Uncertainty Optimization for Robust Dynamic Optical Flow Estimation

Volker Willert, Mark Toussaint, Julian Eggert, Edgar Körner

2007

Uncertainty Optimization for Robust Dynamic Optical Flow Estimation

Volker Willert, Marc Toussaint[*], Julian Eggert, Edgar Körner
HRI Europe GmbH
Carl-Legien-Str. 30
D-63073 Offenbach, Germany
volker.willert@honda-ri.de

Tu Berlin[*]
Franklinstr. 28/29
D-10587 Berlin, Germany
mtoussai@cs.tu-berlin.de

Abstract

We develop an optical flow estimation framework that focuses on motion estimation over time formulated in a Dynamic Bayesian Network. It realizes a spatiotemporal integration of motion information using a dynamic and robust prior that incorporates spatial and temporal coherence constraints on the flow field. The main contribution is the embedding of these particular assumptions on optical flow evolution into the Bayesian propagation approach that leads to a computationally feasible Two-Filter inference method and is applicable for on- and offline parameter optimization. We analyse the possibility to optimize imposed Student's t-distributed model uncertainties, which are the camera noise and the transition noise. Experiments with synthetic sequences illustrate how the probabilistic framework improves the optical flow estimation because it allows for noisy data, motion ambiguities and motion discontinuities.

1. Introduction

The analysis of pixel movement in an image sequence allows to infer the motion of objects as well as the self-motion of the image capturing device. Usually, these movements are estimated by processing spatial and temporal derivatives of image values (like pixel intensities) which hold information about the movement of image structure. The observed image data from which image motion is inferred is noisy and the model assumptions describe only approximations about the real physical relation between spatiotemporal image value changes and image motion caused by a moving threedimensional environment or self-movement of the system within the environment. Technical approaches for the analysis of optical flow need to cope with 1) correspondence problems due to ambiguities (*e.g*. periodicity or lack of texture) in the image structure, 2) camera noise, 3) spatial motion discontinuities, and 4) temporal movement changes.

Recent and very accurate optical flow methods [3] use local spatiotemporal information by means of a structure tensor and combine it with global spatiotemporal smoothness constraints on the flow field. Further on, robust penalty functions or robust statistics [2] improve the estimations especially at motion discontinuities because they are less sensitive to outliers. In addition, there are some new attempts to extract detailed scene-dependent prior knowledge about the underlying spatial statistics of optical flow that hold specific spatial relations between the velocity vectors of the flow field [10]. Mostly, spatial relations - whether they are learned from image data or given by the modeller - are incorporated into the models via *Markov Random Fields* [5, 8, 10].

Image motion is a dynamic feature of an image sequence and the longer the spatiotemporal process is observed the more precise and detailed we can estimate and predict the motion contained in an image sequence. Nevertheless, the majority of optical flow approaches calculate the motion estimations independently for each point in time and the free parameters of the models are optimized anew to achieve the best result dependent on that point in time. Except for [7], the choice of the parameters is done by hand or by learning the optimal prior knowledge for one specific flow pattern or even one specific scene. In fact, the more accurate the methods are, the more specific prior knowledge about the scene has been incorporated. But this is not necessarily a general choice suitable to cover several groups of optical flow patterns usually present in a changing scene, like *e.g*. rotation and expansion, or at least several instances of one group, like *e.g*. expansion with different foci of expansion, but (more or less) only one specific instance.

Starting with the work of Simoncelli *et al*. [11] a number of investigations have been undertaken to make allowance for motion uncertainties and to find proper velocity distributions that are able to represent any kind of motion uncertainty, especially multiple motions [14]. Burgi *et al*. [4] are among the first to express continuous optical flow estimation over time in terms of general Bayesian tracking. In

[15] the focus is on the temporal dynamics of such velocity distributions in a hierarchical probabilistic network, showing the possibility to disambiguate uncertain visual motion estimates over time.

Both, the approaches that incorporate motion estimates from previous timesteps [12, 2, 4, 15] and those that stick to the measurements made at one point in time [3, 10, 11] do not try to systematically optimize and adapt their parameters to a continuously observed data stream, which is one major aspect of the following work.

In this paper we concentrate on 1) a motion estimation model that continuously combines motion information over time with the aim to be the more accurate the more data has been processed by the system as long as the data holds its movement like predicted by the model and 2) being able to cope with motion uncertainty, noisy data and continuous temporal changes of pixel movement. Compared to recent approaches [10] that learn scene specific prior knowledge and are therefore capable to incorporate detailed discontinuity information, our prior assumptions are simpler only imposing some degree of spatiotemporal coherence to constrain the optical flow.

For this reason we propose a probabilistic motion estimation framework which is formulated as a *Dynamic Bayesian Network* [9] to deliver optical flow estimations continuously over long image sequences. After derivation of a special *Two-Filter* [6] inference approach we optimize the free parameters of the built-in uncertainty assumptions to maximize the probability of the smoothed posterior using the EM-algorithm. More specifically, we try to find the optimal parameterization of the camera noise and the transition noise represented by the *covariances* assuming 1) *Student's t-distributions* with fixed *degrees of freedom* ν and 2) the limit to infinity $\nu \to \infty$ which results in a *Gaussian distribution*.

We show that it is possible to optimize the uncertainties for different kinds of noise using *Gaussian* noise to discuss the robustness against camera noise and *Salt&Pepper* noise to analyse the robustness against outliers. Further on, we demonstrate the possibility to disambiguate motion uncertainties because of correspondence problems, like *e.g.* the *aperture problem* and the *blank wall* problem. In addition, *on-* and *offline* optimization are compared by 1) maximizing the smoothed posterior gained from the two-filter inference for the offline case and 2) by maximizing the forward filtered posterior for the online case which leads to an ongoing adaptation of the uncertainties.

