to be estimated, namely the velocity field itself. Consider the general case of an image sequence in which different regions are moving according to different velocity models (for example parametric motion models). In order to partition the image plane into regions of homogeneous velocity, we need to be able to measure how well a certain velocity is in accordance with a given model.

To this end, we assume that the true velocity $\hat{v}$ may deviate from the model velocity v according to a noise model of the form:

$$\hat{v} = v + g(v)\,\sigma\,\bar{\eta}, \tag{9}$$

where $\sigma\bar{\eta} \in \mathbb{R}^2$ is 2-d Gaussian noise of width σ, scaled by a factor $g(v) = \sqrt{1+|v|^2/|v_o|^2}$, which implies that the noise increases with the magnitude of the velocity and is non-zero for zero velocity. For simplicity, we set the normalization constant to $v_0 = 1$. As in equation (2), we assume that the image sequence is generated from a static scene $s(x)$ deformed according to the velocity field $\hat{v}$. Assuming the noise to be sufficiently small and the intensity function to be sufficiently smooth, we can perform a Taylor expansion with respect to the noise:

$$s(x) = I(x+\hat{v}t, t) \approx I(x+vt,t) \,+\, t\,g(v)\,\sigma\,\bar{\eta}^T\,\nabla I. \tag{10}$$

Rearranging terms, we obtain:

$$\frac{I(x+vt,t)-s(x)}{t\,g(v)\,|\nabla I|} \approx \sigma\eta, \tag{11}$$

where η denotes 1-d Gaussian noise. This corresponds to a likelihood function of the form:

$$\mathcal{P}\left(I \,|\, s, v\right) \propto \exp\left(-\frac{1}{2\sigma^2}\int \frac{\big(I(x+vt,t)-s(x)\big)^2}{t^2\,g^2(v)\,|\nabla I|^2}\,dx\,dt\right). \tag{12}$$

Compared to the likelihood (4) for additive noise in the measurement equation, this likelihood includes normalizations with respect to the spatial gradient, the magnitude of the homogeneous velocity and time.

For sufficiently small velocity v, we can expand the intensity function even further:

$$I(x + \hat{v}t, t) \approx I(x, 0) + t\, g(v)\, \sigma\, \bar{\eta}^T\, \nabla I + t\, v^T \nabla I \, + \, t\, I_t. \tag{13}$$

Making use of the fact that $s(x) = I(x, 0)$, we obtain a conditional probability which is independent of the underlying scene function s:

$$\mathcal{P}(I \,|\, v) \propto \exp\left(\frac{-1}{2\sigma^2} \int \frac{(v^T \nabla I + I_t)^2}{g^2(v) |\nabla I|^2} dx dt\right) = \exp\left(\frac{-1}{2\sigma^2} \int \frac{(\bar{v}^T \nabla_3 I)^2}{|\bar{v}|^2 |\nabla I|^2} dx dt\right). \tag{14}$$

Here $\nabla_3 I \in \mathbb{R}^3$ and $\bar{v} \in \mathbb{R}^3$ represent the spatio-temporal image gradient and the homogeneous velocity vector, respectively.

Note that, compared to the likelihood for noise on the spatial and temporal derivative in equation (8), this likelihood function includes an additional normalization with respect to the image gradient. A similar likelihood function (normalized with respect to the spatio-temporal gradient rather than the spatial one) was derived in [2] based on purely geometric considerations. Here the likelihood function with both normalizations is derived from a generative model with noise on the velocity field as introduced above.

In numerical evaluation, we found that these normalizations are important in the case of motion *segmentation* which differs from motion *estimation* in that one needs to associate each image location with one or the other motion hypothesis.

5 Variational Motion Segmentation

In the previous section, we derived the fidelity term (14) for motion estimation. In the present section, we will incorporate this fidelity term into a variational framework for motion segmentation. To this end, we revert to the Bayesian approach introduced in Section 2 and specify a prior $\mathcal{P}(v)$ on the velocity field which enforces the formation of piecewise constant velocity fields. Extensions to models of piecewise parametric motion are conceivable (cf. [3]), they are however beyond the scope of this paper.