2. Dynamic Bayesian Network Model

To derive algorithms for dynamic visual flow field estimation using probabilistic filtering, smoothing and parameter learning methods we specify a complete data likelihood of a sequence $I^{0:T}$ of $T+1$ images. We do this by assuming the generative model for such an image sequence as given by the Dynamic Bayesian Network in Fig. 1 a). Here, I^t is the grey value image at time slice t with entries I_x^t that are the grey values at all pixel locations $x \in X$ of the image. Similarly, V^t is a flow field at time slice t defined over the image range with entries $v_x^t \in W$ at each pixel location x of the image.

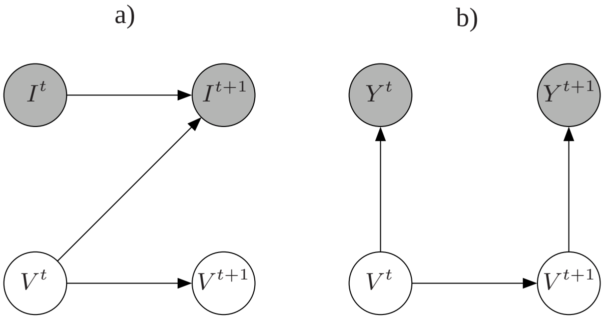

Figure 1. Dynamic Bayesian Network for motion estimation.

To define the model precisely we need to specify 1) the observation likelihood $P(I^{t+1} \mid V^t, I^t)$ of a pair of images I^{t+1} and I^t and 2) the transition probability $P(V^{t+1} \mid V^t)$ of the flow field. To simplify the notation we can introduce an alternative observation variable $Y^t = (I^{t+1}, I^t)$ that subsumes a pair of consecutive images. Since images are observed, the likelihood $P(I^t)$ in the term $P(I^{t+1} \mid V^t, I^t)\ P(I^t) = P(I^{t+1}, I^t \mid V^t)$ is only a constant factor we can neglect. This leads to $P(I^{t+1} \mid V^t, I^t) \propto P(I^{t+1}, I^t \mid V^t) = P(Y^t \mid V^t)$ and the corresponding Dynamic Bayesian Network shown in Fig. 1 b). For both the observation likelihood $P(Y^t \mid V^t)$ and the V-transition probability $P(V^{t+1} \mid V^t)$ we assume that they factorize over the image w.r.t. V^t and V^{t+1}, i.e.,

$$\zeta(V^t) := P(Y^t \mid V^t) = \prod_x \ell(Y^t \mid v_x^t) \tag{1}$$

$$P(V^{t+1} \mid V^t) = \prod_x P(v_x^{t+1} \mid V^t)\ . \tag{2}$$

This allows us to maintain only factored beliefs over V^t during inference, which makes the approach computationally feasible.

2.1. Observation likelihood

We define the observation likelihood $P(Y^t \mid V^t)$ by assuming that the likelihood factor $\ell(Y^t \mid v_x^t)$ of a local velocity v_x^t should be related to finding the same or similar image patch centered around x at time $t+1$ that was present at time t but centered around $x - v_x^t \Delta t$[1]. More rigorously, let $\mathcal{S}(x, \mu, \Sigma, \nu)$ be the Student's t-distribution and

[1] In the following, we neglect dimensions and set $\Delta t = 1$, so $v_x^t \Delta t$ gets v_x^t.

$\mathcal{N}(x,\mu,\Sigma) = \lim_{\nu\to\infty}\mathcal{S}(x,\mu,\Sigma,\nu)$ be the normal distribution of a variable x with mean μ, covariance matrix Σ and the degrees of freedom ν. In the following the covariance is chosen to be isotropic $\Sigma = \sigma^2 E$ (with identity matrix E). We define

$$\begin{aligned}\ell(Y^t|v_x^t) &= \sum_{x'}\mathcal{N}(x',x,\varrho_I)\mathcal{S}(I_{x'}^{t+1}, I_{x'-v_x^t}^t, \sigma_I, \nu_I) \quad (3)\\ &= \sum_{x'}\mathcal{N}(x',x-v_x^t,\varrho_I)\mathcal{S}(I_{x'+v_x^t}^{t+1}, I_{x'}^t, \sigma_I, \nu_I)\,.\end{aligned}$$

Here, $\mathcal{N}(x',x,\varrho_I)$ implements a Gaussian weighting of locality centered around x for I^{t+1} and around $x - v_x^t$ for I^t. The parameter ϱ_I defines the spatial range of this image patch and σ_I the grey value variance. The univariate Student's t-distribution $\mathcal{S}(I_{x'}^{t+1}, I_{x'-v_x^t}^t, \sigma_I, \nu_I)$ realizes a robust behaviour against large gray-value differences within image patches, which means these gray-values are treated as outliers and are much less significant for the distribution.

2.2. Transition probability

Similarly to equation (3), we define the transition probability $P(V^{t+1}\,|\,V^t)$ by assuming that the flow field transforms according to itself. To motivate the definition we assume that the origin of a local flow vector v_x^{t+1} at position x was a previous flow vector $v_{x'}^t$ at some corresponding position x',

$$v_x^{t+1} \sim \mathcal{S}(v_x^{t+1}, v_{x'}^t, \sigma_V, \nu_V)\,. \tag{4}$$

So, we assume robust spatiotemporal coherence because evaluations on first derivative optical flow statistics [10] and on prior distributions that allow to imitate human speed discrimination tasks [13] provide strong indication that they resemble heavy tailed Student's t-distributions. Now, asking what the corresponding position x' in the previous image was, we assume that we can infer it from the flow field itself via

$$x' \sim \mathcal{N}(x', x - v_x^{t+1}, \varrho_V)\,. \tag{5}$$

Note that here we use v_x^{t+1} to retrieve the previous corresponding point. Combining both factors and integrating x' we get[2]

$$P(v_x^{t+1}\,|\,V^t) \propto \sum_{x'}\mathcal{N}(x', x-v_x^{t+1}, \varrho_V)\,\mathcal{S}(v_x^{t+1}, v_{x'}^t, \sigma_V)\,, \tag{6}$$

which is of similar form as (3). The parameter ϱ_V defines the spatial range of a flow-field patch, so we compare velocity vectors within flow-field patches at different times t and $t+1$. The relations for the transition model are sketched in Fig. 2. We introduced new parameters ϱ_V and σ_V for the uncertainty in spatial identification between two images and the transition noise between V^t and V^{t+1}, respectively.

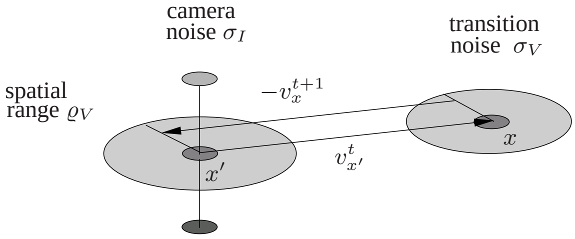

Figure 2. Uncertainty assumptions for the V-transitions and the camera noise.

The robustness against outliers is controlled by ν_V, with smaller/larger ν_V decreasing/increasing the influence of incoherently moving pixels within the observed spatial range ϱ_V. With $\nu_V \to \infty$ the uncertainty for the velocity gets Gaussian distributed and (6) equals the transition probability formulated in [4] which expresses the belief that pixels *are, on average, moving along a straight line with constant velocity*. Therefore, our spatiotemporal transition model can be seen as a generalization of the transition model proposed by Burgi *et al*. [4].

2.3. Two-filter inference

For inference we need to propagate beliefs over the flow field V^t. Storing a distribution over a whole flow field V^t is infeasible if one does not make factorization assumptions. The factored observation likelihoods and transition probabilities we introduced ensure that the forward propagated beliefs will remain factored. However, the standard backward messages do not exactly factor under this model. Hence we follow a two-filter approach [6] where the "backward filter" is strictly symmetric to the forward filter.

The forward filtering equations read

$$\begin{aligned}\alpha(V^t) &:= P(V^t\,|\,Y^{1:t}) = \prod_x \alpha(v_x^t)\,, &(7)\\ \alpha(v_x^{t+1}) &\propto \ell(Y^{t+1}\,|\,v_x^{t+1})\,\alpha^*(v_x^{t+1})\,, &(8)\\ \alpha^*(v_x^{t+1}) &\propto \sum_{V^t} P(v_x^{t+1}\,|\,V^t)\,\alpha(V^t)\,. &(9)\end{aligned}$$

By inserting (6) and (7) in (9) the right side of (9) is

$$\begin{aligned}&\sum_{V^t}\sum_{x'}\mathcal{N}(x', x-v_x^{t+1}, \varrho_V)\mathcal{S}(v_x^{t+1}, v_{x'}^t, \sigma_V)\ \prod_z \alpha(v_z^t) =\\ &\sum_{x'}\ \mathcal{N}(x', x-v_x^{t+1}, \varrho_V)\ \times\\ &\sum_{v_{x'}^t}\ \mathcal{S}(v_x^{t+1}, v_{x'}^t, \sigma_V)\alpha(v_{x'}^t)\ \sum_{V^t\setminus v_{x'}^t}\ \prod_{z\neq x'}\ \alpha(v_z^t)\end{aligned} \tag{10}$$

Note that the summation $\sum_{V^t\in W^X}$ is summing over all possible flow fields (X is the pixel range), i.e. it represents $|X|$ summations $\sum_{v_1^t\in W}\sum_{v_2^t\in W}\sum_{v_3^t\in W}\cdots$ over each local flow field vector. We separated these into a summation $\sum_{v_{x'}^t}$ over the flow field vector at x' and a summation $\sum_{V^t\setminus v_{x'}^t}$ over all other flow field vectors at $x \neq x'$. We can

[2] Due to shortage of space ν_V is omitted in some notations.

use $\sum_{V^t \setminus v^t_{x'}} \prod_{z \neq x'} \alpha(v_z^t) = \prod_{z \neq x'} \sum_{v_z^t} \alpha(v_z^t) = 1$ and the final forward filtering reduces to

$$\begin{aligned} \alpha^*(v_x^{t+1}) \quad \propto \quad & \sum_{x'} \mathcal{N}(x', x - v_x^{t+1}, \varrho_V) \times \\ & \sum_{v^t_{x'}} \mathcal{S}(v_x^{t+1}, v^t_{x'}, \sigma_V, \nu_V)\ \alpha(v^t_{x'})\ . \end{aligned} \tag{11}$$

If we have access to a batch of data (or a recent window of data) we can compute smoothed posteriors as a basis for an EM-algorithm and train the free parameters. In our two-filter approach we derive the backward filter as a mirrored version of the forward filter, but using

$$P(v_x^t \,|\, V^{t+1}) \propto \sum_{x'} \mathcal{N}(x', x + v_x^t, \varrho_V)\ \mathcal{S}(v_x^t, v^{t+1}_{x'}, \sigma_V) \tag{12}$$

instead of (6). This equation is motivated in exactly the same way as we motivated (6): we assume that $v_x^t \sim \mathcal{S}(v^{t+1}_{x'}, \sigma_V, \nu_V)$ for a corresponding position x' in the subsequent image, and that $x' \sim \mathcal{N}(x - v_x^t, \varrho_V)$ is itself defined by v_x^t. However, note that using this symmetry of argumentation is actually an approximation to our model because applying Bayes rule on (6) would lead to a different, non-factored $P(V^t \,|\, V^{t+1})$. What we gain by the approximation $P(V^t \,|\, V^{t+1}) \approx \prod_x P(v_x^t \,|\, V^{t+1})$ are factored β's which are feasible to maintain computationally. The backward filter equations read

$$\beta(V^t) \quad := \quad P(V^t \,|\, Y^{t+1:T}) = \prod_x \beta(v_x^t)\ , \tag{13}$$

$$\beta^*(v_x^t) \quad \propto \quad \ell(Y^t \,|\, v_x^t)\ \beta(v_x^t)\ , \tag{14}$$

$$\beta(v_x^t) \quad \propto \quad \sum_{V^{t+1}} P(v_x^t \,|\, V^{t+1})\ \beta^*(V^{t+1})\ . \tag{15}$$