We discretize the velocity field over a set of disjoint regions $R_i \subset \Omega$ with constant homogeneous velocity $\bar{v}_i \in \mathbb{R}^3$. We now assume the prior probability on the velocity field to only depend on the length of the boundary C separating these regions. Maximizing the conditional probability (1) is equivalent to minimizing its negative log likelihood. Up to a constant, the latter is given by the energy:

$$E(C, \{\bar{v}_i\}) = \sum_{i=1}^{n} \int_{R_i} \frac{(\bar{v}_i^T \nabla_3 I)^2}{|\bar{v}_i|^2 \,|\, \nabla I|^2}\, dx \; + \nu\, |C|. \tag{15}$$

Since we only consider the spatio-temporal image derivatives at a given time instance calculated from two consecutive frames of the sequence, the temporal integration in the likelihood (14) disappears. For an extension of a related approach to the problem of spatio-temporal motion segmentation we refer to [4].

6 Energy Minimization

In order to generate a segmentation of the image plane into areas of piecewise constant motion, we minimize energy (15) by alternating the two fractional steps of updating the motion vectors $\bar{v}_i$ and evolving the motion boundary.

For fixed boundary C, minimization with respect to $\bar{v}_i$ results in the eigenvalue problem:

$$\bar{v}_i = \arg\min_{\bar{v}} \frac{\bar{v}^T M_i \bar{v}}{\bar{v}^T \bar{v}}, \quad \text{with } M_i = \int_{R_i} \frac{\nabla_3 I^T \nabla_3 I}{|\nabla I|^2} \, dx \tag{16}$$

The solution of (16) is given by the eigenvector corresponding to the smallest eigenvalues of M_i, normalized such that its third component is 1.

Conversely, for fixed motion vectors, the gradient descent equation on the boundary C is given by:

$$\frac{dC}{dt} = (e_j - e_k) \cdot n \, - \, \nu\kappa, \tag{17}$$

where n denotes the normal vector on the boundary, κ denotes the curvature, the indices 'k' and 'j' refer to the regions adjoining the contour, and e_i is the energy density given by the integrand in the functional (15).

We implemented this evolution using the level set method [11], since it is independent of a particular parameterization and permits to elegantly model topological changes of the boundary such as splitting and merging. For details, we refer to [2].

7 Numerical Results

7.1 Simultaneous Segmentation and Motion Estimation

Figure 1 presents several steps during the energy minimization for two consecutive images from a sequence showing a rabbit which moves to the right. By minimizing a single cost functional both the boundaries and the estimated motion are progressively improved. The final segmentation gives both an accurate reconstruction of the objects location and an estimate of the motion of object and background.

7.2 Segmenting Multiple Motion

The cost functional (15) permits a segmentation into multiple differently moving regions. Figure 2 shows segmentation results obtained for an image sequence showing two cars moving to the top right, while the background is moving to the bottom left. The original sequence was recorded with a static camera by D. Koller and H.-H. Nagel.[2] To increase its complexity, we artificially translated the

[2] KOGS/IAKS, Univ. of Karlsruhe, `http://i21www.ira.uka.de/image_sequences/`

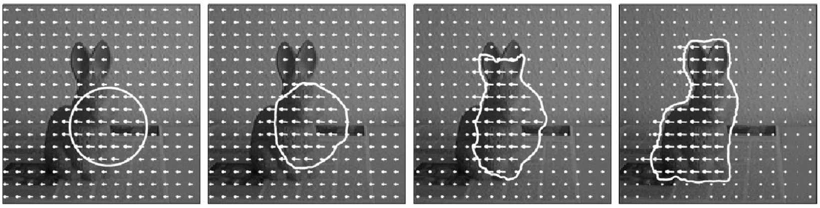

Fig. 1. Motion segmentation obtained by minimizing functional (15) simultaneously with respect to the motion models and the separating motion boundary. During minimization the boundary location and estimated motion are progressively improved. Thus the object's location and motion are simultaneously reconstructed. In contrast to the approach proposed in [3], the present formulation does not require a posterior normalization of the driving terms in the evolution equation.

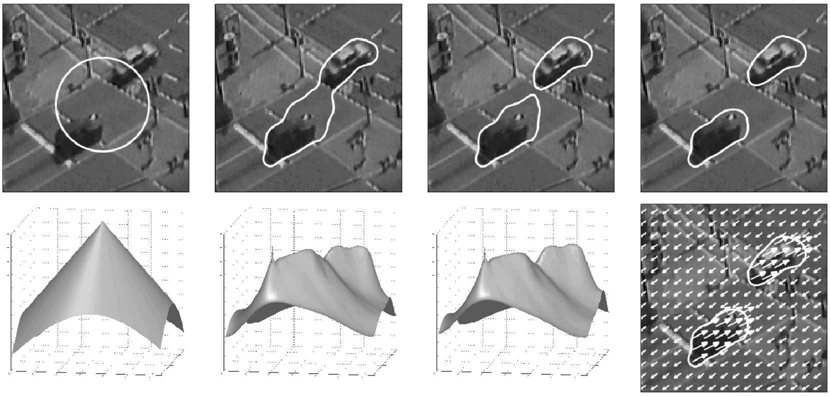

Fig. 2. Evolution of motion boundary (**top**) given by the zero crossing of the level set function (**bottom**) for moving cars on moving background. While the two cars cannot be segmented based on intensity criteria such as edges or gray value homogeneity, the motion segmentation gives an accurate reconstruction of the object location and its motion (**bottom right**).