In analogy to the derivations for the forward filtering (10) we arrive at the final backward filtering equation

$$\begin{aligned} \beta(v_x^t) \quad \propto \quad & \sum_{x'} \mathcal{N}(x', x + v_x^t, \varrho_v)\ \times \\ & \sum_{v^{t+1}_{x'}} \mathcal{S}(v_x^t, v^{t+1}_{x'}, \sigma_V, \nu_V) \beta^*(v^{t+1}_{x'})\ . \end{aligned} \tag{16}$$

To derive the smoothed posterior we need to combine the forward and backward filters. In the two-filter approach this reads

$$\begin{aligned} \gamma(v_x^t) \quad := \quad & P(v_x^t \,|\, Y^{1:T}) = \frac{P(Y^{t+1:T} \,|\, v_x^t)\ P(v_x^t \,|\, Y^{1:t})}{P(Y^{1:T})} \\ = \quad & \frac{P(v_x^t \,|\, Y^{t+1:T}) P(Y^{t+1:T}) P(v_x^t \,|\, Y^{1:t})}{P(v_x^t) P(Y^{1:T})} \\ \propto \quad & \alpha(v_x^t)\ \beta(v_x^t)\ \frac{1}{P(v_x^t)}\ , \end{aligned} \tag{17}$$

with $P(Y^{t+1:T})$ and $P(Y^{1:T})$ being constant. If both the forward and backward filters are initialized with $\alpha(v_x^0) = \beta(v_x^T) = P(v_x)$ we can identify the unconditioned distribution $P(v_x^t)$ with the prior $P(v_x)$.

3. Uncertainty Optimization

Based on the filter equations (11, 16, 17) we can apply the EM-algorithm for an optimization of the free parameters $\theta = \{\sigma_I, \sigma_V\}$. Starting with an initial setting for the parameters θ^o ("o" for old) the offline *E-step* is the evaluation of the smoothed posterior $\gamma(V^t)$ using the two-filter inference as given by (17) keeping the old parameters θ^o fixed. In case of online optimization the forward filtered posterior $\alpha(V^t)$ is evaluated as given by (8) and (11). If only the observed likelihood $\zeta(V^t)$ with its assumed gray-value uncertainty σ_I should be optimized no a priori preference for the velocity is considered (which is equivalent to a time-independent equally distributed prior $\alpha^*(v_x^t) = |W|^{-1}$). The exact *M-step* determines the revised parameter estimate θ^n ("n" for new) by maximizing the expected complete data log-likelihood under the posterior distribution

$$\theta^n = \operatorname*{argmax}_\theta \sum_{V^t} P(V^t | Y^{1:T}, \theta^o) \ln P(Y^{1:T}, V^t | \theta)\ . \tag{18}$$

Now, instead of maximizing this expectation we use the MAP approximation of the M-step, maximizing the log-likelihood only for the flow field $\hat{V}^t$ which maximizes the posterior distribution $P(\hat{V}^t | Y^{1:T}, \theta^o)$. Taking advantage of the assumed factorization of the posterior (7) and that $P(Y^{1:T})$ is constant, we get

$$\theta^n = \operatorname*{argmax}_\theta \sum_{t,x} \gamma(\hat{v}_x^t, \theta^o) \ln \gamma(\hat{v}_x^t, \theta)\ . \tag{19}$$

After assignment of the new parameters $\theta^o \leftarrow \theta^n$ the E- and M-steps are evaluated anew until a convergence criterion is fullfilled. The maximization of (19) for the arguments σ_I, σ_V leads to the following M-step approximations:

$$\sigma_I = \frac{1}{Z} \sum_{t,x} \gamma(\hat{v}_x^t)(I_x^t - (\hat{\mathcal{T}} \circ I_x^{t+1})_x)^2\ , \tag{20}$$

$$\sigma_V = \frac{1}{Z} \sum_{t,x} \gamma(\hat{v}_x^t) || \hat{v}_x^t - (\hat{\mathcal{T}} \circ \hat{v}_x^{t+1})_x ||^2\ , \tag{21}$$

with $Z = \sum_{x,t} \gamma(\hat{v}_x^t)$, $\hat{v}_x^t = MAP(\gamma(v_x^t))$ being the maximum a posteriori estimate of the velocities that describe the estimated flow-field $\hat{V}^t$ and $\hat{\mathcal{T}} \circ$ being the operator for the reverse mapping of image features I_x^{t+1} resp. $\hat{v}_x^{t+1}$ using the estimated flow field $\hat{V}^t$ and bilinear interpolation.

Here, the degrees of freedom ν_I, ν_V which are used for judging the robustness of the analyses are not optimized so far (fixed during EM). Also the covariances ϱ_I, ϱ_V that define to what spatiotemporal extent motion coherence is assumed are specified.

The online M-steps are the same as the offline ones, but with $\gamma(v_x^t)$ replaced by $\alpha(v_x^t)$ and instead of summing over t we update the parameters at each time step with a fixed learning rate. Therefore, adaptation of the parameters over time is possible which is dependent on whether movements within the scene change over time or remain spatially stationary.

4. Evaluation

To evaluate the performance of the proposed Dynamic Bayesian Network two measurements are used. The first one is the well known error measure proposed by Barron *et al.* [1] which is also utilized in [3, 10]. It shows the mean error between the estimated $\hat{V}^t$ and the ground-truth $\check{V}^t$ flowfield

$$\overline{e}_V = \frac{1}{|X|} \sum_x \arccos\left((\hat{v}_x^t)_h^T (\check{v}_x^t)_h\right) , \qquad (22)$$

with $(\hat{v}_x^t)_h$ being the estimated velocity vector and $(\check{v}_x^t)_h$ the ground truth velocity vector both written in homogeneous coordinates. The second criterion has been proposed by Burgi *et al.* [4] to evaluate the evolution of the flow-field estimation over time. It is called the *sharpness* s which is the *Kullback-Leibler divergence* between the estimated probability for a velocity vector $P(v_x^t|Y^{1:T})$ and a uniform distribution U. Usually, the higher the sharpness, the more precise the velocity estimate [4], so we take the spatial mean of the sharpness $\overline{s}$ for the evaluation which is

$$\overline{s} = \frac{1}{|X|} \sum_x \sum_{v_x^t} P(v_x^t|Y^{1:T}) \ln |W| P(v_x^t|Y^{1:T}) \qquad (23)$$

In our momentary implementation every $v_x^t \in W, W = \mathbb{Z}^2$, is a pixel-wise discretized vector of signed integer velocities, which means, the MAP estimator cannot achieve subpixel accuracy.

4.1. Robustness against noise and outliers

To judge the effect of camera noise on the quality of the observed likelihood $\zeta(V^t)$ we use a synthetic test sequence $I^{1:T}$ consisting of a static background pattern $I_b^{1:T} \in [0; 150]$ and a moving circular pattern $I_o^{1:T} \in [200; 255]$ with constant velocity $v_o = 2\,\text{pixels}/\Delta t$. In a first trial we train the corresponding parameter σ_I for increasing ground-truth Gaussian noise $\check{\sigma}_I$ added on the image gray values of the test sequence and for $\nu = 0.1$ and $\nu = \infty$. The spatial weighting of the patches remains fixed at $\rho_I = 5$.

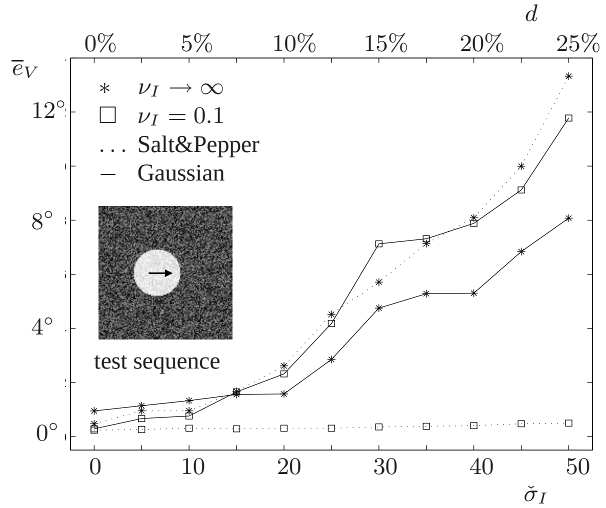

Figure 3. Mean optical flow errors $\overline{e}_V$ for increasing Gaussian noise $\check{\sigma}_I$ and Salt&Pepper Noise d for different degrees of freedom ν_I.

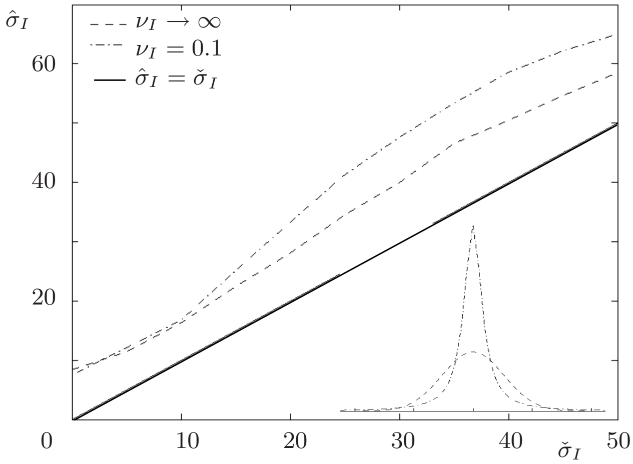

Figure 4. Estimated standard deviations $\hat{\sigma}_I$ for increasing Gaussian noise $\check{\sigma}_I$ on the gray-values of the images I.

In Fig. 3 the calculated mean errors $\overline{e}_V$ for $\nu = 0.1$ (□) and $\nu \to \infty$ ($*$) for Gaussian noise starting with standard deviation $\check{\sigma}_I = 0$ up to $\check{\sigma}_I = 50$ and in Fig. 4 the corresponding optimized standard deviations $\hat{\sigma}_I$ after 10 EM-iterations are shown. Comparing the two solid lines in Fig. 3, for Gaussian camera noise up to $\check{\sigma}_I = 15$ the increase of robustness for $\nu = 0.1$ can clearly be seen. This is because at the motion boundaries along the edge of the moving circular pattern better velocity estimates can be achieved. With increasing Gaussian noise the assumption that the noise is distributed like a Student's t-distribution with heavy tails is no longer sufficient and noisy gray value differences between corresponding image points are increasingly mistaken as outliers. For stronger Gaussian camera noise (starting from $\check{\sigma}_I = 15$) the limit $\nu \to \infty$ which is equivalent to a Gaussian distribution gives better estimation results. This is obvious because Gaussian noise was added to the