second frame thereby simulating a moving camera. An accurate reconstruction of the object location and an estimate of the motion of objects and background (bottom right) is obtained by minimizing the proposed functional. Due to the representation of the boundary as the zero level set of the function shown in the bottom row, the boundary is free to undergo splitting and merging.

8 Conclusion

We addressed the question of choosing appropriate likelihood functions for variational motion segmentation. Motion segmentation differs from gray value segmentation in that the velocity field is not identical with the measured signal, but rather a derived quantity. For this reason, we argued that one needs to break with the "signal plus noise paradigm". We proposed a novel model of image formation in which the velocity is permitted to vary statistically. In contrast to alternative models, we assume that the effect of noise in the measurements is dominated by the effect caused by variations of the true velocity around a particular model velocity. As a consequence, the resulting likelihood function contains additional normalizations with respect to the velocity magnitude and the image gradient. This novel likelihood function is shown to induce highly accurate segmentations of the image plane, obtained purely on the basis of motion information.

References

1. M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise–smooth flow fields. *Comp. Vis. Graph. Image Proc.: IU*, 63(1):75–104, 1996.
2. D. Cremers. A variational framework for image segmentation combining motion estimation and shape regularization. In C. Dyer and P. Perona, editors, *IEEE Int. Conf. on Comp. Vis. and Patt. Recog.*, Madison, Wisconsin, June 2003. To appear.
3. D. Cremers and C. Schnörr. Motion Competition: Variational integration of motion segmentation and shape regularization. In L. van Gool, editor, *Pattern Recognition*, volume 2449 of *LNCS*, pages 472–480, Zürich, Sept. 2002. Springer.
4. D. Cremers and S. Soatto. Combining motion estimation and surface regularization for variational image sequence segmentation. In *ICCV*, Nice, Oct. 2003. To appear.
5. B.K.P. Horn and B.G. Schunck. Determining optical flow. *A.I.*, 17:185–203, 1981.
6. T. Kohlberger, E. Memin, and C. Schnörr. Variational dense motion estimation using the Helmholtz decomposition. In L. Griffin, editor, *Int. Conf. on Scale Space Theories in Comp. Vis.*, LNCS. Springer, 2003. To appear.
7. P. Kornprobst, R. Deriche, and G. Aubert. Image sequence analysis via partial differential equations. *J. Math. Im. Vis.*, 11(1):5–26, 1999.
8. H.H. Nagel. Optical flow estimation and the interaction between measurement errors at adjacent pixel positions. *Int. J. of Comp. Vis.*, 15(3):271–288, 1995.
9. H.H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 8(5):565–593, 1986.
10. O. Nestares, D.J. Fleet, and D.J. Heeger. Likelihood functions and confidence bounds for total-least-squares problems. In *Proc. Conf. Computer Vis. and Pattern Recog.*, volume 1, pages 760–767, Hilton Head Island, SC, June, 2000.
11. S. J. Osher and J. A. Sethian. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. *J. of Comp. Phys.*, 79:12–49, 1988.
12. C. Schnörr. Computation of discontinuous optical flow by domain decomposition and shape optimization. *Int. J. of Comp. Vis.*, 8(2):153–165, 1992.
13. E.P. Simoncelli. *Distributed Representation and Analysis of Visual Motion.* PhD thesis, Dept. of Elect. Eng. and Comp. Sci., MIT, Cambridge, 1993.
14. E.P. Simoncelli, E.H. Adelson, and D.J. Heeger. Probability distributions of optical flow. In *Proc. Conf. Comp. Vis. Patt. Rec.*, pages 310–315, Hawaii, June 1991.
15. J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE–based computation of image motion. *IJCV*, 45(3):245–264, 2001.
16. Y. Weiss and D.J. Fleet. Velocity likelihoods in biological and machine vision. In M.S. Lewicki R.P.N. Rao, B.A. Olshausen, editor, *Probabilistic Models of the Brain: Perception and Neural Function*, pages 81–100. MIT Press, 2001.