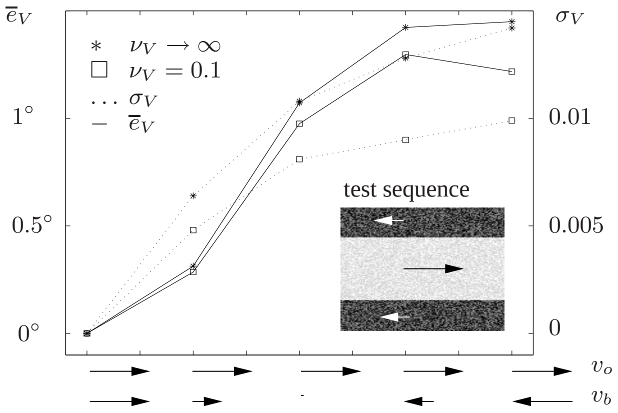

Figure 5. Mean error $\overline{e}_V$ and estimated transition uncertainty σ_V for increasing motion discontinuities along the border of a moving object in front of a differently moving background.

synthetic scene and the errors at the motion boundaries no longer carry weight for the mean error $\overline{e}_V$. The optimization of the standard deviation $\hat{\sigma}_I$ works nicely and reflects the ground-truth Gaussian noise $\check{\sigma}_I$ up to a bias (see Fig. 4) which arises because of numerical errors in the warping procedure (If no Gaussian noise is added $\check{\sigma}_I = 0$ and the image I^{t+1} is warped with the ground truth flow field $\check{V}^t$ to get I^t using the M-step (20) with $\gamma(\check{v}_x^t) = 1$ we get an estimated uncertainty of $\hat{\sigma}_I = 9.59$).

To evaluate the robustness properties of the Student's t-distribution further, the same procedure but for Salt&Pepper noise is also shown in Fig. 3. Here, the mean error for increasing density $d \in [0; 0.25]$ of Salt&Pepper noise is plotted. Comparing the two dotted lines in Fig. 3, up to a density of $d = 0.25$ the mean error $\overline{e}_V$ is not affected for $\nu = 0.1$. For $\nu \to \infty$ the performance deteriorates with increasing noise because the Gaussian noise assumption does not hold anymore for Salt&Pepper noise.

With a second example the behavior at motion discontinuities is evaluated. Here, a random pattern $I_o^{1:T} \in [200; 255]$ moves with constant velocity $v_o = 2\,\text{pixels}/\Delta t$ in front of a moving background $I_b^{1:T} \in [0; 150]$ that changes its velocity $v_b \in [-2\,\text{pixels}/\Delta t; 2\,\text{pixels}/\Delta t]$ in every trial. The more different the velocity of the background v_b compared to the object motion v_o the stronger motion discontinuities are along the object borders and the more the assumption for the flow field evolution is violated that the pixels move coherently within the spatial region $\rho_V = 5$. Fig. 5 shows the mean error $\overline{e}_V$ for robust (□) and quadratic ($*$) penalisation of differences between velocity vectors within flow-field patches. If background and object move totally coherent then the error $\overline{e}_V$ and also the uncertainties σ_I, σ_V tend to zero. For the quadratic penalisation the stronger the motion discontinuities get the higher the error is. But for the robust penalisation the error saturates and even reduces for very large discontinuities. This means if the ratio of the numbers of differently moving pixels within a flow patch differs to a certain degree from one then the pixels that belong to the lower number are treated as outliers.

4.2. Disambiguation of motion uncertainties

Next the properties of the forward- and backward-priors $\alpha^*(V^t), \beta(V^t)$ and in particular the effects of the spatiotemporal coherence assumption for the flow-field patches are discussed. For this purpose, a black square $I_o^{1:T} = 0$ is moved in front of a static background pattern $I_b^{1:T} \in [200; 255]$ with constant velocity $v_o = 2\,\text{pixels}/\Delta t$ for $T = 40$ frames, fixed spatial extensions $\rho_I = 5, \rho_V = 35$, and an equally distributed initialisation of the posteriors $\alpha(v_x^0) = \beta^*(v_x^{41}) = |W|^{-1}$. This synthetic sequence is strongly confronted with the aperture problem along the edges of the square and with the blank wall problem because of the untextured surface of the square. In addition, pixels from the background appear and disappear along the motion boundaries. In Fig. 6 the flow fields for different points in time estimated from the forward filtered posterior $\alpha(V^t)$, the backward filtered posterior $\beta^*(V^t)$, and the two-filter smoothed posterior $\gamma(V^t)$ are shown. At the beginning of forward $\alpha(V^1)$ and backward $\beta^*(V^{40})$ filtering only at the corners of the square the correct velocity can be estimated. With ongoing filtering the motion ambiguities along the edges and within the square are resolved and the square is continuously filled-in with improved velocity estimates. The combination of both in $\gamma(V^t)$ leads to even better results because at every timestep the optical flow is inferred from "having the whole image sequence in mind" with the best result for $\gamma(V^{20})$ because both filtering and smoothing have seen an equal amount of images (which is 20 images). This result is also reflected in Fig. 7 with the mean sharpness $\overline{s}^t$ plotted for the likelihood $\zeta(V^t)$, the forward filtered posterior $\alpha(V^t)$, the backward filtered posterior $\beta^*(V^t)$ and the two-filter smoothed posterior $\gamma(V^t)$. Therefore, the most peaked distributions are gained from two-filter smoothing.

As a first proof of principle for the *online* filtering capabilities of the framework the same sequence is used with Gaussian noise of $\check{\sigma}_I = 10$ added. In Fig. 8 the mean error $\overline{e}_V^t$ and the sharpness $\overline{s}_\alpha^t$ are plotted over time. It can be seen that the error continuously reduces and the sharpness increases.

Also for the Yosemite benchmark [1], as shown in Fig. 9, a continuous reduction of the optical flow errors over time can be achieved with the proposed probabilistic method. To be able to extract subpixel accuracy the MMSE estimator instead of the MAP estimator is applied. Although, a very simple region-based matching measurement (3) is used that exploits the SSD only for signed integer velocities, the rel-

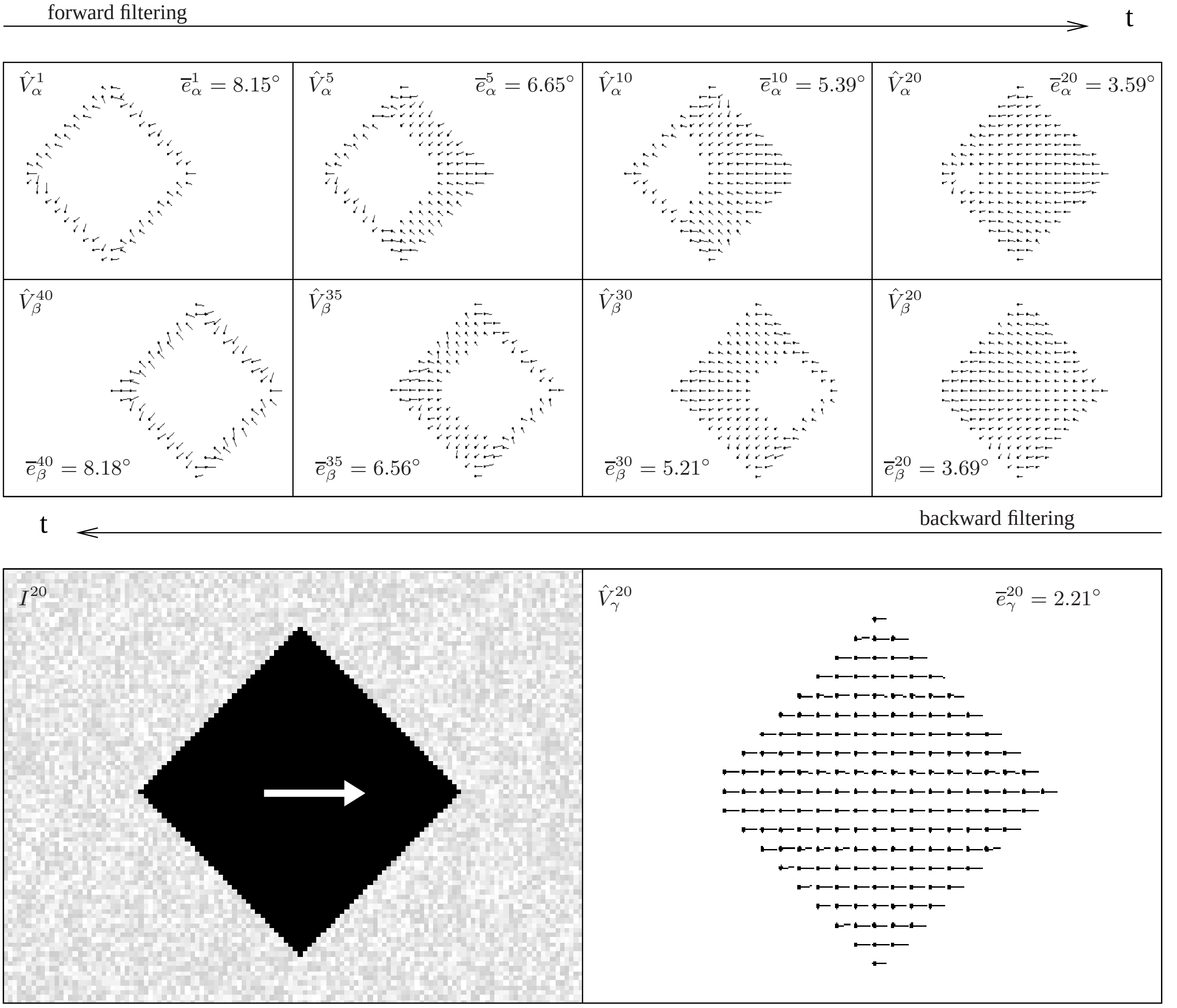

Figure 6. Snapshots of the forward and backward optical flow estimations at different points in time for the complete observed data of 40 images of a sequence. The nice filling-in properties of the model for the regions that lack of texture (solution to the blank wall problem) and the disambiguation at moving edges (solution to the aperture problem) can clearly be noticed.

ative improvement of the flow-field estimate over time is quite good. For the online case the error reduces about 69.6%, from $\overline{e}_\alpha^1 = 39.5°$ to $\overline{e}_\alpha^{13} = 12.1°$, and for the offline case about 73.4%, from $\overline{e}_\alpha^1 = 39.5°$ to $\overline{e}_\gamma^7 = 10.5°$. Recent gradient-based optical flow techniques [3, 10] achieve better accuracy on the Yosemite sequence but are not suited for image sequences where numerical differentiation is impractical [1]. Our result $\overline{e}_\gamma^7 = 10.53°$ compares quite favorably with other region-based matching methods described in [1] with the lowest mean error of $\overline{e}_V = 13.16°$ for Singh's method. Future work will investigate propagating beliefs over a continuous flow-field domain $W = \mathbb{R}^2$ to achieve better sub-pixel accuracy.

5. Conclusions and Future Work

We have presented a robust two-filter inference approach to continuously estimate the optical flow of image sequences. It allows for the optimization of uncertainties that reflect the momentary transition noise of the scene movement and the momentary camera noise on the pixel intensities. Although the transition probability holds only a simple spatiotemporal smoothness constraint the system is able to resolve ambiguous local motion measurements and demonstrates robust behavior at motion discontinuities. No learned scene-specific prior knowledge is incorporated, but an adaptation to the scene by optimizing uncertainty parameters is realized. Because of the linear prediction assump-

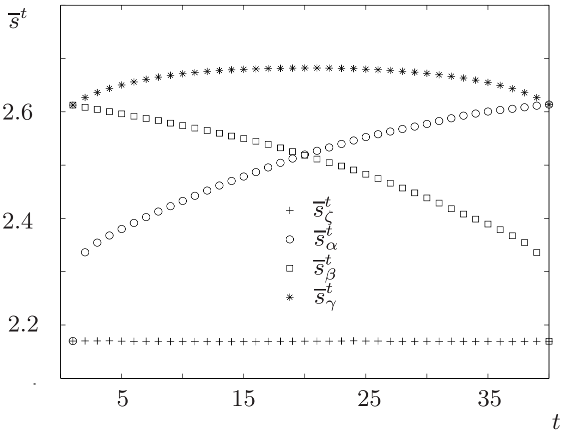

Figure 7. Mean sharpness $\overline{s}^t$ for the observed likelihood, the forward-, backward-, and two-filter inference.

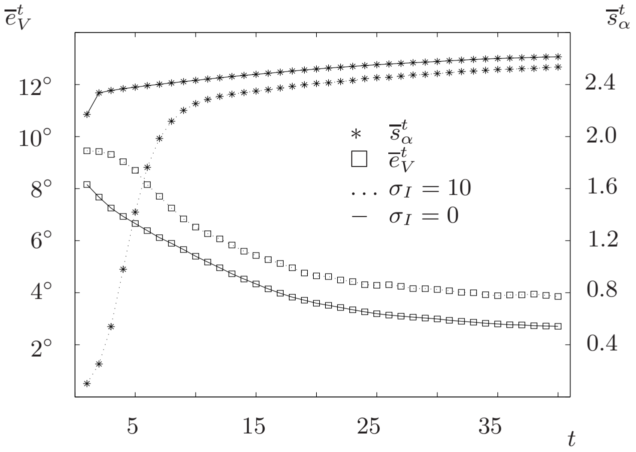

Figure 8. Time dependent mean error $\overline{e}^t_V$ and sharpness $\overline{s}^t_\alpha$ for online parameter optimization.

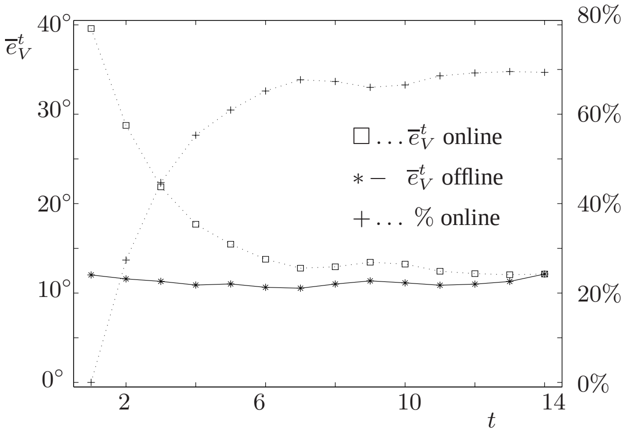

Figure 9. Mean error $\overline{e}^t_V$ for online □ and offline $*$ optimization and the percental improvement $+$ for the online case.

tion rapidly changing movements corrupt the estimation results. As long as the scene movement does not change (or only slightly changes), the optical flow estimation improves over time. Future investigations will include the replacement of the measurement method for the observation likelihood with more accurate gradient-based methods that allow for a continuous flow-field domain and the adaptation of further parameters, like the robustness parameter ν.

References

[1] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *IJCV*, 12(1):43–77, 1994.

[2] M. Black and P. Anandan. Robust dynamic motion estimation over time. In *CVPR*, pages 296–302, Maui, Hawaii, 1991.

[3] A. Bruhn, J. Weickert, and C. Schnörr. Lukas / kanade meets horn / schunk: Combining local and global optic flow methods. *IJCV*, 61(3):211–231, 2005.

[4] P. Burgi, A.L.Yuille, and N. Grzywacz. Probabilistic motion estimation based on temporal coherence. *Neural Computation*, 12(8):1839–1867, 2000.

[5] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. *PAMI*, 15(12):1217–1232, 1993.

[6] G. Kitagawa. The two-filter formula for smoothing and an implementation of the gaussian-sum smoother. *Annals Institute of Statistical Mathematics*, 46(4):605–623, 1994.

[7] K. Krajsek and R. Mester. A maximum likelihood estimator for choosing the regularization parameters in global optical flow methods. In *ICIP*, pages 1081–1084, Atlanta, USA, 2006.

[8] K. Lim, A. Das, and M. Chong. Estimation of occlusion and dense motion fields in a bidirectional bayesian framework. *PAMI*, 24(5):712–718, 2002.

[9] K. Murphy. Dynamic bayesian network: Representation, inference and learning. *PhD Thesis, UC Berkeley, Computer Science Division*, 2002.

[10] S. Roth and M. Black. On the spatial statistics of optical flow. In *ICCV*, pages 42–49, Beijing, China, 2005.

[11] E. Simoncelli, E. Adelson, and D. Heeger. Probability distributions of optical flow. In *IEEE International Conference on Computer Vision and Pattern Recognition, CVPR*, pages 310–315, Maui, Hawaii, 1991.

[12] A. Singh. Incremental estimation of image flow using a kalman filter. In *IEEE Workshop on Visual Motion*, pages 36–43, 1991.

[13] A. A. Stocker and E. P. Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–585, 2006.

[14] Y. Weiss and D. Fleet. Velocity likelihoods in biological and machine vision. In *Probabilistic Models of the Brain: Perception and Neural Function*, pages 77–96. MIT Press, 2002.

[15] V. Willert, J. Eggert, J. Adamy, and E. Körner. Non-gaussian velocity distributions integrated over space, time, and scales. *IEEE Transactions on Systems, Man and Cybernetics - Part B*, 36(3):482–493, 2006.